Species Identification in Ancient and Heated Bone Fragments Using Protein Mass Spectrometry

by

Michael Buckley

A thesis submitted to the University of York in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Biology, Faculty of Science

BioArCh, Departments of Biology, Archaeology and Chemistry, University of York

July 2008

ABSTRACT

The primary aim of this research was to investigate the use of protein mass spectrometry for species identification in heat-degraded (meat and bone meal; MBM), and time-degraded (archaeological and palaeontological) bone material. As protein degradation in a burial environment over time is related to the burial temperature (its 'thermal age'), the degradation of buried bone over time should follow a broadly similar pattern to the heat-treatment of bone. An opportunity to use proteins for species identification stems from difficulties with using degraded DNA sequences that are due to diagenetic alteration of DNA and succeptibility to contamination. The principal aim of this research is to develop a method of analysing peptide markers that persist in ancient bone beyond the limits of amplifiable ancient DNA. Following an in-depth introduction to bone structure, proteins and soft-ionisation mass spectrometry (Chapter 1), the previously established archaeological protein species-specific marker osteocalcin was more thoroughly investigated in archaeological samples (Chapter 2) and applications to MBM (Chapter 3). Due to limitations of osteocalcin analyses highlighted by these studies, characterisation of MBM was carried out using LC-MS-based methods as well as amino acid analyses (Chapter 4). As collagen (I) was clearly identified as the dominant protein in MBM and is known to survive in palaeontological bone, two methods to simplify the analyses of species-specific collagen peptides were investigated (Chapters 5-8). One of these methods was to use the enzyme bacterial collagenase and solid phase extraction (SPE) to purify the collagen α2 (I) chain carboxyterminal telopeptide (Chapter 5) and the survival of this telopeptide marker in archaeological samples devoid of ancient DNA was confirmed (Chapter 6). An alternate method uses the enzyme trypsin and SPE to isolate four collagen α2 (I) chain peptides from the helical region of the molecule (Chapter 7), which complement the telopeptide approach by increasing the taxonomic resolution of the species identification, in particular allowing distinction between morphologically-similar archaeological sheep and goat bone fragments (Chapter 8). These collagen-based methods of species-identification, including LC-MS-based approaches were then applied to both the heat-rendered MBM (Chapter 9) and a range of archaeological and palaeontological samples (Chapter 10). Possible problems with sequencing exceptionally-ancient proteins are considered in a reply to a recent paper claiming to sequence *Tyrannosaurus rex* collagen (Chapter 11). Chapter 12 presents a general discussion of the outcomes of this thesis.

LIST OF CONTENTS

LIST OF TABLES

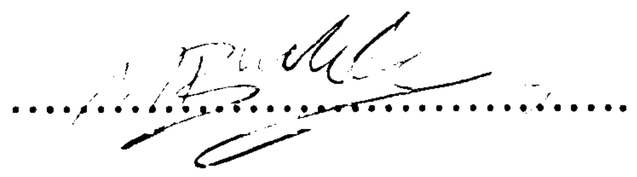
LIST OF FIGURES

ACKNOWLEDGEMENTS

This project was funded by the Natural Environment Research Council (NERC), with additional funding from Prosper De Mulder (PDM) the NSF grant EAR-0309467 to Matthew Collins and Peggy Ostrom. I would like to thank supervisors Matthew Collins and Jane Thomas-Oates for their guidance throughout this research, and John Currey, Kirsty Penkman and Hannah Koon for advice and proof-reading. I would like to thank Camilla Nichols, Hannah Koon, Jerry Herman, Nigel Larkin, Adrian Lister, Sarah Kansa, Paul Langthorne, Andrew Chamberlain, and Naiomi Sykes for their important samples. Specifically, I would like to acknowledge the help of Brian Raney for sequence acquisitions in Chapters 2, 5, 6 and 9, Matthew Collins for his work in EST sequence compilation in Chapter 9 and Sarah Howard, Stuart Campbell and Sarah Whitcher Kansa for their help in the work and writing of Chapter 8. Additional thanks must go to Peggy Ostrom, whom took me under her wing during the early stages of this research, and Terry O'Connor for enduring my numerous bizarre sample requests and discussions regarding further potential archaeological applications of species identification using biomolecules. Thanks must also go to the dedicated staff of the Department's Technology Facility (TF), namely Jerry Thomas, Adam Dowle, Dave Ashford (Analytical Biochemistry), Jarred Cartwright, Mick Miller (Protein Production), Andrew Leech and Berni Strongitharm (Molecular Interactions).
I also thank Sarah for encouragement and inspiration.

AUTHOR'S DECLARATION

Three chapters presented in this thesis were published prior to submission; I proposed the research, designed the experiments, carried out the osteocalcin and amino acid analyses and wrote the manuscript that constitutes Chapter 2; I proposed the research, designed the experiments, carried out the analyses and wrote the manuscript that constitutes Chapter 5; Matthew Collins proposed the idea and wrote the manuscript, whereas I carried out amino acid analyses, re-interpretation of data and contributed to the writing of the comment that constitutes Chapter 11. I hereby certify that the work described in this thesis is my own, except where otherwise acknowledged. This work has not been submitted previously for a degree at this, or any other university.



1 Literature Review

This thesis investigates the use of protein fragments to recover species identity in degraded (archaeological/fossil and heat treated) bone. The literature review first considers the importance of species identification in archaeology and industrial rendering practices. This is followed by a description of the structure of bone and the relative survival of biomolecules in ancient bone, as it is the most common organic tissue found in archaeological and palaeontological deposits. This includes a detailed description of the two most dominant proteins, collagen and osteocalcin (OC), which are the focus biomolecules of this thesis. The problems of identification due to protein degradation (such as natural degradation in the environment and the hydrolysis of proteins in the process of Meat and Bone Meal (MBM) rendering) are then considered. Finally, a detailed introduction to soft-ionisation forms of mass spectrometry for protein analysis is then described.

1.1 Species Identification from Biomolecules

In zoology, the taxon genus is a collection of species that are similar and presumed to have a common phylogenetic origin, having evolved from a common ancestor and all species in a genus are considered to occupy a similar niche though one that is broader than that of a single species (Mayr *et al.* 1953). The classification of animals is codified and the procedures for making changes are subject to rules governed by the International Code of Zoological Nomenclature. Scientific nomenclature is dynamic and changes are made even among the best-known groups of animals (Reitz and Wing 1999). The naming of species is based on Linnaeus' Systema Naturae (1758 cited in Reitz and Wing 1999), where the binomial nomenclature refers to a hierarchical system with clearly defined species diagnoses based on their similarities. These similarities were originally observed via morphological characteristics but more recently the study of molecular similarities has proven useful in analysing relationships between species. Some groups of animals are more subject to classification change than others as more is learned about them and their phylogenetic relationships are better understood; microbiological studies of DNA reveal relationships between animals that are not clear on morphological grounds alone. Not only has research focussed on identifying relationships between extant taxa, but also in resolving such relationships in extinct taxa.

Likewise, similar morphological criteria for species distinction are applied to remains of extinct species, but these are somewhat more limited because often only the hard skeletal tissues remain for analysis, and thus much less morphological information is available. Even this is further complicated by taphonomic and diagenetic processes, often rendering many fragments of bone undiagnostic. More recently however, investigations into molecular similarities have proven very useful in taxonomic identifications and phylogenetic analyses of ancient and extinct species. The widespread potential of biomolecular species identification is apparent in the many palaeontological site species lists that have a significant proportion of samples remaining unidentified (Stuart 1975; Stuart 1982).

1.1.1 Species Identification in Ancient Bone

The past two million years, known as the Quaternary Period, are of particular interest to mankind because it covers much of hominin evolution. The Quaternary Period, divided into the Pleistocene and Holocene, is characterised by extraordinary fluctuations in global climate; in northwestern Europe, the latter half of this period (the last 750,000 years) have been characterised by long periods (ca. 100,000 years) of cold climate interspersed with shorter periods (ca. 10-15,000 years) of warmer conditions. A chronology of these different stages using oxygen isotope stages (OIS) was proposed by Emiliani (1955), later modified by Shackleton (1967), where odd numbers are given to interglacials and even numbers given to the glacials. This Quaternary climate change has produced a geological record dominated by sediments deposited under glacial, periglacial and temperate environmental conditions. However, it is extremely unusual to find neatly ordered deposits of alternating glacial (cold stage) and interglacial (temperate stage) sediments, but partial records that are difficult to piece together and which remain the subject of much debate.

The identification of both animal and plant species are often used as indicators of climate as well as other environmental conditions and have been particularly widely investigated in the British record (e.g. Currant and Jacobi 2001; Stuart and Lister 2001). Schreve (2001) demonstrated that biostratigraphical evidence from fossil mammal assemblages is an effective tool for establishing the number and nature of different climatic episodes in the Middle Pleistocene in Britain. With the mammalian faunas in particular, each interglacial

stage was shown to have a characteristic fauna (Table 1.1). The suitability of mammals for this purpose stems from several factors (see Lister 1992): the rapid turnover of many mammalian lineages during the Pleistocene through origination and extinction of species, the quantifiable evolutionary trends shown in many of these lineages, and the ability of mammals to track the complex climatic and environmental fluctuations of the Pleistocene through means of migration (Schreve 2001). The specific combination of past geographical, environmental and biological factors which leads to a particular group of very diverse mammals coexisting in Britain is most unlikely to have ever been repeated (Schreve 2001). Schreve (2001) identifies three distinctive temperate-climate mammal assemblage-zones (MAZ) that are believed to correspond with three discrete climatic cycles between the Anglian glaciation and the last (Ipswichian) interglacial using such mammalian biostratigraphy (Fig. 1.1 & Table 1.1).

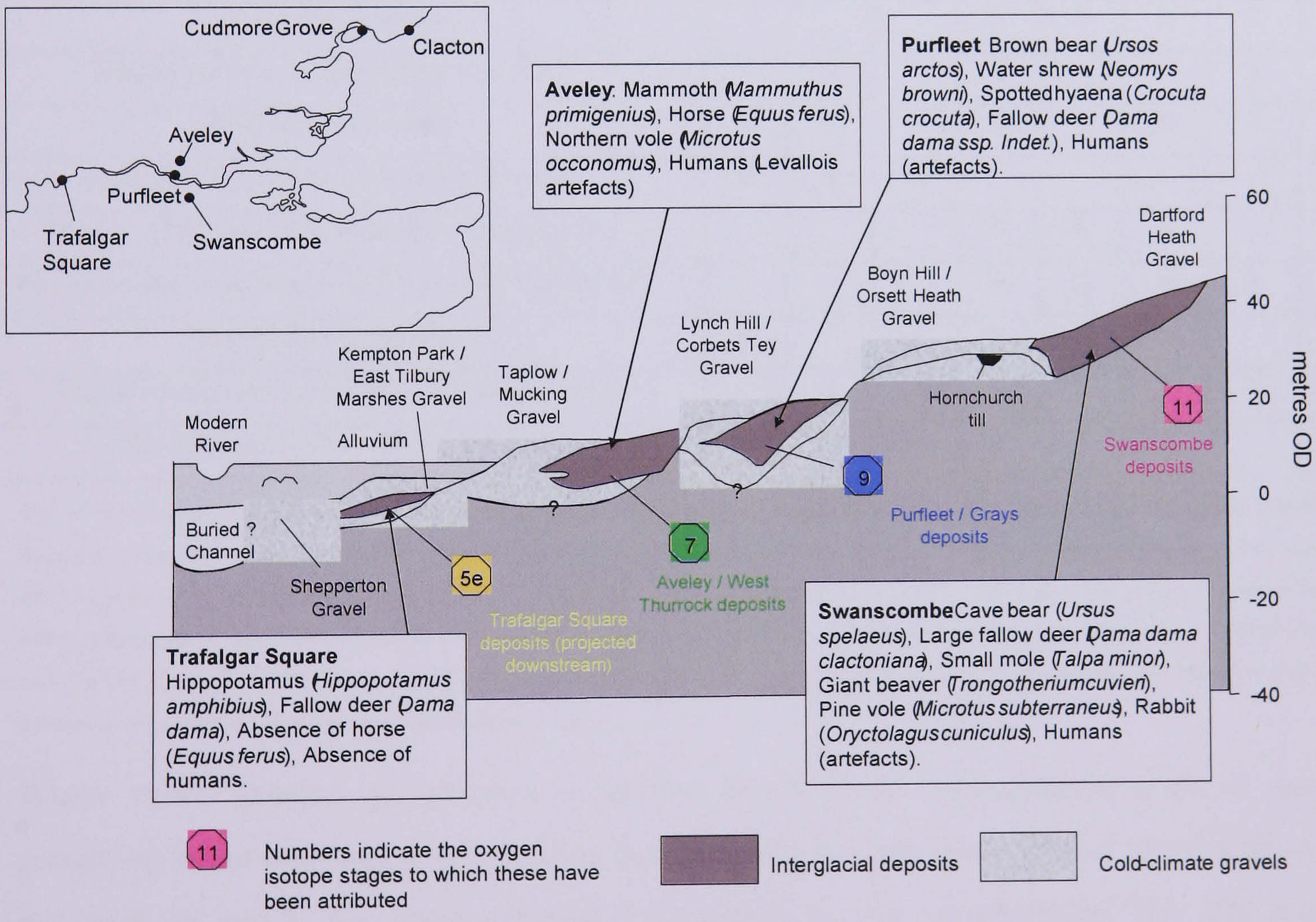


Figure 1.1 – Transverse section of the Lower Thames terraces (Modified from Bridgland et al. 1994 in Schreve 2001). Proposed correlations with the oxygen isotope record and biostratigraphically diagnostic features of the mammalian assemblages from each interglacial are shown.

Table 1.1 - Ungulate faunas from British interglacials tabulated by likely oxygen isotope stage (from Bradshaw et al. 2003)

Species	Oxygen Isotope Stage							
	17	15	13	11	9	7	5e	1
Approximate age of start (in thousands of years)	700	620	520	410	330	240	130	11.5
Giant deer *(Megaloceros verticornis)*	X	X	X					
Giant deer *(Megaloceros savini)*	X	X						
Giant deer *(Megaloceros dawkinsi)*		X	X					
'Irish' Giant deer *(Megaloceros giganteus)*				X	X	X	X	
Broad-fronted moose *(Alces latifrons)*	X		(X)					
Moose *(Alces alces)*								X
Red deer *(Cervus elaphus)*	X	X	X	X	X	X	X	X
Fallow deer *(Dama dama)*	X	X	X	X	X		X	
Roe deer *(Capreolus capreolus)*	X		X	X	X	X	X	X
Bison *(Bison priscus/Bison schoetensacki)*	X	X	X	X	X	X	X	
Aurochs *(Bos primigenius)*				X	X	X	X	X
Hippopotamus *(Hippopotamus amphibius)*		X					X	
Wild boar *(Sus scrofa)*	X	X		X	X	X	X	X
Rhinoceros *(Stephanorhinus hundscheimensis)*	X	X	X					
Merck's rhinoceros *(Stephanorhinus kirkbergensis)*				X	X	X		
Narrow-nosed rhinoceros *(Stephanorhinus hemitoechus)*				X	X	X	X	
Wild horse *(Equus ferus)*	X		X	X	X	X		
Straight-tusked elephant *(Palaeoloxodon antiquus)*		X	(X)	X	X	X	X	
Steppe mammoth *(Mammuthus trogontherii)*	X	X	(X)			X		

'OIS 17' based on the West Runton Freshwater Bed, Norfolk (Stuart, 1996) : 'OIS15' on Pakefield/Kessingland, Suffolk (Stuart and Lister 2001); 'OIS13' on Boxgrove, Sussex, and Westbury-sub-Mendip, Somerset, (Andrews *et al.* 1999; Roberts *et al.* 1999); OIS 11 on Swanscombe, Kent and Hoxne, Suffolk (Schreve 2001; Stuart *et al.* 1993; Sutcliffe 1964); OIS 9 on Purfleet, Essex (Schreve 2001); OIS 7 (later part) on Aveley, Essex and correlated deposits (Schreve 2001); OIS 5 on Barrington, Cambridgeshire, Trafalgar Square, London and correlated deposits (Stuart 1976); OIS 1 on Star Carr , 'Yorkshire', ca. 9500 BP (Fraser and King 1954; Legge and Rowley-Conwy 1988). *Palaeoloxodon* and *Mammuthus* in OIS 13 (in brackets) are not present at Boxgrove or Westbury, but abundant at correlative sites on the adjacent continent (Stuart and Lister 2001).

Where more detailed information is known about faunal assemblages, such as those pertaining to archaeological sites rather than palaeontological sites, species identification is not used as much for environmental interpretations, but incorporated into Minimum Number of Individuals (MNI) and/or Number of Individual Specimen (NISP) records. These are used to investigate animal management strategies at individual sites, as the majority of archaeozoological research questions focus on themes that benefit from accurate

identifications, such as the 'secondary products revolution' (Sherrat 1981). For example, it was argued by Sherratt (1983) that during the 4th-3rd millennia BC, stock farming in Europe was transformed by the exploitation of animals for their secondary products (milk and wool/hair) and by the development of a 'pastoral' economy. During the later Neolithic and Early Bronze Age (5th-3rd millenia BC) settlements became more spread into marginal hilly areas and caves, which would add support to the pastoral point of view. A counter view to prehistoric pastoralism is that pastoral economies are dependant on the exchange of animal products for staple grain foods and so would have required the development of 'market' institutions. Halstead (1996) argued that a reliance on livestock by Greek pastoralists is founded on keeping a large number of animals per person and scheduling pasture localities to maintain high levels of nutrition and productivity. These both require seasonal mobility and specialisation in one species. He also argues that for pastoralism, there would be a particular emphasis on specific secondary products with a high exchange value which would result with the distinct 'milk' or 'wool' mortality models presented by Payne (1973).

In order to establish whether or not an archaeological site's faunal assemblage represents a particular secondary products model, accurate species identification of all samples would be preferred, particularly with the young individuals that would be intentionally killed off in such modes of husbandry. However, two of the most interesting species for secondary products are sheep (*Ovis sp.*) and goats (*Capra sp.*), which are morphologically very similar and thus are difficult to distinguish using some of the skeletal elements. Species discrimination is further complicated by intentional fragmention (ie., butchery, bone tools) and unintentionally fragmented (ie., trampling) bones, resulting in material where morphologically diagnostic criteria are no longer present. As a result of these difficulties in identification, the use of biomolecules for making objective identifications have been considered in detail (Hyland *et al.* 1990; Lowenstein 1985; Nielsen-Marsh *et al.* 2002; Ostrom *et al.* 2000). Current research into retrieving amplifiable DNA sequences from Near Eastern archaeological sites often fail (Bar-Gal *et al.* 2003; Larson *et al.* 2007) due to the poor survival of biomolecules in warm climates. This is one example where proteinaceous species identification could be useful and should be explored further.

1.1.2 Previous Approaches to Biomolecular Species Identification in Ancient Bone

Species identification of fragmented ancient bone is thus limited when the surviving bones are undiagnostic due to fragmentation or modification. Methods to overcome these limitations using molecular information have been increasing in number over the past few decades but these are also affected to a greater or lesser extent by preservation. These methods originally focused on DNA-based approaches following the introduction of the polymerase chain reaction (PCR) to molecular biology. PCR allowed for the amplification by replication of minute amounts of DNA into much greater quantities, therefore allowing for subsequent sequencing, even in archaeological samples (Brown and Brown 1992; Jones 1985). By the very nature of PCR, contamination issues in archaeological samples have always been a major problem and so other methods of species identification have arisen, such as stable isotope analysis (Angerbjorn *et al.* 1994; Balasse and Ambrose 2005), as well as improvements in morphological criteria for species identifications (Halstead *et al.* 2002).

With the recent improvements in soft-ionisation mass spectrometry, attempts at obtaining sequence information from ancient proteins in order to identify species have been on the increase (Nielsen-Marsh 2002; Nielsen-Marsh *et al.* 2002; Nielsen-Marsh *et al.* 2005; Ostrom *et al.* 2006; Ostrom *et al.* 2000). Because the inorganic mineral fraction of bone survives much greater lengths of time than the protein-containing organic fraction, proteins that could associate with the mineral fraction were believed to be the most likely to survive (Masters 1987). These initial attempts have all focused on the highly-abundant small non-collagenous protein (NCP) osteocalcin because of its potentially long survival properties, as well as the ease of analyzing and sequencing such a small protein. Once the mineral-binding properties of OC were realised in the 1980s (Hauschka 1980; Hauschka and Wians 1989), it was soon studied in detail for survival rates in very ancient samples by immunological reactivity (Collins *et al.* 2000a; Muyzer *et al.* 1992). Several other NCPs have been searched for in ancient remains, ranging from osteonectin (Termine *et al.* 1981b), proteoglycans (Fisher *et al.* 1983), sialoproteins (Franzen and Heinegard 1985; Heinegard *et al.* 1989) and Matrix Gla Protein (MGP) (Price *et al.* 1983). There were also reports of haemoglobin and albumin in ancient calcified tissue (Cattaneo *et al.* 1990) but none has progressed onto direct

sequencing. The vastly dominant protein collagen was initially considered as a possible target; however, early immunological work by Shoshani *et al.* (1985) described that collagen evolves much slower than other proteins such as albumin, and were not able to distinguish between the collagens of the three elephantids (Asian elephant, African elephant and Mammoth). Also, the direct sequencing of the randomly degraded collagen found in fossils is thought virtually impossible due to the redundant Gly-Xaa-Yaa sequence, where Xaa is often Pro and Yaa often Hyp (hydroxyproline) (Kadler *et al.* 1996), precluding localisation of peptide fragments within the large protein (approximately 3300 residues per molecule), in other words making any small peptides sequenced hard to localise. With these complications, much of the initial work on analysing ancient proteins for phylogenetic information has been on small NCPs found within the bone matrix, the most abundant of which being OC (Nielsen-Marsh *et al.* 2002; Ostrom *et al.* 2000).

The previous archaeological applications of the sequencing of OC make use of the slight variations in amino-terminal sequences that result in a genus-specific molecular mass. Such a difference in sequence (and mass) would be ideal for archaeologists to distinguish between morphologically similar animals (such as sheep and goat bones), especially in pre- and neo-natal bones. Although ancient DNA has already been applied to such studies (Newman *et al.* 2002), it is subject to more limitations including low preservation rates as well as contamination problems. If the only information required is species identification of an archaeological specimen, preference should be given to species-specific proteins that have the potential to survive millions of years, theoretically longer than ancient DNA.

1.2 The Rise of BSE and the Need for Species Identification in MBM

After various crises in the British food industry, including the 'mad cow'/bovine spongiform encephalopathy (BSE) epidemic emerging in 1986 and resurfacing in 2000, the major concerns for people involved were no longer placed on productivity but on the protection of human and animal health. BSE is a chronic, degenerative disease that affects the central nervous system of cattle and belongs to a group of progressive degenerative neurological diseases known as transmissible spongiform encephalopathies (TSEs) that occur in a wide

variety of mammals. Symptoms of this disease include a change in body temperature, abnormal posture and difficulty in standing, lack of coordination, decreased milk production and weight loss, and are always fatal (Vermeulen *et al.* 2005). Other TSE diseases include scrapie, which occurs in sheep and goats, feline spongiform encephalopathy which occurs in cats, and chronic wasting disease found in elk and deer (Vermeulen *et al.* 2005). Creutzfeldt-Jakob disease (CJD) is a TSE that afflicts humans over 60 years old and occurs in one to two people per million worldwide per year. After the diagnosis of BSE in cattle, a new variant of CJD was observed in humans and called variant Creutzfeldt-Jakob disease (vCJD; formerly nvCJD). This disease is different from CJD as it afflicts younger people and has pathological similarities to BSE: BSE was therefore suspected to have infected human beings, with the infectious agent being a modified form of a normal cell protein known as a 'prion' (Ocana *et al.* 2004), a term for 'proteinaceous infectious particles' (Prusiner 1982). These TSEs were known to involve the abnormal accumulation of particular proteins in the nerve cells, which change structure from a soluble form into an insoluble form, making them resistant to the normal processes of degradation (White *et al.* 2000). Although apparently not the case with BSE, which is thought to be only transmitted by contaminated MBM, scrapie and chronic wasting disease are both thought to involve a soil reservoir (Schramm *et al.* 2006). It was noticed that healthy sheep bred in pastures previously frequented by scrapie-infected sheep could become infected without contact with the former flock (Palsson 1979), and that the infectious agent could remain active after three years in soil (Brown and Gajdusek 1991) and recently shown to persist for at least 16 years in the environment (Georgsson *et al.* 2006).

Early work in the 1960s suggested that the scrapie agent could replicate without nucleic acid, and mechanisms for the self-replication of proteins were described soon after (Griffth 1967). Decades later, Prusiner adapted this mechanism to propose that prions multiply by converting normal protein molecules into dangerous ones, by inducing the benign molecules to change their shape from one characterised by alpha helix motifs to one characterised by beta sheet formation. It is this change in conformation that is associated with protease resistance and the accumulation of insoluble aggregates of prion protein (White *et al.* 2000). The initial cause of the BSE epidemic is thought to have been the use of 'processed animal protein' (PAP) or MBM made from scrapie-infected sheep as animal feed. The possible transmission via feed was confirmed experimentally by feeding domestic goats with brain

tissue collected from BSE-positive cattle (Foster *et al.* 1993). Following this line of thought in 1994, the main BSE control measures specify that PAP (including MBM, blood products, hoof meal, horn meal, poultry offal meal, feather meal, fish meal, dicalcium phosphate and gelatine) cannot be fed to animals (mammalian) kept for the production of food (94/381/EC). In 1996, legislation 96/449/EC stated that mammalian by-products had to be pressure-cooked when processed into MBM for feeding such animals, however this was later repealed to completely restrict the production of MBM and tallow by legislation 1999/534/EC. This total ban was introduced due to the inadequate feed controls introduced by the first ban in 1994, as demonstrated by the 'second BSE crisis' which saw a general increase in cases of BSE at the end of the 1990s due to cross-contaminations of feeds in the production and transport of MBM (Vermeulen *et al.* 2004). To enforce and police this ban, the UK Department for Environment, Food and Rural Affairs (DEFRA) carries out feed sampling programmes to test for contamination from animal sources using conventional microscopy methods as well as methods using immunoassay or PCR (Bellagamba *et al.* 2001; Krcmar and Rencova 2003; Lahiff *et al.* 2001; Ocana *et al.* 2004; Tartaglia *et al.* 1998). The PAP regulations stipulate that there is zero tolerance for the presence of processed animal proteins in feed. However, it is generally accepted that an analytical method for detection should show sufficient sensitivity at a concentration of 0.1%, with a rate of false positives less than 5%. Animal By-Product (ABP) regulations (legislation 1774/2002; 811/2003) also require analytical methods to detect the presence of species-specific processed animal proteins (e.g. poultry, pigs). The potential methods must also be able to tolerate the extreme conditions used in the production of the processed animal by-products (20 min., 133°C, 3 bar pressure), used to sterilise and/or destroy dangerous pathogens present in the processed residues following the directive implemented (in the UK) by the ABP Order of 1992. This was because the more heat-resistant strains of scrapie were not completely inactivated by conditions less rigorous than 133°C for a minimum of 20 minutes, although all rendering processes investigated resulted in some degree of inactivation. As the heat-processes involved in rendering at most plants cannot be carried out at these temperatures precisely, the standard sample sets for analytical research include a range of temperatures above this value. One of the most important aims for the food and agricultural industry is to develop a method of identifying contaminant tissues in MBM so that the total ban on the use of MBM can be relaxed into species-specific bans of MBM in particular animal feed.

1.2.1 Previous Approaches to Species Identification in MBM; the STRATFEED Project

Subsequent to the BSE outbreak in UK cattle herds in 1986 and its associations with variant CJD in humans in 1996 (Murray *et al.* 2001), the use of bovine offals and tissues from cattle over thirty months of age have been banned (since 1996) from the UK food chain. This was later extended to cover all European Union member states from 4 December 2000 (Murray *et al.* 2001). The use of MBM in all animal feeds was temporarily banned, while fishmeal was banned from being used for ruminant feeds, but not pig and poultry diets (2000/766/EC). Hence large amounts of MBM are being produced as slaughter wastes that have no value or purpose. Due to the expected fraudulent economic adulteration within the animal feed industry, the EU instigated a three-year research program called STRATFEED (Dardenne 2000) to explore new strategies for the detection of mammalian tissues in feedingstuffs in order to help police the industry. Although several sophisticated methods exist to detect banned ruminant tissue in feeds, they are costly and time-consuming, and often defeated by thermal damage to the target biomolecule. The study presented here was to establish whether or not the most thermostable proteins present in MBM are bone proteins and the efficiencies of species identification using protein sequences in such degraded material.

The STRATFEED project was a research project supported by both the Department for Environment Food and Rural Affairs (DEFRA) and the EU commission and dedicated to the investigation of new methodologies to detect the illegal addition of mammalian tissues in feedingstuffs. The microscopic methods currently in place are limited by slow speed, a commitment to skilled staff, high costs, inaccurate quantification of illegal MBM, and has great difficulties in differentiating anything more than land-based from marine-based species. Alternative techniques that permit a rapid control system were investigated by several laboratories in conjunction with STRATFEED for species differentiation and quantification of contaminants in MBM. The current microscopic methods of species identification, as well as methods of near-infrared spectroscopy (NIRS) (Baeten *et al.* 2001; Behnam *et al.* 2002) and near-infrared microscopy (NIRM) (Baeten *et al.* 2001) are cumbersome, require a specially trained expert and struggle to differentiate between terrestrial mammals. Current biomolecular methods included DNA analysis by PCR (Gizzi *et al.* 2004), and immunological

methods, such as enzyme-linked immuno-sorbent assay (ELISA)(Chen *et al.* 2004) which can sensitively identify the species of origin in raw MBM (rendered at ~100°C) using the thermostable muscle protein troponin. However, the immunological methods are significantly less sensitive with severely heat-treated material because of the alteration of the specific epitopes (Hofmann 1973), and considered as a method of surveillance of proper rendering protocols (von Holst *et al.* 2000). Although most immunological work so far on MBM has failed on the heat-treated material (von Holst *et al.* 2001), one laboratory succeeded in the detection of ruminant and porcine proteins at 0.125% concentration in MBM heated to 130°C (Stahl *et al.* 1988). However, the temperature ranges involved in heat-treatment of MBM for this research are much higher than those previously studied, with a minimum of 133°C up to maximum of 145°C/3 bar pressure, due to the previous legislations in place for destroying dangerous pathogens like the scrapie agent and these biomolecular methods of species differentiation have not been reported to be successful at the higher temperatures. PCR-based methods of species identification appear much more successful at determining species, but have only been shown to be successful at the minimum temperature requirements (133°C-138°C/3 bar pressure for 20 minutes) (Lahiff *et al.* 2001; Tartaglia *et al.* 1998). Also, whilst some PCR-based methods achieved a sensitivity of approximately 0.125% (Tartaglia *et al.* 1998; Wang *et al.* 2000), other PCR-based methods obtained detection limits of 1-5% (Lahiff *et al.* 2001). However, because these biomolecular methods have not been shown to succeed at the higher temperatures (>138°C)(Baeten *et al.* 2004), the investigation into identifying proteins and species-specific markers in the higher-temperature rendered MBM samples by protein mass spectrometry is of great interest.

The recently emerged field of study called 'proteomics' encompasses the identification and quantification of proteins by soft-ionisation mass spectrometry following protein/peptide separation techniques. Proteomics has great potential to supplement the fast moving genomics work by characterising post-translational modifications not coded for by DNA. Hereafter the term 'protein mass spectrometry' will be used to imply proteomics-style research, but where a single protein has been targeted and isolated prior to mass spectrometric analysis. Research into protein analysis by Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI-MS) would be ideal for this species differentiation for its ability to sequence proteins in the presence of impurities. However, the first work

done using MALDI-MS for animal protein detection was by Ocana *et al.* (2004) which aimed to detect gelatine-derived peptides in animal feed. The select few peptides studied from hydrolysed samples could only indicate presence of the protein and the potential for species-identification was not considered. Some recent applications of MALDI have been done in the field of archaeological science, which targeted the bone protein OC for information such as phylogeny and preservation status of the sample (Nielsen-Marsh *et al.* 2002). Here, the potential applications of this technique for targeting the OC protein will be studied using the STRATFEED sample set, as well as other proteins found to survive the extremes of temperature and pressure involved in the rendering processes of MBM. The chosen methods will then be applied to a range of archaeological and palaeontological samples from different burial environments and ages.

1.3 Bone – The Most Durable Organic Tissue

1.3.1 The Structure of Bone

Bone not only appears to be the most common tissue found in ancient deposits but fragments of it also survive the high-temperature rendering processes of the MBM industry. Thus the structure of bone needs to be considered in order to find the most appropriate approach to biomolecular species determination and possible reasons for preferential survival over other tissues. Fossils not only provide the geologic record of evolution, but the abundance of mineralised tissues preserved is a reminder of the crucial role that they play in the biology of organisms. The evolution of exoskeletons (shells, scales, etc.) some 500-600 million years ago during the "Cambrian explosion" enhanced the preservation of this event in the form of fossils in the rock record (Boskey 2007). The subsequent development of endoskeletons (bones and teeth) gave vertebrates improved mobility and mechanical competence. Bone, one of the most abundant mineralised tissues in the fossil record, is a specialised form of dense connective tissue in vertebrates. It has the mechanical function of giving the skeleton the necessary rigidity to be an attachment and lever for muscles, supports the body against gravity, protects the internal organs and is a ready source of the key regulatory inorganic ions calcium, magnesium and phosphate. Bone also contain cells and growth factors that, in turn, control tissue properties (Boskey 2007). The two main forms of

bone tissue are dermal bone, which forms near the surface (skin) cells of the animal (such as those making up the mammalian skull), and endochondral bone, which replaces a cartilaginous framework (such as long bones in mammals) (Lyman 1994). To fulfil its functions, these can be made up of two distinct types of bone; compact (or cortical) and cancellous (or trabecular). Cortical bone is hard and compact in structure and found in the shafts of long bones surrounding the marrow cavity. Cancellous bone is spongy, consisting of fine interlacing partitions (trabeculae), which contain marrow. It has a large surface area to volume ratio to allow for participation in metabolism, especially in the regulation of calcium homeostasis, for these reasons it is found in vertebrae, most flat bones and the ends of long bones. At a lower level, bone is considered to be organised into two fairly distinct forms: either lamellar bone or woven bone. A mixture of both known as fibrolamellar bone is also observed. Woven bone is laid down rapidly and with variable collagen orientation (Boyde and Jones 1998) and quite porous at the micron level (Currey, 2006). Lamellar bone is laid down much more slowly and the collagen fibrils are more precisely arranged with their associated mineral layed down as sheets (lamellae). Lamellar bone is considered less mineralised than woven bone (Currey, 2006). Bone is a natural composite material, consisting of about 70% inorganic material, predominantly the calcium phosphate hydroxyapatite (HAP), and an organic (mostly protein) component that is predominantly the structural protein collagen. The inorganic phase gives the tissue resistance to compression forces whereas the organic phase gives resistance to tension forces (Currey 1984). Bone cells control the initial production of the mineralised tissue. Cells called osteoblasts control the mineralization of the extracellular collagen protein matrix. When these osteoblasts become engulfed in mineral, they become a different type of cell, called osteocytes, which communicate with each other via interconnecting channels (canaliculae) throughout the tissue (Boskey 2007). Ultimately, osteoclast cells remove bone mineral and bone matrix. In this way, bone cells regulate the formation and turnover or resorption of bone, a key step in regulating body calcium, magnesium and phosphate levels.

The inorganic bone mineral component is composed of a carbonate hydroxyapatite, similar to dahllite, with an approximate stoichiometry $Ca_5(OH)\{[PO_4]_{0.5-0.9}[(CO_3)(OH)]_{0.1-0.5}\}_3$ although the unit cell formula is often generalised as being $Ca_{10}(PO_4)_6.(OH)_2$. It can accommodate a large number of trace elements by surface and lattice substitutions, whereby

the Ca^{2+} ions are sometimes replaced by Na^{+}, K^{+}, Mg^{2+}, Sr^{2+} or Pb^{2+}, OH^{-} groups by CO_3^{2-}, HPO_4^{2-}, Cl^{-} or F^{-}, and the PO_4^{3-} groups can also be replaced by CO_3^{2-}groups (Glimcher and Krane 1968; Millard 2001; Spadaro *et al.* 1970). This mineral is not the same as any geologically-formed apatite, so it is termed 'bioapatite'. The mineral phase has been demonstrated to consist of small tabular or plate-shaped crystals (Landis 1996; Lees 1979; Weiner and Price 1986) with dimensions typically ranging from 2-5 x 20-50 x 12-20 nm, although this varies with age and species (Glimcher 2006; Millard 2001). The small size of these crystallites gives bone mineral a very large surface area of 85-170 m^2g^{-1} (Lowenstam and Weiner 1989), and makes it chemically very reactive. In bone, these apatite crystals develop with their long *c*-axes parallel to the collagen fibril. The collagen, along with associated proteins, plays an important part in determining nucleation, growth and proliferation of these crystals. Initially, the mineral crystals are formed in an environment rich in the Small Integrin-Binding Ligand N-linked Glycoprotein (SIBLING) proteins, and as bone crystals grow with age, association with proteins that regulate remodelling, such as osteocalcin, increases (Boskey 2007). This mineral phase contributes between 60-80% (by weight) of the dry weight of bone (Glimcher and Krane 1968).

The organic protein component provides flexibility and forms the matrix upon and within which mineral crystals are grown. In bone the protein phase accounts for 25-30% (by weight), of which collagen predominates accounting for about 90% (by weight) of the constituents in the organic matrix (Millard 2001). Collagens are the most abundant structural protein in the animal kingdom and of the more than 27 types of collagen, the fibrous collagen type I (hereafter written as 'collagen (I)') is prevalent, particularly so in bone. Fibrous proteins have long polypeptide chains that are arranged into long strands or that impart strength and/or flexibility to the structures in which they are found. Fibrous proteins are notably insoluble in water due to the large number of hydrophobic residues in their primary structure. As it is fibrous in nature, collagen is a very good candidate for preservation in the burial environment. Fibrous collagens are comprised of three polypeptide chains wound into a triple helix. To distinguish one collagen type from another, vertebrate collagens were assigned Roman numerals in order of discovery (I, II, III, etc). To distinguish between collagens' chains, individual chains were called α1, α2, α3, etc., followed by the collagen type in which they occur. For example, collagen (I), which consists of two identical

chains and one dissimilar chain, all of which are unique to collagen (I), can be written [$\alpha1(I)_2$ $\alpha2(I)$] (Kadler 1994). Each alpha chain coils in a left-handed alpha helix, whereas the combined three chains supercoil in a right-handed manner to form tropocollagen triple helix molecules. These tropocollagen molecules combine and line up head-to-tail to form fibrils, which also combine to form repeating arrays of fibers that give flexibility to the nonmineralised tissues. When reinforced with mineral particles, the resulting composite increases in strength and becomes capable of bearing weight. Spaces between the molecules, and between the fibrils, can accommodate the mineral particles (Hodge 1989; Weiner and Traub 1986). The apatite crystals appear to deposit first within the holes and then spread throughout the matrix (Fig. 1.2)(Lees 1979).

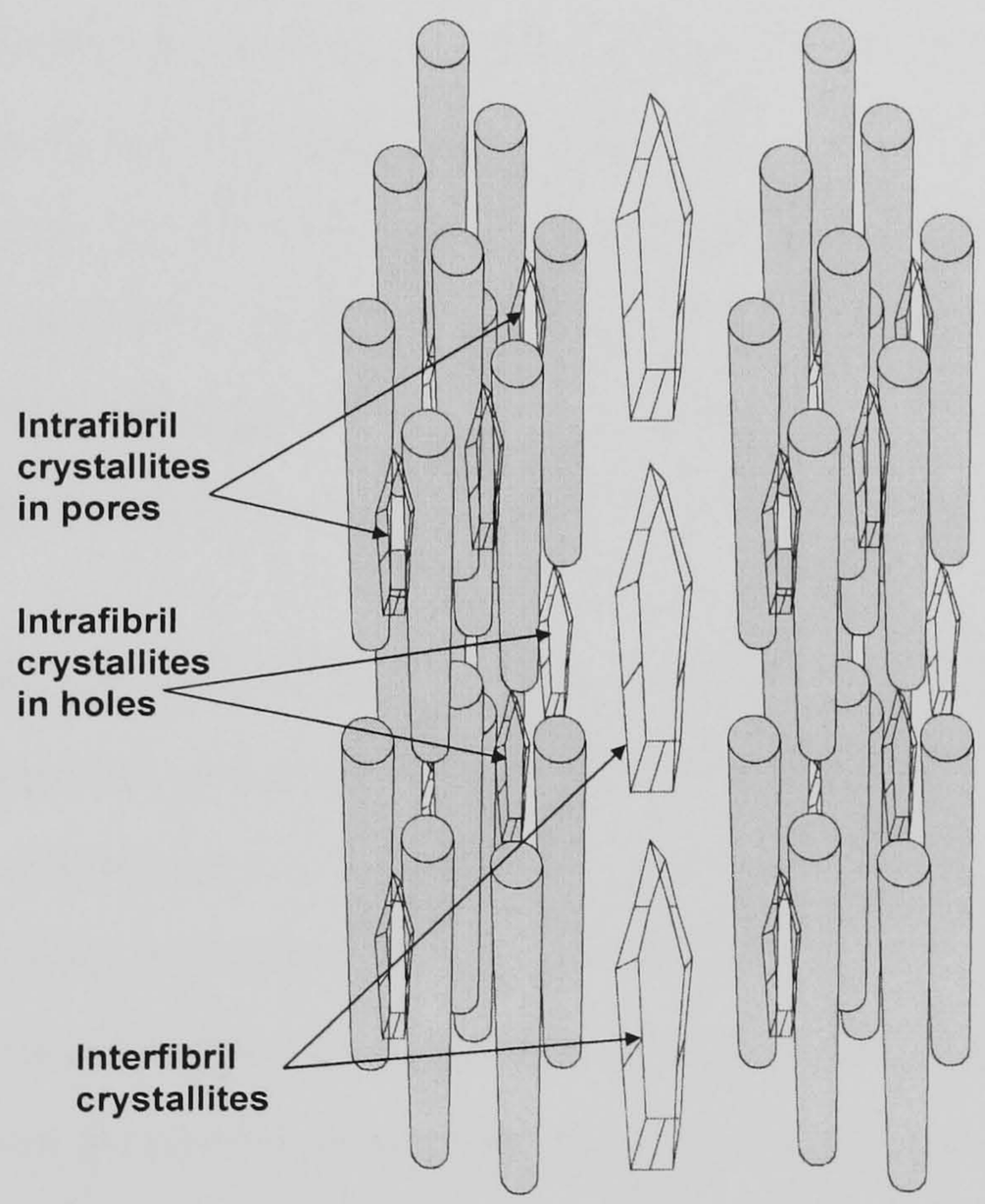


Figure 1.2 - Distribution of apatitic crystallites in collagen matrix showing orientation of the c-axis (Modified from Lees 1979).

The mineralisation process of bone is mediated by collagen and molecules which are bound to collagen (Lowenstam and Weiner 1989). Associated amongst the complicated networks of collagen molecules are many other proteins loosely termed 'non-collagenous proteins' (NCPs). If the mineral is removed from the collagen matrix with solvents that also extract the NCPs, the matrix cannot be remineralised (Termine *et al.* 1981a). Many NCPs are phosphorylated proteins, suggesting a role for the protein-linked phosphate-ester groups in

the mineralisation process. Several families of proteins associated with the collagen matrix are involved in regulation of the mineralised process, although recent research into the characterisation of bone matrix has emphasized its complexity, with Schreiweis *et al.* (2007) identifying 133 proteins, and Jiang *et al.* (2007) showing as many as 2479 unique proteins associated with bone. Some of these proteins have multiple functions beyond their role in mineralisation (Zhu *et al.* 2007). The proteins include phosphorylated proteins, proteoglycans, glycoproteins, and gamma-carboxy-glutamic acid-containing (Gla) proteins. Among the phosphorylated proteins, the SIBLINGs are the most widely studied (Qin *et al.* 2004); all have cell-binding domains and all interact with fibrillar collagen. Some of these proteins act as both inhibitors and promoters of mineralisation (Boskey 2007), depending on the extent of post-translational modification and/or concentration. Similarly, small leucine-rich proteoglycans (SLRPs) interact with fibrillar collagen. The Gla-protein family has fewer members, but one, the small mineral-binding protein OC, is the most abundant of all NCPs found in bone, making up approximately 20% by weight of the total NCPs. Although collagen and OC are present in equimolar amounts (Prigodich and Vesely 1997), the large difference in abundance by weight is due to the difference in molecular mass – collagen molecules (tropocollagen) being approximately 280 kDa and OC molecules around 5.6 kDa.

Even with the completion of the human genome (IHGSH 2004) as well as a growing number of other mammalian genomes, the precise functions of many of these NCPs are still uncertain, despite several attempts of knock-out experiments (for example see Ducy *et al.* 1996). More recently, Jiang *et al.* (2007) confirmed the presence of various types of protein in bone. Not only were many bone-specific proteins identified, but this study also showed that many of the NCPs were present in minute amounts: the bone morphogenetic protein (BMP) content in bone tissue was ~ 2 ng/kg wet bone tissue (Jiang *et al.* 2007). Among the bone-specific proteins, many proteins embedded in the bone matrix were identified, including OC, osteonectin (SPARC), bone sialoprotein, fibronectin, MGP, BMPs, growth factors, cytokines, and proteoglycans (like perlecan and biglycan). Other NCPs (not specific to bone), such as osteopontin were also identified. The identified proteins exhibited a broad spectrum of functions, including the control of cell proliferation, cell-matrix interactions, and mediation of hydroxyapatite deposition. Several serum-derived proteins including serum albumin, haemoglobin, myoglobin, alpha-2-Heremans Schmid glycoprotein (A2HS), are

reported to bind to the mineral component (Triffitt *et al.* 1976; Weiner *et al.* 1976). Two identified proteins, biglycan and creatine kinase, are involved in bone growth and differentiations (Nogami *et al.* 1987; Wallace *et al.* 2006). Furthermore, many proteins associated with bone matrix degradation, such as cathepsin, matrix metalloproteinases (MMPs), and plasminogen were also observed (Jiang *et al.* 2007). The lipid component of bone makes up only about 0.1% (wt) of the tissue, three-quarters of this is triglyceride (triacylglycerol), with the rest being predominantly cholesterol (Williams and Elliot 1989).

1.3.2 Bone Preservation

The preservation of bone into the fossil record is influenced by a number of factors, primarily biological, environmental and, in the case of archaeological assemblages, cultural factors. Fossils form as a result of the accidental burial of a dead plant or creature in sediment, such as sand, grit, clay or mud. If the carcass is not buried, the normal processes of decay or scavenging by other organisms take place. Land-living creatures are thus far more poorly represented than sea-living creatures in the fossil record because the animal's carcass would need to be rapidly covered in sediment, either following the rapid disposition into a river and then down to such a lake or sea, or get buried in collapsing cave sediments. However, the vast majority would rot or get scavenged before burial could occur (Behrensmeyer *et al.* 2000; Norman 1994; Shipman 1981).

Taphonomy is the science of the laws of burial, as coined by Efremov (1940), and thus is a branch of both palaeontology and archaeology which studies the details in the transition of organics from the biosphere into the lithosphere (Lyman 1994). Taphonomic processes directly influence the preservation or destruction of buried tissues (principally bone), and whether or not they get buried at all (Lyman 1994; Shipman 1981). The major processes recognised by palaeontologists include digestion, trampling, burning and exposure (Brain 1981; Lyman 1994). Many of these processes also cause physical changes, such as those which are characterised in the weathering stages of Behrensmeyer (1978). As well as the taphonomic processes described above, bones found in many archaeological sites are often processed by humans or other animals, such as for marrow extraction or the creation of bone tools (Brain 1978). Such processes can leave the surviving bone fragments undiagnostic

of species. Following initial burial, further sediment deposition entombs the remains of the organism which becomes protected from further disturbances. The soft tissues continue to rot and eventually disappear and the increasing weight of the sediment gradually crushes the body flat, with the exception of the more robust skeletal parts, such as shells, teeth and some of the bones (Behrensmeyer *et al.* 2000; Norman 1994). Eventually, the particles of sediment surrounding the remains also become compressed by the increasing weight above and compacted into sedimentary rock. Once fully entombed within rock, the skeleton becomes a true fossil. Mineralised fossils are composite materials that potentially contain the original chemicals which formed the skeleton of the animal, but the minute spaces which were originally filled with collagen or soft tissues may become filled with another mineral, carried into the hard part of the skeleton by water. Berna *et al.* (2004) note that bone mineral is in fact much more soluble than synthetic hydroxyl apatite and observed that between a pH 7.6 and 8.1 (at 'steady state' conditions) it tends to dissolve and reprecipitate as a more insoluble form of carbonated hydroxyl apatite. This will undoubtedly affect the preservation potential of remaining biomoleucules. Stuart (1982) notes that with the exception of Lower Pleistocene crag material, which is heavily infilled with iron oxides, and bones from cave sites that contain calcium carbonate, most British Pleistocene vertebrate material is only slightly mineralised/unaltered and contains some collagen.

Although there are many factors that effect the preservation of bone, the most influential factors are likely the activity of saprophytic organisms, which is in turn influenced by environmental conditions. Saprophytic organisms, such as earthworms, insects, fungi, bacteria, arthropods and microbes, obtain their nutrients (nitrogen, potassium, phosphates) by breaking down the organic matter of deceased organisms. Carbone and Keel (1985) noted four environmental factors that can influence their action; soil acidity, aeration, relative humidity, and temperature. Because most of these saprophytic organisms live in alkaline soils and are intolerant of highly acidic soils, organic components of tissues, such as bone collagen, tends to preserve better in acidic conditions, but chemical decomposition of the mineral phase tends to occur, the extent of which would depend on the ion content, ie., an abundance of phosphate ions in solution will affect the solubility of the hydroxyapatite (O'Connor 2000). Conversely, alkaline conditions tend to preserve the mineral component of bone, but the organic components remain subject to attack from various types of

saprophytic organisms, the extent to which is effected by exclusion by the mineral phase (Gordon and Buikstra 1981; Grupe 1995). Skeletal element size and mineral density, the larger of which can endure more damage from saprophytic organisms and from chemical decomposition, and the burning of bones, which removes protein and alters the mineral components, can also affect the preservation state of particular bones. The bones of large animals, and particular large skeletal elements (femur, tibia, humerus, calcaneus, astragalus) are thus preferentially recovered from ancient sites (Sobolik 2003). Other environmental factors may include the physical destruction through seasonal freeze/thaw cycles, preservation in dry/arid or dry/frozen regions, or anaerobic environments (waterlogged, bogs, under clay, or silt) that decrease saprophytic activity. However, observations reported by Nicholson (1996) could not separate the diverse range of several of the factors mentioned above, highlighting the difficulties in recording the individual effects of each on bone (O'Connor 2000).

As bone tissue is demonstrated to be the most commonly found organic tissue in palaeontological and archaeological deposits, as well as being a major constituent of MBM, it is the focus material of this thesis.

1.3.3 Ancient Biomolecules

Increased interest in studying biomolecules has followed the incredible recent advancements in technology that allow biomolecular analysis in much greater detail, at lower concentrations and at much faster rates than ever before. As the amount of sample material required for molecular work has decreased dramatically with improved technology, archaeologists and palaeontologists are more likely to part with 'precious' samples in order to gain useful information, be it the age of sample (Agarwal and Sohal 1994; Aiello and Molleson 1993; Ajie *et al.* 1991; Bada 1985; Csapo *et al.* 1998), kinship analyses (Brown and Brown 1994; Gerstenberger *et al.* 1999; Hummel and Herrmann 1996) or phylogenetic analyses (Burger *et al.* 2004; Greenwood *et al.* 2001; Helgen 2003; Orlando *et al.* 2003). As the greatest hindrance to the biomolecular archaeologist is the non-uniform preservation of these ancient biomolecules, bone diagenesis and biomolecular degradation needs to be understood in greater detail.

In brief, an ancient biomolecule is an organic molecule from a sample of ancient biological tissue. Within biological tissues there are four main types of biomolecule; carbohydrates, lipids, proteins and DNA. Two of the most important factors regarding the analysis of these ancient biomolecules are the preservation potential of the molecule and the information content preserved. Carbohydrates, such as sugars and starches, have high preservation potential but very low information content. Proteins (collagen, OC and haemoglobin) and lipids (waxes, fats and oils) have a moderate preservation potential and some contain useful information, sometimes as much as species specificity (to genus) (Nielsen-Marsh *et al.* 2002). The biomolecule with the greatest information content is DNA, but it has a very low potential for preservation (Curry 1988). Although the biomolecules may be listed in rank order of decay resistance (e.g., lipid > carbohydrate > protein > nucleic acid), and by implication, preservation potential, this is an oversimplification as the preservation potential of a molecule may be increased if it becomes incorporated into structural tissues (Briggs 1993).

It has been shown that the fossilisation potential of organic molecules is extremely variable, as some compounds common in living organisms are extremely rare in the fossil record and only preserved under exceptional conditions (Curry 1988). For example, despite its abundance in living tissues, DNA is rarely preserved in recognisable form in fossils, partly due to it being concentrated in vulnerable soft-tissues, as well as the molecule itself being highly reactive and easily broken down. However, the recognisable remains of molecules such as proteins and lipids are extremely abundant in the fossil record (Wyckoff and Davidson 1973). This observation is believed to be due to the constituent nucleic acids of DNA being much less stable than the amino acids which make up proteins. Initial work by Abelson (1956) showed that amino acids survive in ancient fossils such as Devonian (416-359 Ma) fish, Jurassic (199-145 Ma) and Cretaceous (145-65 Ma) dinosaurs and Oligocene (33.9-23 Ma) to Pliocene (5.3-1.8 Ma) horses. The preservation of biomolecules, e.g., collagen and NCPs, is also enhanced by protection within biominerals, e.g., in bones, shells, and teeth.

When considering the type of ancient biomolecule to study for any particular research, the preservation potential should be considered with respect to the information content required. The focus of the research presented here is the matter of species-identification, a frequent problem faced by archaeozoologists with fragmented or morphologically similar material. In the case of species-identification, DNA would be the ideal biomolecule to target as it can offer highly defined phylogenetic analyses as well as the unambiguous identification of species. One decade ago, the hope that original DNA and protein would be widely persistent in geological samples fuelled biomolecular studies of fossil material. Ancient DNA led the way with spectacular claims of sequences being retrieved from tissues millions of years old, including Cretaceous dinosaur bones (Woodward *et al.* 1994). However, one by one, the claims for exceptional preservation were questioned and the field of aDNA research became focused on analyses of more recent material (see Willerslev and Cooper 2005 for review). However, it can rarely be extracted and amplified from fossil bones that have not been preserved under a combination of unusual conditions (Geigl 2002). Based on experiments of purified DNA in aqueous solution at physiological pH (Lindahl 1993), it was extrapolated that DNA could not 'survive' longer than 100,000 years, although this will greatly depend upon the fragment size being amplified. This timescale is greatly reduced when the samples are from warm climates, and as one of the most important topics for archaeozoology is the onset of agriculture and its spread from the Near East approximately 8-10 ka, relevant samples rarely yield DNA for species identification. The next most informative biomolecule to consider is thus proteins, and models put forward by Collins *et al.* (1998) hypothesise that the proteins collagen and OC should survive much longer than DNA (Table 1.2).

Table 1.2 - Concentrations and survival rates of osteocalcin, collagen and DNA in bone (Nielsen-Marsh 2002).

Biomolecule	**Concentration in bone (by dry weight)**	**Method**	**E_a (kJmol-1)**	**Detection limit (years x 10^3 BP)**		
				0°C	**10 °C**	**20 °C**
DNA	0.001%	Limit of amplification using E_a for DNA depurination	127	125	17.5	2.5
Collagen	22%	Laboratory measured rates of gelatinisation (M. Collins unpublished work)	173	2700	180	15
Osteocalcin	0.2%	Laboratory measured rate loss of mid-region epitope (M. Collins unpublished work)	175	110000	7500	580

Like DNA, proteins contain fundamental genetic information that is the key to phylogenetic reconstruction, but they exist in larger quantities and have much greater potential for preservation (Curry 1988; Logan *et al.* 1991; Robbins *et al.* 1993). Proteins have been shown to escape chemical or microbial attack and thus complete degradation by hydrolysis when physically encapsulated in hydrophobic networks of humic acids or other refractory organic matter, such as in compost (Zang *et al.* 2000). Bada *et al.* (1999) report proteins to hydrolyze within 100,000 to 1 million years, a time period that goes beyond archaeological time scales and into geological time scales.

The main advantage of protein sequencing to palaeontology over aDNA sequencing is that endogenous aDNA is less likely to be found, especially with the contamination issues involved with spectacular fossil material that has been handled by many individuals over time since it was excavated, sometimes hundreds of years ago. However, the limitations of protein sequencing for species identification will be the resolution of the sequence data which will be explored further within this thesis. From protein sequence studies, relationships can be inferred and the numbers of effective amino acid substitutions can be related to base changes in the genetic code of the DNA molecule. Because bone is considered one of the most durable organic tissues, being commonly found in palaeontological sites, it is the tissue investigated throughout this thesis.

1.4 Bone Proteins and Amino Acid Racemisation

1.4.1 Amino Acids

There are 20 different amino acids found in proteins, where each consists of an amino group ($-NH_2$), a carboxyl group (-COOH), a hydrogen atom (H) and a side chain (R), that are attached to a central carbon atom known as the alpha carbon (Table 1.3)(Fig. 1.3). In proline the $-NH_2$ group is incorporated into a five-membered ring but still has the same structure as described above. Some of these R groups are small and relatively simple, such as a single hydrogen atom (glycine) or methyl group (alanine). Others are large complex aromatic side chains (e.g., phenylalanine, tryptophan and tyrosine). At neutral pH, amino acids take the ionic form of zwitterions, where most have no net charge with a positive charge on the

amino group (NH_3^+) and a negative charge on the carboxy group (COO^-). When the pH is high, the carboxyl group is anionic and the amino group is not charged (NH_2), resulting in a net negative charge; when the pH is low, both the carboxyl group and the amino groups are protonated, resulting in a net positive charge. However, at neutral pH there are two that are negatively charged (aspartic acid and glutamic acid) and three which are positively charged (arginine, lysine and histidine). Some are polar (e.g. serine and threonine), others less polar (e.g. alanine, leucine and valine). With the exception of the simplest amino acid, glycine (R = H), all the amino acids found in proteins contain at least one asymmetric carbon atom and hence are optically active. The vast majority of proteins synthesised *in vivo* have been found to have the L configuration (Wilson and Walker 2005).

*Table 1.3 -Symbols and side chains of the protein amino acids (from Wilson and Walker 2005). * Cyclic structure of proline side chain incorporates a-carbon*

Name	**Symbol**	**Residue mass**	**Side-chain**
Alanine	A, Ala	71.079	CH_3
Arginine	R, Arg	156.188	$HN=C(NH_2)-NH-(CH_2)_3$
Asparagine	N, Asn	114.104	$H_2N-CO-CH_2$
Aspartic acid	D, Asp	115.089	$HOOC-CH_2$
Cysteine	C, Cys	103.145	$HS-CH_2$
Glutamine	Q, Gln	128.131	$H_2N-CO-(CH_2)_2$
Glutamic acid	E, Glu	129.116	$HOOC-(CH_2)_2$
Glycine	G, Gly	57.052	H
Histidine	H, His	137.141	$Imidazole-CH_2$
Isoleucine	I, Ile	113.160	$CH_3-CH_2-CH(CH_3)$
Leucine	L, Leu	113.160	$(CH_3)_2-CH-CH_2$
Lysine	K, Lys	128.17	$H_2N-(CH_2)_4$
Methionine	M, Met	131.199	$CH_3-S-(CH_2)_2$
Phenylalanine	F, Phe	147.177	$Phenyl-CH_2$
Proline	P, Pro	97.117	$*C_3H_6$
Serine	S, Ser	87.078	$HO-CH_2$
Threonine	T, Thr	101.105	$CH_3-CH(OH)$
Tryptophan	W, Trp	186.213	$Indole-NH-CH=C-CH_2$
Tyrosine	Y, Tyr	163.176	$4-OH-Phenyl-CH_2$
Valine	V, Val	99.133	$CH_3-CH(CH_2)$

1.4.2 Protein Structure

Like other biological macromolecules such as polysaccharaides and nucleic acids, proteins are essential in all living organisms. The functions of proteins vary, some being enzymes that

catalyze biochemical reactions, whereas others are structural proteins that form a system of scaffolding, such as cytoskeletal proteins and some extracellular collagens. Protein is also a necessary component of diet in order to obtain the essential amino acids that cannot be synthesised by the organism. Proteins are large polymers of amino acids joined by peptide bonds, amide linkages joining the amino group of one amino acid to the carboxyl group of another (Fig. 1.3). When several amino acids are joined together by peptide bonds, the result is a peptide, and longer chains make polypeptides. Proteins consist of one or more polypeptide chains held together, usually by non-covalent interactions. All polypeptides have an amino-terminus (where the amino group of the terminal amino acid is unlinked) and a carboxy-terminus (where the carboxyl group of the terminal amino acid is unlinked).

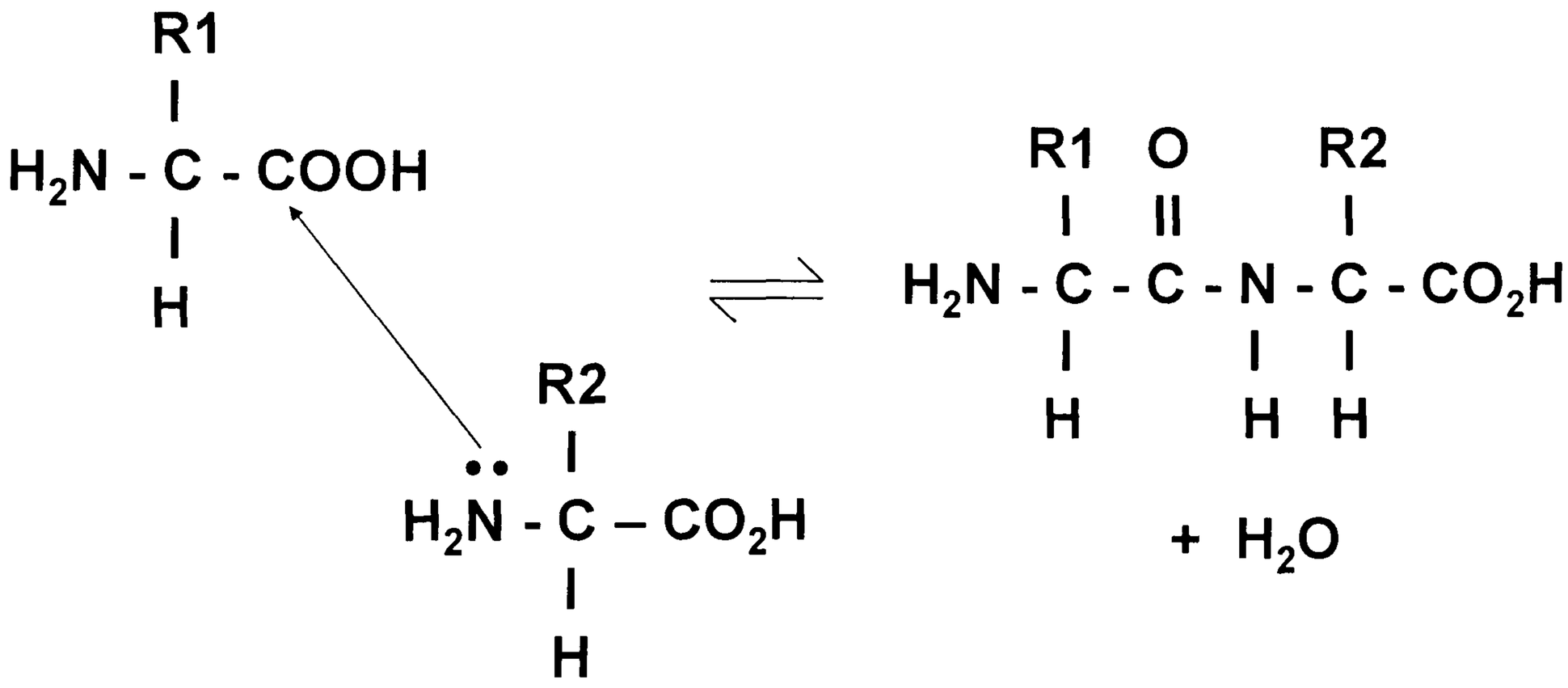


Figure 1.3 – The linkage of two amino acids into a dipeptide chain by condensation (Penkman 2005)

The sequence of amino acids in a polypeptide, known as its primary structure, governs the structural basis of a protein's function. The planar groups of atoms involved in two adjacent peptide bonds can rotate relative to one another about the C_α – N and C_α – C bonds. Steric features of the side chains limit rotation to a few favourable configurations that result in particular secondary and tertiary structures (Wood *et al.* 1981). In this way the primary structure of a protein determines its three-dimensional conformation and resulting functional specificity. Different sequences of amino acids result in different combinations of chemical reactivities, dictating the overall structure and surface positions of chemically reactive groups and thus determining the chemical properties of the protein (Brown 2002).

1.4.3 Amino Acid Composition Analysis

The determination of which of the 20 possible amino acids are present in a particular protein, and in what relative amounts, is achieved by completely hydrolysing the protein followed by separating and quantifying them chromatographically. Hydrolysis is achieved by heating the protein with 6 M hydrochloric acid (HCl) for at least 14 h at 110°C. However, this hydrolysis procedure modifies the asparagines and glutamines into their corresponding acids (Asp and Glu) (which are quantified with them respectively) and completely destroys tryptophan (Wilson and Walker 2005). Amino acids must be derivatised for detection at analytical levels, which is possible to do either before or after separation on the column. In more recent methods of amino acid analysis, the amino acids in the hydrolysate are quantified by pre-column derivatisation with an appropriate reagent that, as well as being hydrophobic and thus allowing for separation by reversed-phase HPLC, reacts with the amino groups of amino acids to produce a coloured or fluorescent product (see Kaufman and Manley 1998). The effluent passes through the column and then an appropriate detector (colorimeter or fluorometer), and the amount of each amino acid is recorded as a separate peak and the area under the peak being proportional to the amount of that amino acid present in the sample analysed (Wilson and Walker 2005).

Much work has been done on hydrolysing bone proteins, and Hare (1980) demonstrates that the relative amino acid composition of NCPs after separation from collagen, contains no hydroxyproline or hydroxylysine, relatively much less glycine and proline, and more glutamic and aspartic acid residues than collagen. The total bone amino acid composition includes this NCP fraction, but because collagen is so abundant, the overall bone amino acid composition resembles that of collagen. Unlike modern bone, fossil bone shows extreme variations in amino acid composition ranging from collagen-like to non-collagen amino acid patterns. A nitrogen content of about 4% for modern bone corresponds to approximately 2500 nanomoles of amino acids per milligram of bone, and as long as the amino acid content of bone remains above about 250 nanomoles (~0.4% nitrogen content), the relative amino acid pattern resembles that of collagen, with high amounts of glycine and substantial amounts of proline and hydroxyproline (Hare 1980). Below about 50 nanomoles of amino acids per milligram of fossil bone, the relative amino acid pattern looks very different from that of

collagen. Some consider this to be evidence of contamination (Wyckoff 1972), others believe it to be indigenous NCPs considered highly resistant, previously masked by high concentrations of collagen (Hare 1980).

1.4.4 Amino Acid Racemisation

Most amino acids (except glycine) exist in one of two possible isomeric forms designated the D- and L-amino acid. These amino acids can undergo structural change via a process called amino acid racemisation (AAR). The L- form is exclusively used for protein biosynthesis, but after death of the organism it undergoes racemisation to the D- form until the two isomers are present in equal amounts. An equal mixture of the D and L isomers show no optical activity (rotation of polarised light) and is referred to as a racemic mixture. The racemisation of both free and peptide-bound amino acids is believed to occur via a carbanion intermediate (Fig. 1.4) (Bada and Schroeder 1975; Neuberger 1948), where the abstraction of the proton by a hydroxide ion (or other base) leads to the formation of a planar carbanion. Readdition of the proton occurs by the reaction of water with the carbanion. The rate of free amino acid racemisation is largely determined by properties of its side-chain; those with a greater electron-withdrawing capacity have more stable carbanion intermediates and thus faster racemisation rates (Bada and Shou 1980), larger side chains can retard proton removal and thus slow racemisation (Smith and Evans 1980). Smith *et al.* (1978) determined the order of racemisation for free amino acids at elevated temperatures to be Ser>Thr>Asp>Phe>Ala>Glu>Leu>Ile>Val. When peptide-bound, the racemisation rates are not only 2-3 times faster (at pH 7.6, 20°C) (Bada and Schroeder 1975), but have different relative rates of racemisation, influenced by neighbouring amino acids. When peptide-bound, the racemisation of Asp is much higher than those of the other amino acids due to its ability to racemise via the formation of a cyclic succinimide (Fig. 1.5) (Geiger and Clarke 1987; Radkiewicz *et al.* 1996).

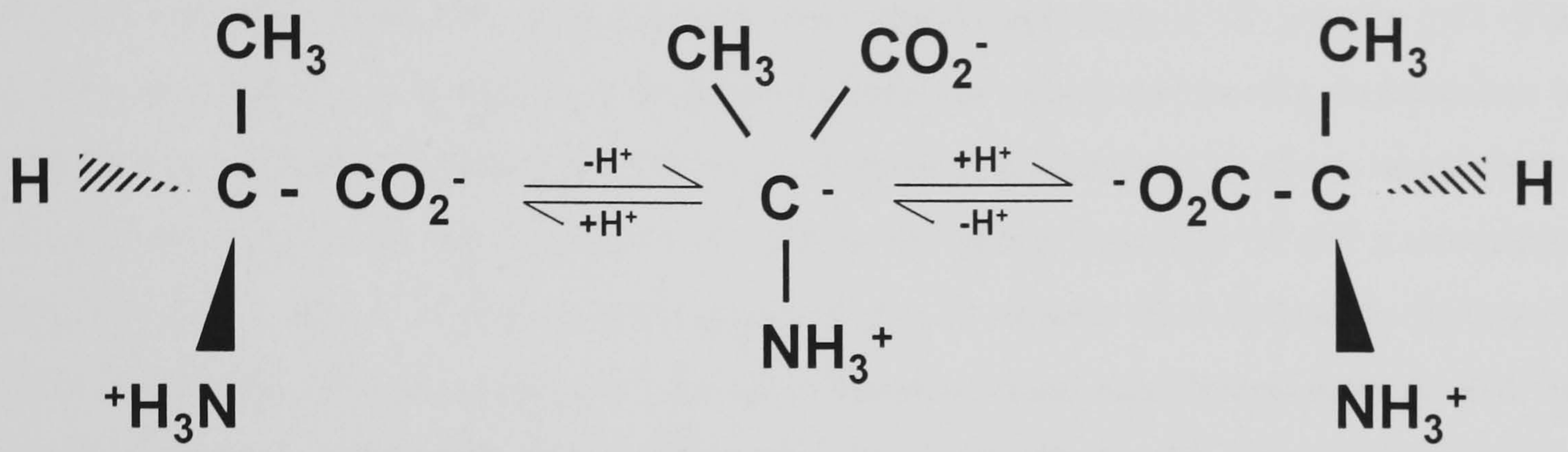


Figure 1.4 - Racemisation mechanism of free amino acids in solution (Neuberger 1948 in Penkman 2005). Amino acids exist as one of two possible optical isomers, the L-enantiomer and the D- enantiomer, named by the direction by which they rotate plane-polarised light.

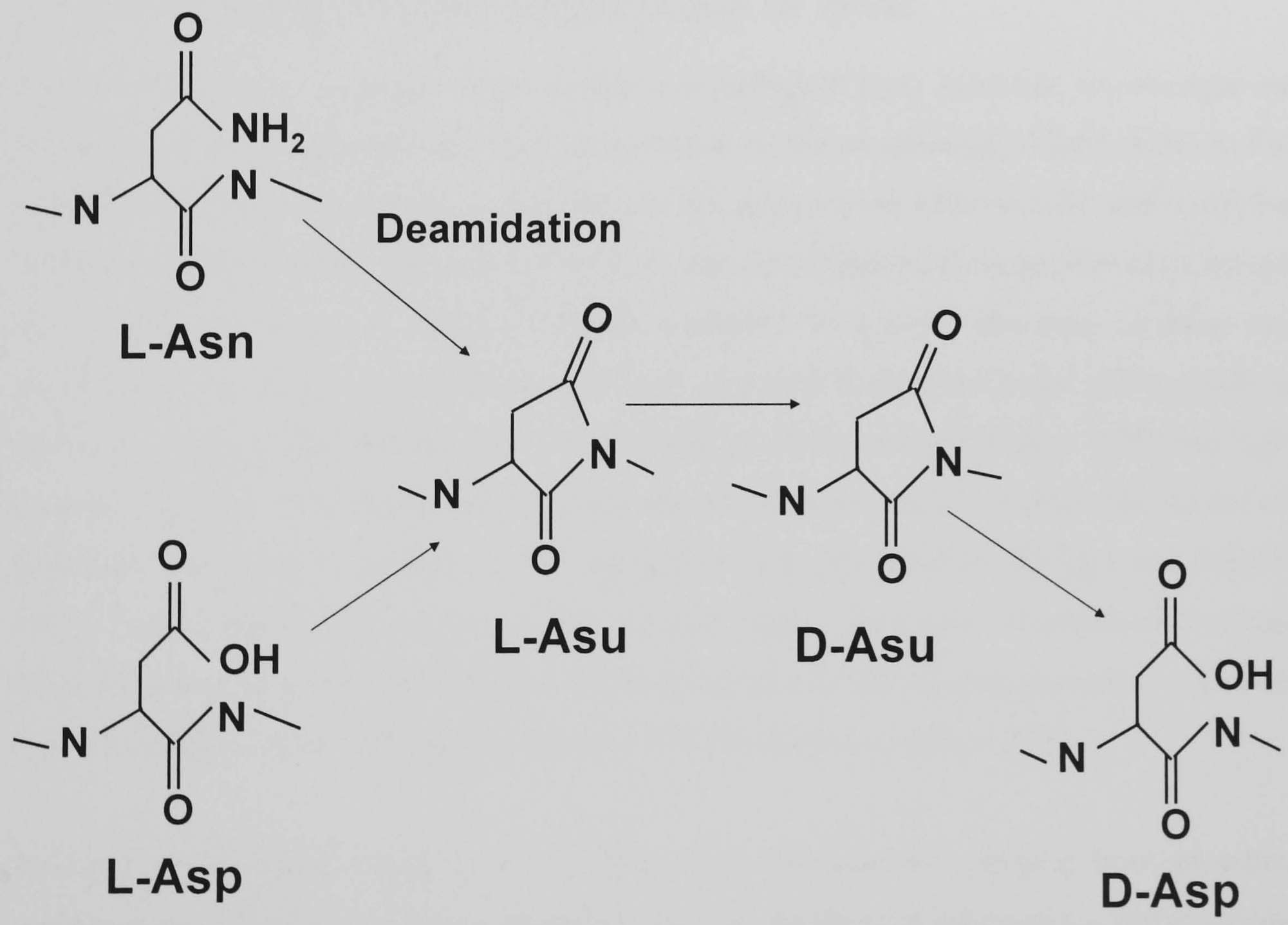


Figure 1.5 - Racemisation of L-aspartic acid (L-Asp) and deamidation of L-asparagine (L-Asn) via a succinimide intermediate (Asu) to form D-aspartic acid (D-Asp)(Modified from van Duin and Collins 1998).

When aspartic acid lies between two different amino acids, a reaction with the carboxy-terminal nitrogen allows the formation of a cyclic succinimide (Asu). L-aspartic acid converts to L-succinimide, which then converts to D-succinimide and then to D-aspartic acid (Fig. 1.5). In the case of L-asparagine, a deamidation reaction occurs before the intermediate is formed. The removal of hydrogen from the chiral carbon is facilitated by the partial electron delocalisation from the ring formation, as well as the planar structure of the succinimide. Both D-aspartic acid and D-asparagine appear to give D-aspartic after hydrolysis (Brunauer and Clarke 1986). Bound asparagine is found to racemise faster than bound aspartic acid via a more rapid succinimide formation (Brennan and Clarke 1993), although its deamidation reaction is affected by neighbouring residues, where smaller residues, such as glycine, make the reaction faster (Geiger and Clarke 1987).

1.5 Monitoring Protein Degradation in Bone

Preservation can be assessed either on the morphological level (whether macroscopic or microscopic) or on the molecular level. Preservation at the morphological level refers to the overall preservation of the bone tissue down to the preservation of bone cells which may be observed by histological examination. However, the assessment of gross preservation, by eye or by light microscopy, is a highly subjective process. In general, the rates of decay are considered to be inversely proportional to bone size (von Endt and Ortner 1984), possibly due to the high internal surface area (~85-170 m^2g^{-1})(Lowenstam and Weiner 1989) and high porosity of bone. Any change in this pore size distribution could influence the extent of diagenetic alteration via changes in the interaction with groundwater (Hedges and Millard 1995; Nielsen-Marsh and Hedges 1999). Several studies have also emphasised the link between biomolecule survival (notably DNA) and degree of histological preservation (Bailey *et al.* 1996; Colson *et al.* 1997; Hagelberg *et al.* 1991; Richards and Sykes 1995).

Molecular preservation refers to the survival of the biomolecules ranging from proteins, nucleic acids and lipids to polysaccharides, but the standard measurements for assessing biomolecular preservation tend to focus on the extent of degradation presented by the amino acids of proteins. The two most common methods of assessing the extent of protein degradation are by nitrogen content and amino acid analyses. Measuring the amount of nitrogen is now commonly used to estimate the amount of remaining protein in

archaeological bone (Bailey *et al.* 1996; Colson *et al.* 1997; Hedges *et al.* 1995) which is thought to decrease with geological age. However, these values can be complicated by the presence of contamination, especially in older specimens. Analysing the composition of amino acids within a sample has also been extensively used; amino acid analyses of Pleistocene bones and teeth were used to suggest that collagen is capable of surviving the vicissitudes of fossilisation and geological time (Dungworth *et al.* 1975; Wyckoff 1972). Most analyses of more ancient fossils published by various authors showed significant differences in their compositional profiles to that of collagen (Armstrong 1966; Wyckoff 1972). The extent of amino acid racemisation can give further information regarding the extent of degradation of proteins.

The initial chemical reaction in the diagenesis of proteins in bone is the progressive hydrolysis of the peptide bonds between adjacent amino acids in the protein chain (Goodfriend *et al.* 1997). An increase in peptide fragments caused by hydrolysis increases the number of amino-terminal amino acids available for more rapid racemisation (Goodfriend *et al.* 1997). This hydrolysis lowers the molecular weight of the protein and increases its solubility. Some amino acids, such as valine and isoleucine, form more stable peptide bonds than others (such as serine, alanine and glycine) which will influence the susceptibility for hydrolysis (Hare *et al.* 1975). Although water is necessary for the hydrolysis of the peptide-bound amino acids and for racemisation, the indigenous water in bone is usually sufficient for these reactions to occur; if excess liquid water is present, varying amounts of leaching can occur in addition to the hydrolysis and racemisation reactions, as bone is an 'open' system. Modern bone generally contains 7-10% indigenous water, which is reversibly bound and can be removed by heating to 110°C. In most natural environments, water is present in excess, at least periodically, resulting in the leaching of soluble components.

Hare (1980) simulated the effects of environmental leaching by heating bone samples in various amounts of water; in one experiment, the water was frequently changed so that bone fragments were in contact with fresh water most of the time. He concluded that the sample heated under anhydrous conditions looked similar to the unheated bone, with no nitrogen loss and little racemisation. In contrast, the 'closed-system' sample heated in indigenous water or water vapour showed the highest degree of racemisation, although very little

nitrogen loss. With liquid water present, further dramatic effects due to leaching were apparent, with about 85% protein leached out when the water:bone ratio was 10:1. This increased to 95% with the continuous water replacement/leaching experiment. However, as more protein material was leached out, the proportion of D- to L- amino acids diminished due to the removal of the soluble fraction preferentially removing the D- amino acids (Hare 1980). Goodfriend *et al.* (1997) state that the leaching of free amino acids causes the decreased apparent racemisation of total amino acids because the free amino acids are more racemised than peptide bound amino acids. However, this would not account for the racemisation patterns seen in polar amino acids; aspartic acid, asparagine, glutamic acid, glutamine, serine and threonine, in which side chain reactions effect racemisation in peptide-bound residues. Due to the highly constrained triple helical structure of bone collagen (I), it is believed that in-chain racemisation cannot occur. Even the succinimide reaction, allowing the in-chain racemisation of aspartic acid, is unfavourable in the tightly-constrained helix (van Duin and Collins 1998) and thought only to occur in the telopeptides of collagen and the much less abundant NCPs (Collins *et al.* 1999). Upon heating beyond the melting temperature of collagen, the hydrogen bonds that stabilise the molecules of collagen are broken and shrinkage occurs. Upon shrinkage, the aspartic acid residues no longer strained undergo a succinimide reaction, resulting in aspartic acid racemisation. The extent of this racemisation is believed to be associated with the extent of denaturation of the collagen triple helix and thus the degradation of the sample.

1.5.1 Amino Acid Racemisation in Fossils

In bone this measure of protein degradation is complicated by the inability for amino acids such as Asx, to racemise whilst constrained within the triple helical structure of collagen. The amino acid racemisation measurements in fossil bone samples are thus considered to reflect the extent of collagen denaturation and racemisation of non-triple helical amino acids such as collagen telopeptides and NCPs. Armstrong *et al.* (1983) carried out early work analyzing proteins in vertebrate fossils from sediments ranging from 1300 years B.C. to approximately 400 million years ago. The amino acid analyses revealed little evidence of intact collagen in fossils of Tertiary (65 – 1.8 Ma), Mesozoic (251 – 65 Ma), or Palaeozoic (542-251 Ma) age. Although the Palaeozoic material studied contained a general background of amino acids

common to both fossils and sediments, the degree of amino acid racemisation was used to measure the modern contamination of geologically older samples. Other uses of AAR in bone samples have been to estimate age at death (Ritz-Timme *et al.* 2000), and attempts to date fossil bones using AAR. However, the latter has been shown to be invalid: results of an extinct goat from Mallorca, *Myotragus balearicus*, did not agree with the carbon-14 (^{14}C) dates obtained (Turekian and Bada 1972). Thus the main role of AAR in bone samples has been for methods of screening biomolecule survival. Poinar *et al.* (1996) suggested that if the D/L value of aspartic acid is greater than 0.08, bone samples are unsuitable for DNA analysis, founded on the concept that the depurination of DNA and amino acid racemisation take place at similar rates in aqueous solution at neutral pH and are affected by similar factors (Bada *et al.* 1994), although this suggested cut-off value was later increased to between 0.12 (with aDNA success from a Neanderthal sample) and 0.15 (the lowest Asx D/L value for a sample not yielding aDNA in their study)(Poinar and Stankiewicz 1999).

1.6 Bone Collagen

1.6.1 Collagen Synthesis

Most of the collagen (I) in mammals is produced by fibroblasts and osteoblasts. Once the three peptide chains are formed within the cell (known as the preproprotein), they are subject to post-translational modification, where some of the proline and lysine residues are hydroxylated (a process dependent on ascorbic acid/vitamin C) and enzymatic glycosylation of the individual chains also occurs. Following this, the triple helical structure is formed and known as the procollagen molecule, which is then secreted from the cell by exocytosis. Outside the cell, registration peptides that were present in the preprocollagen, are cleaved by procollagen peptidase and tropocollagen (TC) is formed (Fig. 1.6). TC molecules self-assemble to form collagen fibrils, and collagen fibrils in turn form collagen fibers (Kadler 1994).

1.6.2 Tropocollagen molecule

The basic collagen (I) molecule is approximately 300 nm in length and 1.5 nm in diameter (Kadler *et al.* 1996). It consists of three α-chains, two identical α1 (I) chains and one

genetically different α2 (I) chain (Vuorio and de Crombrugghe 1998) that together form a right-handed supercoiled triple helix (Rich and Crick 1961)(Fig. 8). Each helically-wound alpha chain contains 338 triplets of amino acids constructed from repeating Gly-Xaa-Yaa triplets, where Xaa and Yaa can be any amino acid but frequently the imino acids proline and hydroxyproline (Hulmes 1992). The glycine at every third position of each chain is a prerequisite for the folding of the three chains into such a tight triple helix. The pyrrolidine ring of proline introduces a left-handed twist in the peptide backbone of each α-chain, placing glycine residues into the centre of a triple helix as the tight packing of the protein strands can accommodate no other residue. To accommodate glycine residues in different chains into the triple helix, adjacent chains are mutually staggered by one residue. For each Gly-Xaa-Yaa triplet, one hydrogen bond forms between the amide hydrogen atom of glycine in one chain and the carbonyl oxygen of residue Xaa in an adjacent chain (Horton *et al.* 2002). The thermal stability of the triple helix is enhanced by the hydroxyl group on the pyrrolidine ring of Hyp residues (Berg and Prockop 1973). This extra stability is thought to arise from hydrogen bonds mediated by a network of bridging water molecules. These typically consist of two water molecules that link a Hyp side-chain of one strand to a main-chain carbonyl of another strand (Bella *et al.* 1995).

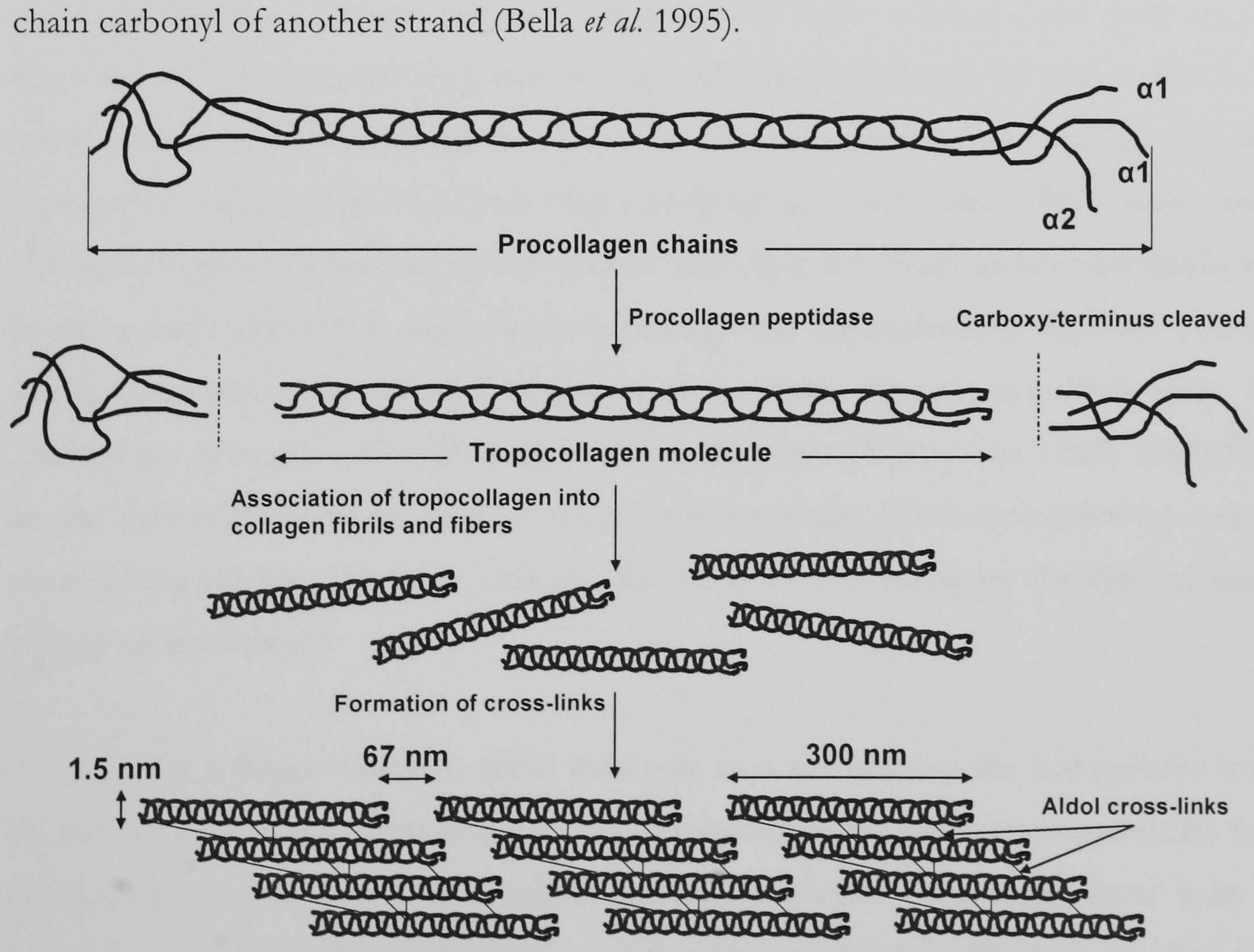


Figure 1.6 - Formation of collagen fibers (Klug and Cummings 1997)

Collagen (I) molecules contain an uninterrupted triple helix of approximately 300 nm in length and 1.5 nm in diameter flanked by short nonhelical telopeptides. The telopeptides, which do not have a repeating Gly-Xaa-Yaa structure and do not adopt a triple helical conformation, are essential for fibril formation (Kadler *et al.* 1996). The telopeptides are the most immunogenic regions of the collagen (I) molecule and the most carboxy-terminal part of the carboxy-terminal telopeptide of the collagen (I) α1 chain D-G-G-R-Y-Y is also highly conserved as an activator for polymorphonuclear leucocytes (Monboisse *et al.* 1990). The α2 chain carboxy-terminal telopeptide which does not have this sequence also lacks this property. In addition, a specific property of the α1 chain carboxy-terminal telopeptide is that it adopts a folded conformation with a sharp hairpin turn around residues 13 and 14 of the 25-residue telopeptide (Orgel *et al.* 2000). Both α1 chain carboxy-terminal and amino-terminal telopeptide and the α2-chain amino-terminal telopeptide regions take part in cross-link formation and some aspartyl residues in the telopeptides are isomerised (α→β) and racemised (L→D).

To form fibrils and fibers, the collagen molecules form a linear array with the amino-terminus of one molecule lying near to the carboxy-terminus of the next molecule. These linear arrays are stacked alongside each other in a staggered fashion, the gaps in adjacent linear arrays being offset by 67 nm (Fig. 1.6) (Prigodich and Vesely 1997). Many individual collagen molecules associate by end-on and side-on interactions to generate fibrils that are approximately cylindrical, and in bone have diameters approximately 30-80 nm (Tzaphlidou 2005) and surrounded by and impregnated by small mineral crystallites, only tens of nanometers in length and width and 2-4 nm in thickness (Weiner and Traub 1992). Collagen *in vivo*, generally occurs as white, opaque, non-branching fibrils embedded in a matrix of mucopolysaccharide and other proteins, the amounts depending on the type of tissue and the age of the animal.

In order for collagen fibrils to fulfill their role as a scaffold for the extracellular matrix of connective tissues, the fibrils provide sites of attachment for other macromolecules from the extracellular matrix, and cells. Members of several classes of proteins have been shown experimentally to bind to collagen (I), and to immunolocalise to fibrils *in situ*. For example,

decorin is a small proteoglycan that binds at two sites on collagen (I), situated at 50 nm and 100 nm from the amino-terminal end of the triple helix (Kadler 1994).

1.6.3 Post-Translational Modifications

1.6.3.1 Hydroxylation

The collagen molecule has some unusual post-translational modifications (PTMs) which are the hydroxylation of some proline and lysine residues. There are three enzymes involved in the conversion of peptidyl lysine to hydroxylysine, peptidyl proline to 4-hydroxyproline and some peptidyl prolines to 3-hydroxyproline as well as requiring molecular oxygen and ascorbate. The prolyl-4-hydroxylases (P4Hs) all require molecular oxygen, L-ascorbic acid, 2-oxoglutarate, and Fe^{2+} to catalyse the hydroxylation of Pro in peptides (Hirsila *et al.* 2003; Kukkola *et al.* 2003). Hydroxylation of peptidyl proline by prolyl 4-hydroxylase, which requires the proline to be in the context -Gly-Xaa-Pro-, induces conformational changes in the nascent collagen peptide backbone that promotes triple helix formation *in vitro* (Kadler 1994). The presence of hydroxyproline residues also increase molecular stability by allowing a second hydrogen bond per Gly-Xaa-Yaa triplet involving a water bridge, connecting the hydroxyl (-OH) group on a Hyp residue in the Yaa position of a triplet, with an –NH group in a neighbouring chain (Brodsky 1999; Brodsky and Ramshaw 1997). In fibril-forming collagens, approximately 50% of the proline residues contain a hydroxyl group at position 4 and the extent of prolyl-hydroxylation is species-dependent (Stanlotter *et al.* 1993). The organisms living at lower environmental temperatures show a lower extent of hydroxylation (Glimcher *et al.* 1990). The presence of 4-hydroxyproline is essential for intermolecular hydrogen bonds and thus contributes to the thermal stability of the triple helix (Stanlotter *et al.* 1993). In addition to 4-hydroxyproline, proline residues in collagen are also hydroxylated at position 3 (see Fig. 1.7). 3-Hydroxyproline occurs infrequently and only at the Xaa position and when 4-hydroxyproline occurs at the Yaa position of the Gly-Xaa-Yaa repeating structure. Although it is clear that 4-hydroxyproline is required for the stabilisation of the triple helix by hydrogen bonding, little is known about the functions of 3-hydroxyproline.

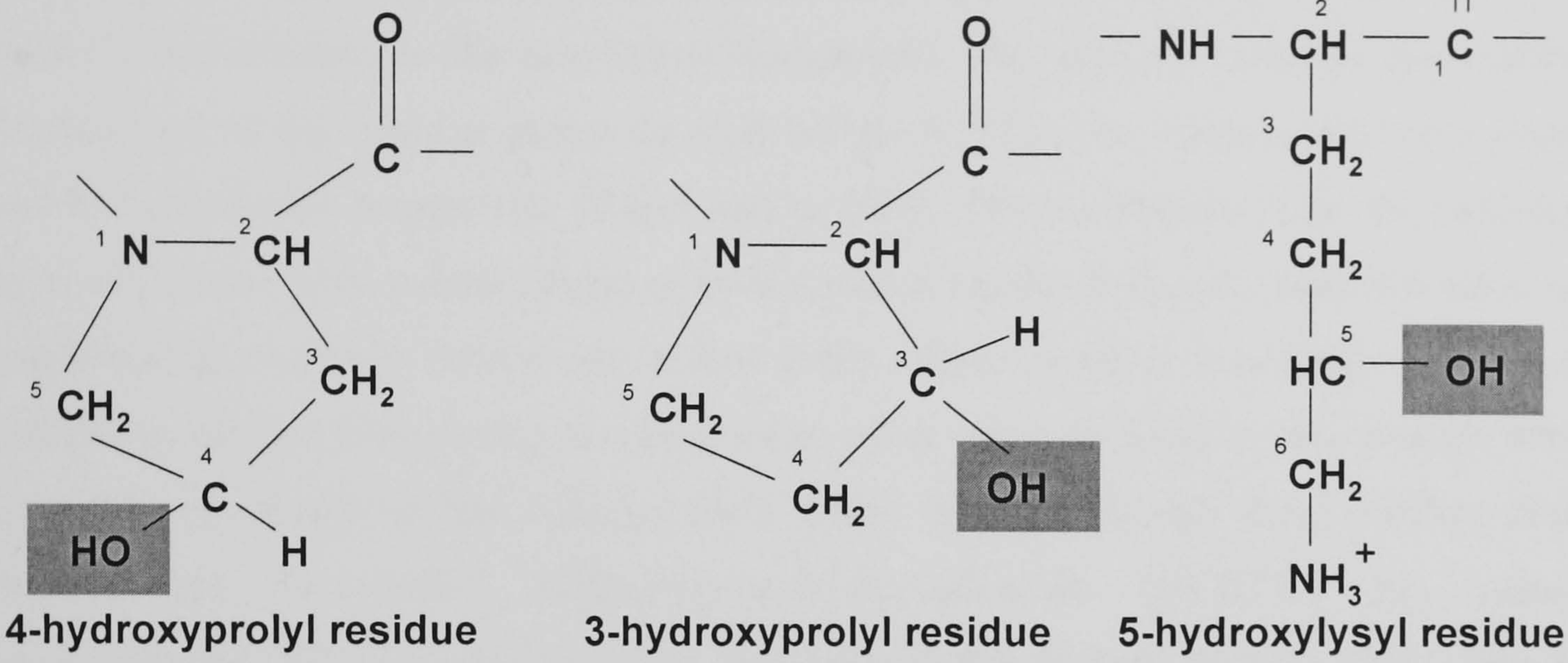


Figure 1.7 - Structures of the three most common post-tranlsational modifications in bone collagen (I) with additional hydroxyl groups shaded grey.

The hydroxylation of peptidyl lysine by the enzyme lysyl hydroxylase is commonly reported to require the lysine to be in the context –Gly-Xaa-Lys- but Royce *et al.* (1985) highlight the inability of this enzyme to hydroxylate the telopeptide lysine residues. This had led to the speculative presence of another independent enzyme that caries out the telopeptide hydroxylations (Knott and Bailey 1998). The hydroxyl groups of hydroxylysine serve important functions in providing attachment sites for glycosyl residues and in the formation of interchain cross-links.

1.6.3.2 Cross-Links

The assembly of types I, II and III collagen into fibrils is accompanied (or followed immediately) by the formation of inter- and intramolecular covalent cross-links between α-chains which confer high tensile and mechanical strength on the fibrils (Kadler *et al.* 1996). In bone the majority of cross-links are believed to be via certain lysine residues of collagen after the post-translational modification into hydroxylysine by lysyl 5-hydroxylase. Hydroxylysine is almost exclusively found in the Yaa positions of the repeating Gly-Xaa-Yaa sequences in the triple helix as well as the non-helical telopeptides found at the ends of each molecule (Knott and Bailey 1998). The two important functions of hydroxylysyl groups are; to stabilise the intramolecular and intermolecular cross links, and serve as attachment sites for carbohydrates (Kivirikko and Pihlajaniemi 1998) which are considered to be linked to the

regulation of collagen molecules into supramolecular assemblies (Notbohm *et al.* 1999). The cross-linking is initiated by the enzyme lysyl oxidase which modifies particular lysine and hydoxylysine residues in the non-helical telopeptides. The enzyme catalyzes the oxidative deamination of the ε-amino group on each residue to form the reactive aldehydes allysine and hydroxyallysine respectively (Kagan and Li 2003). These aldehydes form the cross-links by condensation with another lysine or hydroxylysine residue within the helical portion of a neighbouring molecule (Knott and Bailey 1998). These covalent cross-links prevent the collagen molecules from sliding past each other under stress, in doing so they provide tensile strength and stability to the collagen fibril (Bailey 2001). Although these divalent bonds (usually the keto-imines hydroxylysino-5-ketonorleucine (HLKNL) or lysino-5-ketonorleucine (LKNL)) are heat labile and hydrolysable in acid (Robins 1988), they can stabilise *in vivo* by maturation, which involves the further condensation of the immature cross-link with another hydroxyallysine aldehyde to form tri-valent cross-links (Robins and Duncan 1987).

Newly formed cross-links occurring in young tissue contain secondary amide or aldehyde groups that can be examined by mild reduction in borohydride and are frequently referred to as 'reducible' or immature cross-links (Fig. 1.8). The intermolecular cross-links in older tissue cannot be reduced by borohydride and are referred to as 'non-reducible' or mature cross-links (Kadler 1994). During cross-link maturation *in vivo*, native collagen fibres display a progressive conversion of the reducible cross-links into non-reducible cross-links.

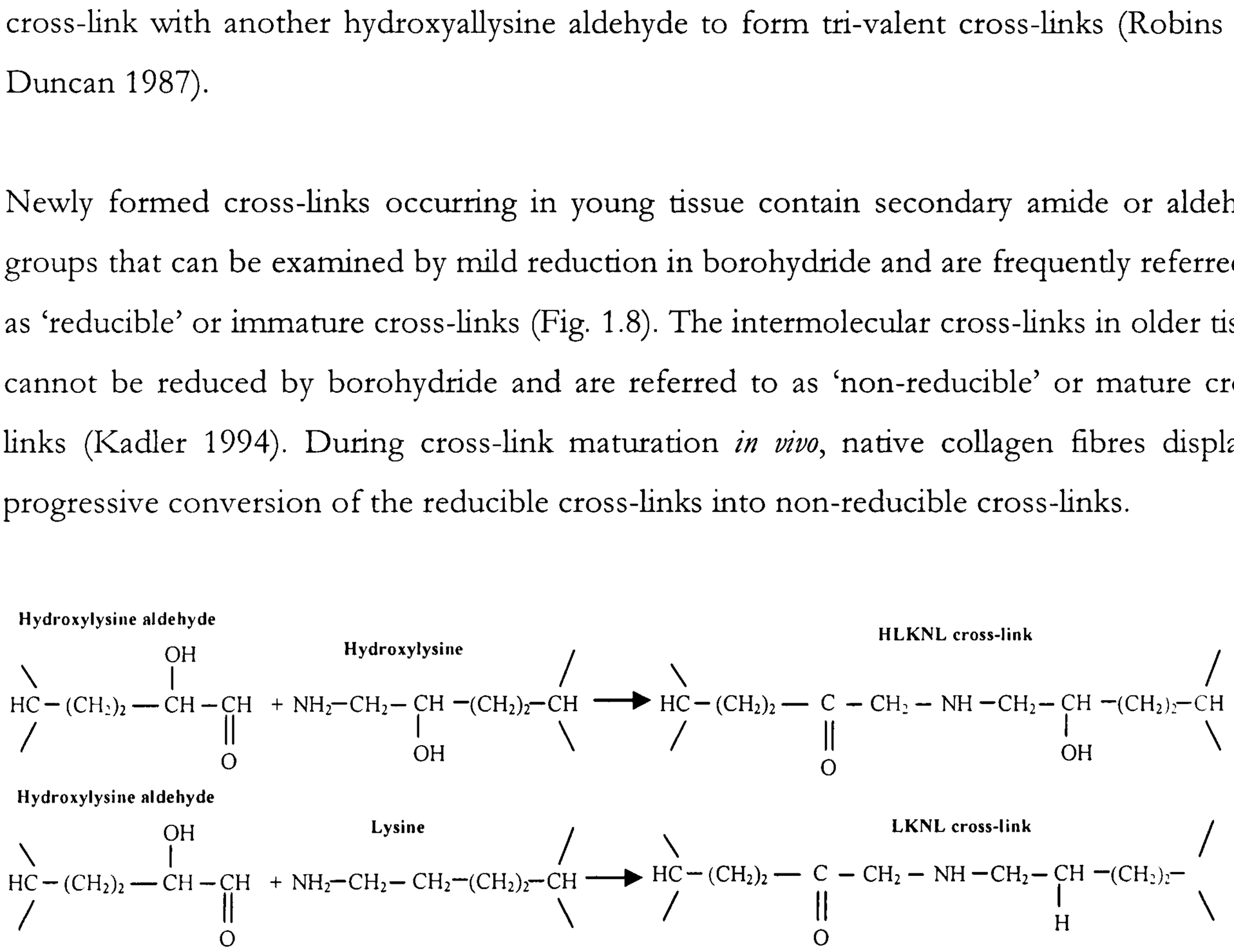


Figure 1.8 - Formation of the immature ketoimine cross-links by condensation and subsequent Amadori rearrangement, of a telopeptide hydroxylysine aldehyde with either a helical hydroxylysine or a helical lysine residue to produce HLKNL and LKNL cross-links respectively (Redrawn from Knott and Bailey 1998).

Collagen is known for possessing additional types of cross-links involving saccharides that make the molecules of collagen much more stable at higher temperatures. Some hydroxylysyl residues are glycosylated to galactosylhydroxylysine and glucosylgalactosylhydroxylysine by the enzymes hydroxylysyl galactosyltransferase and galactosyl hydroxylysl glucosyltransferase. The first of these enzymes adds galactose to the hydroxylysyl residues and the second adds glucose to the galactosylhydroxylysine residues (Kadler 1994). These early glycation products are reversible and do not accumulate in most proteins. In long-lived proteins like collagen, these glycation products are able to undergo a series of reactions that result in more persistent advanced glycation end products (AGEs or Maillard Products). The bone of young individuals contains mainly the enzymatic forms of saccharide cross-linkages (glycosylations), but as the individuals' age, the amount of additional glycation increases, resulting in more AGEs. Vertebrate collagens contain oxygen-linked (O-linked) monosaccharide galactose and the disaccharide glucosyl galactose that are covalently attached to hydroxylysyl residues in the triple-helical domains of the α-chains (Kadler 1994). The nonenzymic addition of sugars (notably glucose) to the ϵ-amino group of lysyl residues (the Maillard reaction of 'browning') occurs in most proteins and has been observed in proteins from haemoglobin to collagen (Kadler 1994). Although in non-mineralising tissues like skin and tendon, this causes a significant increase in the stiffness of the tissue, in bone, which is turned over more quickly, AGEs do not accumulate with age to the same extent (Knott and Bailey, 1998).

1.6.4 Collagen Degradation into Gelatine

The gelatins are a class of proteinaceous substances that do not exist in nature but are derived from the parent protein collagen, by any one of a number of procedures involving the destruction of the secondary structure of collagen and in some aspects the primary and tertiary structures (Veis 1964). The collagen-gelatine transition is the process whereby the highly organised, quasi-crystalline, water-insoluble collagen fiber is transformed from an asymmetric network of linked TC molecules to a system of water-soluble, independent molecules with a much lower degree of internal order. The acidic and basic functional groups of the amino acid side chains confer polyelectrolyte characteristics on the random gelatine chain. These electrically charged sites govern to some extent the interactions

between gelatine molecules and also between gelatine molecules and the solvent. They affect the viscosity and all other hydrodynamic properties. Hence to understand and characterise a gelatine system one must also take into consideration the net charge of the gelatin molecule, the total charge, the nature of the ionisable groups and their internal distribution.

1.6.4.1 Unmineralised Collagen-Gelatine Transition

Demineralised bone collagen fibers are highly insoluble due to the many intra and intermolecular cross-links in contrast to collagens in soft connective tissues (Eyre 1988). Although mature mammalian collagens contain a small soluble collagen component, convertible to gelatin by heating briefly at 40-45°C, the bulk of the unmineralised collagen in skin, sinew and bone is prevented from dissolving by its covalent cross-linking and spatial confinement of the molecule within the fibrils and fibers (Miles and Ghelashvili 1999). In order to explain the large increase in thermal stability offered by the cross-linked fibril, Miles and Ghelashvili (1999) proposed a 'polymer-in-a-box' model whereby the collagen molecule is confined within a 'box', the walls of which are formed by adjacent TC molecules within the fibril. This model suggests that as the available space around the collagen molecule is reduced (resulting in a loss of configurational entropy) there is a corresponding increase in the amount of energy required to change state (Gibbs free energy of activation), and thus interpreted as being more thermally stable. The increased stability induced by cross-linking is thus attributed to close-packing and reduction in inter-molecular spacing as the cross-links hold the molecules closer together. The packing of collagen molecules into fibrils and fibers allows for the denaturation temperature of mammalian skin, tendon and demineralised bone to be approximately 25°C higher than the denaturation temperature of the individual TC molecule, which is close to body temperature (Miles and Bailey 1999).

One part of the conversion process, thermal denaturation, results in the breakdown of hydrogen and electrostatic bonds, so disarranging the collagen helix that the three entwined protein chains are freed from each other and pass into solution as more random coils. Warming to approximately 40°C is sufficient to release 'parent' gelatine (collagen alpha chains) from newly formed collagen which has no cross-links, but thermal denaturation alone is insufficient for the release of gelatin from more mature collagen because of the stabilising effect of cross-links. Warming to higher temperatures (65°C) disrupts the triple

helical structure of the collagen as the hydrogen bonding between the molecules is disrupted, but the presence of the cross-links will keep the chains together in 'cross-linked' gelatine. The other part of the conversion process is the hydrolytic breakdown of covalent bonds. Those covalent bonds likely to be broken include peptide bonds as well as those forming cross-links. More severe temperatures greater than 65°C are necessary to break covalent bonds, and gelatinisation is usually carried out at pH away from neutral (e.g. the presence of chemical treatments such as acid or lime) in order to break the cross-links.

The transition is accelerated by elevated temperature, extremes of pH, organic acids and damage to the collagen structure (Miles *et al.* 2000). Collins *et al.* (1995) put forward a model for collagen degradation which is driven by the chemical hydrolysis of peptide bonds and assumes that water is in excess in order to make the rate of hydrolysis a 'first-order' reaction. However, despite earlier assumptions that the degradation of random TC molecules resulted in the collapse of the fibril at a 'first-order' rate (Collins *et al.* 1995), Koon (2006) observed that collagen degradation is dependent on the packing of the collagen molecules into microfibrils and fibrils where particular sections of the fibril denature, which is quickly followed by a rapid collapse of the entire fibril. This is consistent with the model of denaturation proposed by Miles and Ghelashvili (1999), where once the three alpha chains at the 'thermally labile region' are uncoupled, the whole structure becomes unstable and unzips. This thermally labile region in type 1 collagen was determined to be a hydroxyproline-free sequence of 65 amino acid residues near the carboxy-terminus of the molecule (Miles and Bailey 2001; Miles *et al.* 1995).

1.6.4.2 Mineralised Collagen-Gelatine Transition

The collagen triple helix is surrounded by water molecules that are nucleated at hydroxyproline side chains (Bella *et al.* 1995). Miles *et al.* (2005) showed that reducing the water content of fibers reduces the available space for the individual molecules to collapse into, resulting in elevated denaturation temperatures. In bone, the transition from collagen to gelatine is not only retarded by the close packing of the fibrils but also by the presence of the surrounding hydroxyapatite mineral. It is this replacement of some water molecules with the mineral that is thought to increase the characteristic 65°C endothermic transition which occurs in untreated non-mineralised collagen to 155°C in mineralised tissues (Kronick and

Cooke 1996). The mineralisation of bone collagen is consistent with, and supplementary to, the 'polymer-in-a-box' theory as the mineral further reduces the amount of intermolecular space available for the collagen molecules to move into because the positions of water molecules are also being replaced by mineral (Fratzl *et al.* 1992). As a consequence, more energy is required for denaturation to occur, thus the fibrils are more thermally stable. This model is also consistent with the observations of Koon (2006), implying an all-or-nothing process of fibril decay, rather than the random polymer scission model proposed by Collins *et al.* (1995), a process which does, however, appear to be consistent with the observed pattern of DNA degradation in bone.

1.6.5 Collagen Variability

Due to the large size of the triple-stranded subunit of collagen, the primary structure has only been determined for a few samples (Fietzek and Kuhn 1976; Hulmes *et al.* 1973; Kadler 1994). This is largely due to the very large size of the two primary structures involved, the complexity of the post-translational modifications involved, and because the Gly-Xaa-Yaa repeating pattern of collagen gives the impression that it would exhibit little species variability. Alignments of the currently available sequences (see Appendix 1.1 for alignment of published sequences) make it clear that the α1 (I) chain is highly conserved.

Table 1.4 - Number of amino acid differences for collagen a1/a2 (I) chains in published sequences of human, cattle, dog, mouse and chicken from the UniProt protein database.

Species	Human	Cattle	Dog	Mouse	Chicken
Human	0	25/68	16/56	44/96	70/160
Cattle		0	32/58	53/98	87/169
Dog			0	37/91	70/170
Mouse				0	74/163
Chicken					0

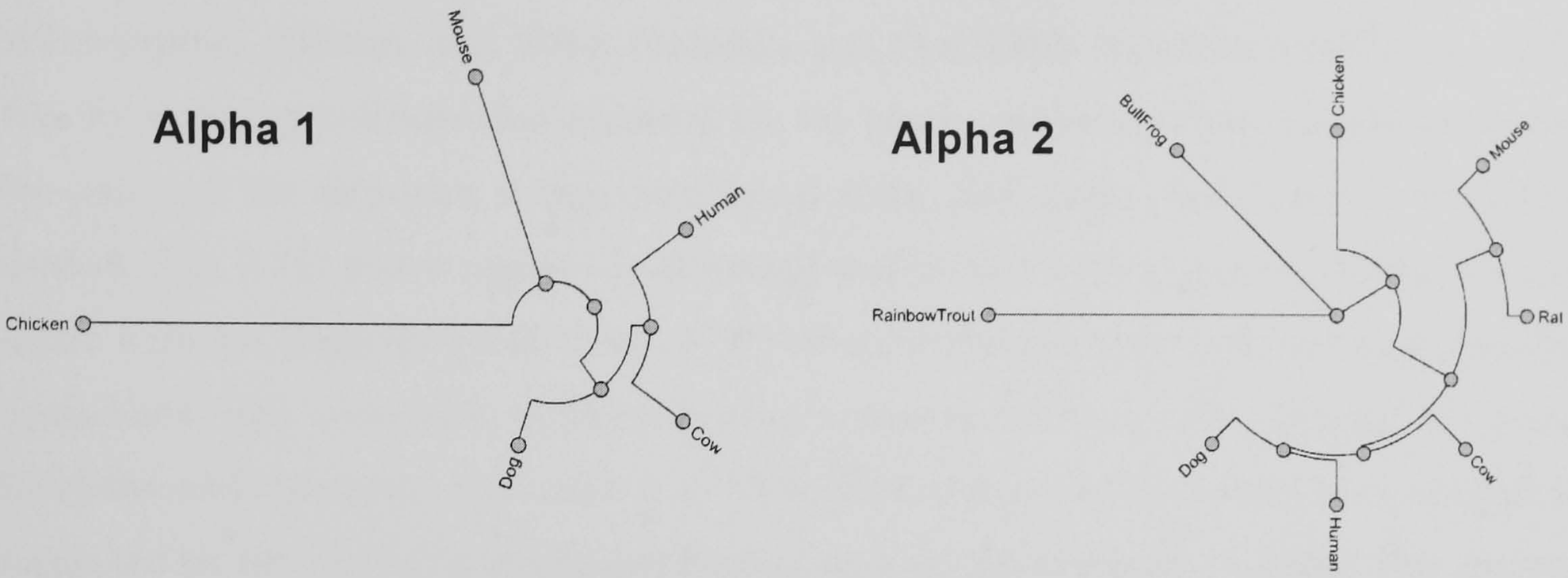


Figure 1.9 - Circular tree layout phylogram using collagen a1(I) and a2 (I) chain sequences (rooted by chicken and rainbow trout sequences respectively) published in Uniprot, aligned in Clustal X and created in Geneious 3.5.6.

The α2 (I) chain is much more hydrophobic, has a lower imino acid content than the α1 (I) chain (McBride *et al.* 1997) and shows a much greater amount of variability (Table 1.4). The heterotrimeric composition of type 1 procollagen in a wide range of vertebrate species, including mammals, birds, and some fish species (Kelly *et al.* 1988; Lewis and Piez 1964) suggests that it supplies a selective advantage to vertebrates for tissue specialization (McBride *et al.* 1997). With the variability of collagen being potentially sufficient to identify species (Fig. 1.9), the greatest limitation is the ability to retrieve and handle data from protein mass spectrometry with such a large and complex protein as the target (particularly with a very limited number of published sequences available). Hence this thesis focuses at developing methods for distinguishing between particular species of interest including cattle, pig, chicken, and the morphologically similar sheep and goat.

1.7 Osteocalcin

The vitamin K-dependent Ca^{2+} binding protein OC is the most abundant NCP of adult bone, which is secreted by the osteoblasts, and present in bone matrix and serum (Colombo *et al.* 1993). Its name derives from its abundance in osseous tissue (10-20% of the noncollagenous protein; up to 2.2 molecules per 40 x 8 x 3 nm of bone mineral) and its affinity for Ca^{2+} (Hauschka 1981). Isolated osteocalcins are small acidic proteins (pI ≈ 4.0) containing 46-50 amino acid residues, three of these amino acids are Gla residues at positions 17, 21 and 24 (Hauschka *et al.* 1989) that appear to be responsible for the Ca^{2+}

binding properties of OC to phospholipids vesicles (Gendreau *et al.* 1989) and hydroxyapatite (Delmas *et al.* 1984; Hauschka and Carr 1982; Hauschka and Wians 1989). The three most prominent characteristics are the seven-residue disulfide bridged loop near the centre of the sequence; a high number of acidic side chains, including the three Gla residues (Fig. 1.10) and a region of alternating proline and hydroxyproline residues in the amino terminus (Prigodich and Vesely 1997) which is speculated to be involved in specific interactions with membrane surfaces and/or receptors of bone cells essential for bone formation and resorption (Colombo *et al.* 1993). The affinity of OC for the bone mineral is supported by the clustering of calcium binding sites on the molecules' surface. This derives from the presence of the doubly charged Gla residues and charged Asp residue all being concentrated in a common surface region on the same side of the molecule (Frazao *et al.* 2005).

The Gla helix domain is likely to be of ancient evolutionary origin as it occurs in a subset of proteins and peptides which interact in a particular fashion with Ca^{2+} ions and Ca^{2+} mineral surfaces. The staggering of Gla residues in primary sequence by three or four positions allows them to project from one face of the alpha helix (Fig. 1.10), seen in both OC and MGP in bone (Price and Williamson 1985), and elsewhere such as the blood proteins prothrombin and Factor X (Nelsestuen *et al.* 1974; Schwalbe *et al.* 1989). Due to this ability to bind to the HAP, early studies suggested a physiological role for OC as a matrix signal for the recruitment and differentiation of osteoclasts (Glowacki *et al.* 1991; Liggett *et al.* 1994; Malone 1982). However, Ducy *et al.* (1996) showed that bone mineralisation is not affected in OC-deficient mice. They did however show that the OC-deficient mice had a higher bone mass, and as the osteoblast surface was not increased, each osteoblast must have theoretically produced more hydroxyapatite/bone mineral. This more recent evidence indicated that the normal function of OC was not to allow mineral deposition to occur, but possibly to limit bone formation without impairing bone resorption or mineralisation (Dowd *et al.* 2003).

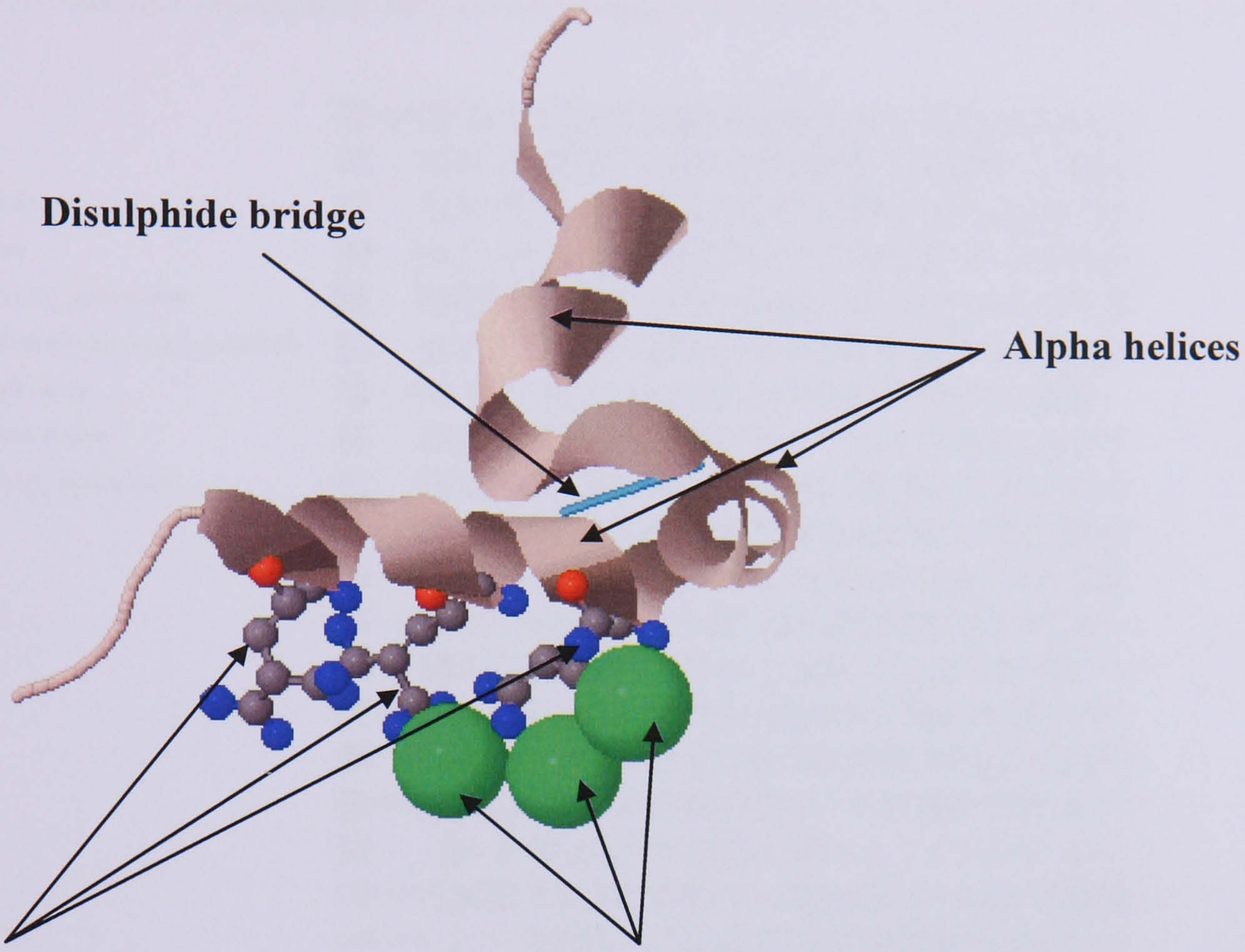


Figure 1.10 - Illustration of the docking of porcine OC on the HAP, showing the affinity of the 3 Gla residues at positions 17, 21 and 24 to bind to the calcium ions (in Green) of the HAP (Created in jmol using 1q3m.pdb from Hoang et al. 2003).

The OC precursor is synthesised during progressive mineralisation and maturation of the bone tissue and would be bound to either free Ca^{2+} ions and then incorporated onto the surface of the HAP crystals, or bound directly to calcium ions already present as part of a solid mineral phase (Glimcher 1989). Prigodich & Vesely (1997) showed that the addition of collagen reduces the concentration of soluble OC due to binding between the two proteins. The high affinity of OC for collagen as well as the mineral hydroxyapatite is a characteristic postulated to allow preservation in ancient bone (Collins *et al.* 1998; Collins *et al.* 2002).

1.7.1 Osteocalcin Sequence Variability for Species Identification and Phylogenetic Applications

The primary structure of OC has been determined for at least twenty different vertebrate species (Table 1.5) and all show extensive sequence homology suggesting a functional preservation throughout evolution (Hauschka *et al.* 1989).

Table 1.5 - Osteocalcin sequences taken from UniProt and aligned in Clustal and converted to rich-text format in BioEdit.

In the consensus sequence '*' = conserved residue, ':' = strong group of aminos conserved, '.' = weaker group of conserved aminos. In the actual sequences for each species, '.' indicates a conserved residue, and the shading indicates similarity using BioEdit's greyscale (black is conservative, white is non-conservative).

The strongest homology is in residues 20-25, where five of these six amino acid residues are identical and two are Gla. Conserved residues are Cys_{29}, Asp_{34}, Gly_{37}, Ala_{41}, Tyr_{42}, Tyr_{46}, and Gly_{47}, with Pro_{13} conserved throughout the mammals and Asp_{30} and Glu_{31} being conserved throughout mammals, birds and amphibians. A pair of basic residues at positions 43-44 exists in most mammalian osteocalcins. Dog and rabbit OC are unique in that residues 43-44 are Gln-Arg; and only the fishes have no basic residues at either of these positions. Among mammalian osteocalcins, hydroxylation of Pro_9 occurs in all species except dog, human,

chimpanzee and orangutan (Colombo *et al.* 1993; Nielsen-Marsh *et al.* 2005). The sequences around this Pro show a Val instead of an Ala at residue position 10, suggesting that the specificity of the osteoblast hydroxylase may be defined by a sequence of Xaa-Pro-Yaa, where Yaa can be Ala but not a bulkier side chain (Colombo *et al.* 1993). Also, Pro-48, when present, occurs in a sequence Gly-Pro-(Val/Ile) and is not hydroxylated (Colombo *et al.* 1993). Price *et al.* (1988) recognised that the sequence Glu-Xaa-Xaa-Xaa-Glu-Xaa-Cys is conserved in the vitamin K-dependent proteins of bone (OC and MGP) and blood (prothrombin, factors X, IX, and VII, and proteins C, S, and Z) and suggest it to be a recognition site for the γ-carboxylase. Colombo *et al.* (1993) recognised an additional conserved sequence in these bone and blood Gla proteins which has an aromatic cluster Ala-Tyr-Xaa-Xaa-(Phe/Tyr)-(Tyr/Phe) at residues 41-46, possibly associated with membrane interactions, which remains conserved in the additional OC sequences obtained since.

The variability of OC has the potential to be used for species identification and phylogenetic applications to ancient samples where DNA does not survive. From phylogenetic analyses of the sequences already present in protein databases (NCBi and UniProt) there are clearly significant differences between the classes of fish and birds, and between the classes of birds and mammals (Fig. 1.11). At a higher taxonomic level, OC primary sequences differ between some of the orders such as lagomorphs (rabbits), rodents, artiodactyls (cattle, sheep, pigs) and perissodactyls (horses); however, OC-based phylogeny does not appear to be consistent with expected phylogenies, e.g. human is closer to gorilla than to chimpanzee/orangutan, and mouse is closer to rabbit than to rat. Analyses at a higher level than order are more difficult using this approach but some distinctions are still possible using as little as one amino acid substitution, such as the sequence difference present between the cattle and bison OC sequences.

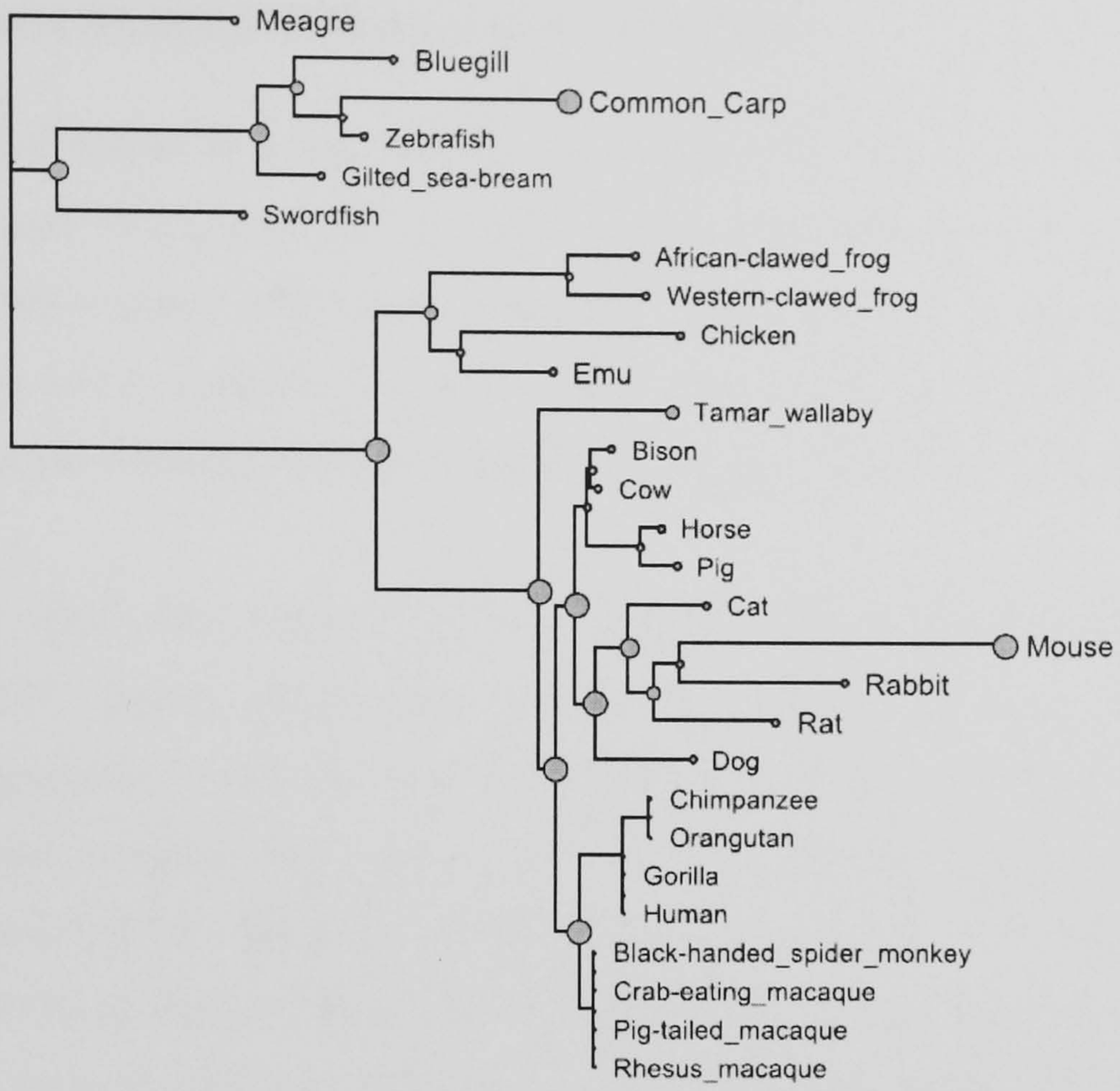


Figure 1.11 - Phylogram of all known OC sequences taken from the UniProt, aligned in Clustal X and made in Geneious 3.5.6.

However, most of the higher taxonomic differences lie within the amino-terminal region of the molecule, shown by McKnulty *et al.* (2002) and Collins *et al.* (2000) to be most labile and least protected against long term survival. Collins *et al.* (2000) used monoclonal antibodies against the three regions of the OC molecule, one against either end of the polypeptide chain, and the third against the Gla-rich mid-region. They observed a significant difference between using reconstituted "bone" (Collins *et al.* 1998) and bone powder in that not only was the OC more than one order of magnitude more detectable in the bone powder samples, but also, the decay of the mid-region epitope (from Pro_{15} to Glu_{31}) was much slower than either of the terminal regions, and much more sensitive to temperature. The aminoterminus (from His_4 to Hyp_9) proved to be the least stable, supported by the findings by Huq *et al.* (1990) that the yields from amino-terminal sequencing are much lower than concentrations estimated immunologically by using polyclonal antiserum. The carboxy-terminal region was proposed to be more stable than the amino-terminus due to its hydrophobicity (Collins *et al.* 2000).

1.7.2 Previous Studies of Osteocalcin Survival

OC is considered simple to isolate because of its unusually small size (46-50 amino acids), and of relative ease to sequence. OC is small enough to be isolated using a preparative (and inexpensive) reverse phase solid phase extraction (SPE) cartridge, a reproducible isolation procedure being first reported by Colombo *et al.* (1993). OC is also exclusive to vertebrates, eliminating possible contamination by microbes, invertebrates, or plants (Muyzer *et al.* 1992).

The protection of OC from thermal decarboxylation via mineral binding proposed by Poser and Price (1979) gained wide-spread interest in archaeology and palaeontology by theoretically improving its abilities for long-term survival. Much early work on detecting OC in ancient bone samples was carried out using immunological techniques such as Radioimmunoassay (RIA) (Huq *et al.* 1984), ELISA (Ulrich *et al.* 1987) and even the preservation of OC in dinosaur bone was reported (Muyzer *et al.* 1992). Gurley *et al.* (1990) supported this work by showing that amino acids in dinosaur bone were not only distinct from those in the surrounding matrix but also showed relative abundances that suggested preservation of a protein other than collagen. However, initial studies of archaeological bone controversially suggested that the preservation of Gla amino acids (exclusive to OC and MGP) is no better than collagenous amino acids, indicating that the hydroxyapatite does not significantly protect the Gla residue (Collins *et al.* 2000). In addition, Gla was observed to be rapidly leached from bone (King and Bada 1979). The criticism that immunological methods are difficult to apply to ancient samples because the assays are pushed to their limits (concentrated extracts, tested with antibodies at low dilution) and are therefore prone to yield false positive results (see Montgelard *et al.* 1997) induced doubt into the authenticity of previous results.

With the introduction of soft-ionisation mass spectrometry to analysing proteins and peptides, Ostrom *et al.* (2000) were able to directly observe the effects of heating OC for up to 200 hours at 100°C. Although intact OC was still observed throughout the heating experiments, studying the hydrolysis fragments they were able to derive the most thermally labile peptide bonds in the molecule. Their findings supported those of Collins *et al.* (2000), where the amino-terminus is apparently the least stable, and the carboxy-terminus is the next

fragment to be created via hydrolysis at more extensive heating regimes than for the amino-terminus. Collins *et al.* (2000) modeled the survival of the alpha helices in OC and found that by extrapolating to lower temperatures (<10°C), the alpha helix has the potential to survive for millions of years (Collins *et al.* 2000) However, the relationship between the state of mineral recrystalisation and protein survival needs to be further explored. For example, most fossil dinosaur bone is highly crystalline francolite (Hubert *et al.* 1996) and thus OC survival would not be anticipated in such ancient bones. This appears to be confirmed by the low concentrations and high levels of racemisation of amino acids reported from some dinosaur bones (Bada *et al.* 1999).

The greatest potential for OC sequencing from samples of ancient bone was to support phylogenetics and attempts were made for bovines (Nielsen-Marsh *et al.* 2002) and hominids (Nielsen-Marsh *et al.* 2005). However, there has not yet been report of an improvement of the phylogenetic record over that already known from DNA, where the bos/bison difference is a single amino acid substitution and more informative ancient DNA was also reported from the same ancient sample, and only the highly conserved mid-region of OC was reported in the ancient Neanderthal specimen analysed. OC was also studied in archaeological material for other purposes than phylogenetic information as it was considered more suitable for obtaining accurate ^{14}C age estimates and $\delta^{13}C/\delta^{15}N$ values for palaeodietary reconstructions (Ajie and Kaplan 1991). Further to this, Ritz *et al.* (1996) attempted to use the aspartic acid racemisation of purified OC as an improved method for age at death determinations in forensic bone.

1.8 Degradation of Bone

1.8.1 Bone in the Burial Environment

Bone survives better than most other tissues due to the mutual stability induced by association of water-insoluble protein to thermostable mineral, constrained further by packaging arrangements. As this is a mutual relationship, once the degradation of one phase occurs, the other will likely soon follow. For example, once the mineral phase is damaged and altered, the organic phase will also begin to degrade. The term diagenesis covers a multitude of processes which occur in buried bone and are responsible for the destruction of the bone at a variety of structural levels, studies of which can also add information about the burial history of the bone. Bone degradation is considered to occur mainly by two processes; one mediated by microorganisms and fungi (Bell *et al.* 1996; Hackett 1981), and the other much slower process of chemical degradation. Regarding microbial attack, within a very short time of death, decay processes take over a body, with gut organisms invading tissues following the circulatory system. Their initial attack on the body is primarily the soft tissues, because these are more accessible and easier substrates for enzymatic degradation. Only secondarily does attack start to occur on the bones, usually by specialised organisms (such as fungi and bacteria) capable of degrading collagen (Garland 1987), where bone can be histologically altered by microbial attack within three months of being subject to decay (Bell *et al.* 1996).

Even with the inhibition of microbial degradation, both the organic biomolecules and the inorganic bioapatite mineral remain subject to chemical degradation. All living bones contain very small and thermodynamically unstable crystals of carbonate hydroxyapatite, hence crystallinity changes after burial are likely. Crystallinity is a concept that includes crystal size, crystal order, and the number of defects in the crystal. The mineral alterations that occur in the burial environment result in re-crystalisation into larger crystals via a process of Ostwald ripening. A number of different 'crystallinity indices' were proposed based either on X-ray diffraction peak shapes (Bartsiokas and Middleton 1992; Sillen 1989) or on the splitting of two peaks in the infra-red spectrum (Weiner and Bar-Yosef 1990). Collagen decay and loss is

the primary diagenetic change to the organic phase of bone which is almost certainly demineralised before it can be altered, and then undergoes a variety of hydrolysis reactions (Collins *et al.* 1995). Although a number of micro-organisms can produce collagenases, which are capable of hydrolysing collagen and, under laboratory conditions, degrade bone in a matter of days (Child and Pollard 1991), in general, mineralised bone collagen is considered not susceptible to enzymatic degradation. Therefore non-enzymatic hydrolysis, followed by loss of hydrolysed peptide fragments, must be a major pathway for collagen degradation and this has been modeled by Collins *et al.* (1995). Recent belief was that the thermal history of a sample (the integrated time-temperature history) is the key factor influencing survival (Smith *et al.* 2003). Although collagen yields between bones of the same thermal age may vary, the integrity of surviving collagen will theoretically be the same (Nielsen-Marsh *et al.* 2002).

1.8.2 Heated Bone

As the degradation of bone in the burial environment is heavily influenced by temperature, there are suggested parallels between the diagenesis of ancient bone and the degradation of heated bone (Roberts *et al.* 2002). For this reason, heating experiments on modern bone have been widely used to gain an understanding of ancient buried bone degradation under archaeological time-scales not possible to re-create in laboratory conditions. In both forms, archaeological and heated bone, different degrees of crystallinity can be observed which is seen to increase by a process of Ostwald ripening (Tuross *et al.* 1989; Weiner and Bar-Yosef 1990). The changes in crystallinity can be observed via a number of techniques, mostly X-ray diffraction, infra-red spectra, and microscopic analysis of the crystallites (Sillen 1989; Weiner and Bar-Yosef 1990). Roberts *et al.* (2002) observed that as bone is boiled, protein is lost (measured by nitrogen content), as well as the increase in crystallinity and porosity of the bone. Nielsen-Marsh & Hedges (2000a; Nielsen-Marsh and Hedges 2000b) observed that in bones of low nitrogen content, the crystallinity of the bones is higher; also supporting the proposal that cooking could be used as an analogue for the effects of long-term burial. Koon (2006: 2003) investigated the effects of heating bone on the structure of the collagen fibrils that dominate the protein fraction of bone and found collagen to collapse at the fibrillar level rather than molecule by molecule. Some studies argue that collagen is preferentially lost relative to NCPs (Masters 1987; Weiner and Bar-Yosef 1990). These NCPs are generally

more soluble but also much more acidic, which display a higher affinity for hydroxyapatite than collagen and hence persist in bone even after the collagen has been lost (Masters 1987).

Building upon models of bone diagenesis where the most influential factor of degradation is temperature (Collins *et al.* 1995), several authors have likened the degradation of bone in the burial environment to the taphonomy of cooked bone (Koon *et al.* 2003; Roberts *et al.* 2002). During the heating process, all components of bone, organic, inorganic and water are affected. Although the temperature, duration of heating, and method of heating (ie., wet, dry) may vary, the key changes in bone result from loss of water, loss of organic content via gelatinisation or combustion, and alteration of the mineral component (Koon 2006). Roberts *et al.* (2002) attempted to characterise the physico-chemical effects of boiling by measuring a number of 'diagenetic factors' on modern bone boiled for increasing numbers of hours, such as the infrared splitting factor (SF) using Fourier transform infrared spectroscopy (see Weiner and Bar-Yosef 1990), insoluble collagen extraction yield (following Collins *et al.* 1995), nitrogen analyses and histology (see Hedges and Millard 1995) and mercury intrusion porosimetry (HgIP). Their results indicated that the boiling process does mirror the diagenesis of archaeological bones inasmuch that similar processes occur, such as loss of collagen, increasing crystallinity and increasing porosity. Koon (2006) furthered this research by visualising the effects of boiling modern bone on the insoluble bone collagen fibrils using transmission electron microscopy (TEM) with comparisons to potentially cooked and uncooked archaeological bone. As bone is heated in water, the bone collagen is transformed by a process known as gelatinisation. In this denatured form, the collagen/gelatin will be leached out of the bone more as the boiling process continues. Collagen loss from bone is one of the simplest degradation processes to study because the phase transition (hydrolysis) from insoluble collagen to soluble (more biodegradable) gelatin impacts upon the physical properties of bone and the survival of bone proteins. Roberts *et al.* (2002) showed that bones had to be boiled for 81 hours before a significant loss of the organic content of bone was noticed. Bada *et al.* (1989) was able to detect the effects of boiling on bone based upon the extent of racemisation of the aspartic acid. Since cooking can be considered an analogue of burial diagenesis, the extraction and sequencing of degraded proteins from ancient burial environments could be applied to heat-treated modern industrial samples, such as the possibilities of detecting contamination MBM.

1.8.3 Roasted Bone

If a bone is roasted (or baked), once the water within the bone has boiled off, the temperatures are able to climb so that the combustion of organic matter can occur. This process deposits carbonised material on the bone surface (charring), indicating breakdown and removal of proteins and fats present in the bone (Koon 2006). Modification and loss to the organic fraction has been investigated by analyzing the elemental composition using carbon, hydrogen, and nitrogen (C, H, & N) concentrations (Nicholson 1998; Stiner *et al.* 1995) and carbon/nitrogen (C/N) ratios (Brain and Sillen 1988). At a macroscopic level, thermally-induced colour change, associated with the combustion of the organic matter, has been found to correlate to some degree with increasing temperature (Bennett *et al.* 1999; Nicholson 1993; Shipman *et al.* 1984).

At high temperatures, modification and decomposition of the mineral also occurs due to the disruption of the bonds holding the apatite crystal structure together. The result is the formation of larger, more uniform and thermostable crystals, which also causes a reduction in space between these components (Shipman *et al.* 1984). Impurities in the crystal structure are also removed (Wright and Schwarcz 1996) and pyrophosphate is introduced, ultimately leading to the formation of β-tricalciumphosphate, a homogenous mineral (McKinley 2000). By measuring the temperature of bone in insulated samples cooked at 180°C and 220°C for the duration of time needed to cook the insulating meat, Koon *et al.* (2003) found that it rarely reached 100°C by the time that the meat was cooked, supporting the notion that roasting does not heat flesh-insulated bone efficiently as the presence of moisture prevents temperatures from climbing above the boiling point of water.

1.8.4 Carcass Rendering and the Production of MBM

The production of animal feed is one of the most important activities in agriculture and about 120 million tons of feedstuffs are produced annually in the European Union (EC 2004). Meat and bone meal, a form of animal feed, was used to supply essential amino acids to lactating and fast-growing animals such as cattle and sheep.

In the food and agricultural industries, rendering refers to the crushing of animal by-products (fat, bones and internal organs), heating them to drive off the water content, and separating the residue into fat (often called 'tallow') and high-protein solids (often called 'greaves'). Tallow can be used for soap manufacture and for human consumption, whereas the greaves are used in fertiliser or animal feed. After further processing, the solid residue/greaves could be ground to produce MBM. Renderers deal almost exclusively with either red meat material such as cattle, sheep and pig, or poultry material. In 1986, the material processed by renderers mostly came from slaughter-houses and consisted of the parts of animals that were unsuitable for food or that people in the UK choose not to eat, such as; offal no longer of use (bladder, diaphragm and udder, intestines, kidneys, spleen, blood, stomach, heart, liver and lungs), the head, hooves, bone and tails, edible fat and any carcasses deemed as unfit. In addition, renderers received waste fat and bones from butcher's shops and from food factories.

Although there are several different rendering systems used for MBM, all can be classified as either batch or continuous. Batch rendering systems were used exclusively until the 1970s when the first continuous systems were introduced into the UK. By the mid 1980s, continuous systems were used to produce over 75% of the annual MBM, and the reduction in the use of batch systems was believed linked to the emergence of the BSE epidemic. In a typical continuous rendering system, a continuous supply of raw material was delivered into bunkers in the floor where it was transported to a crusher and pulverised. Following the crusher, it was transported to the cooker, which was heated at 100°C until the moisture was driven off in the form of steam and released through vents. Once this was done, the temperature would rise to about 140°C or more, which would cause the cell structure of the residue to break down and release fat as tallow (Fig. 1.12). The workings of the heat stage varied according to different rendering plants, the most extreme of which being held at maximum temperatures around 135°C to 145°C for around 30 minutes. Most renderers chose maximum temperatures below 140°C because at that temperature the vitamins and trace elements in the solids were not too much affected, but the solids sufficiently crisp to make the following grinding into easier (White *et al.* 2000).

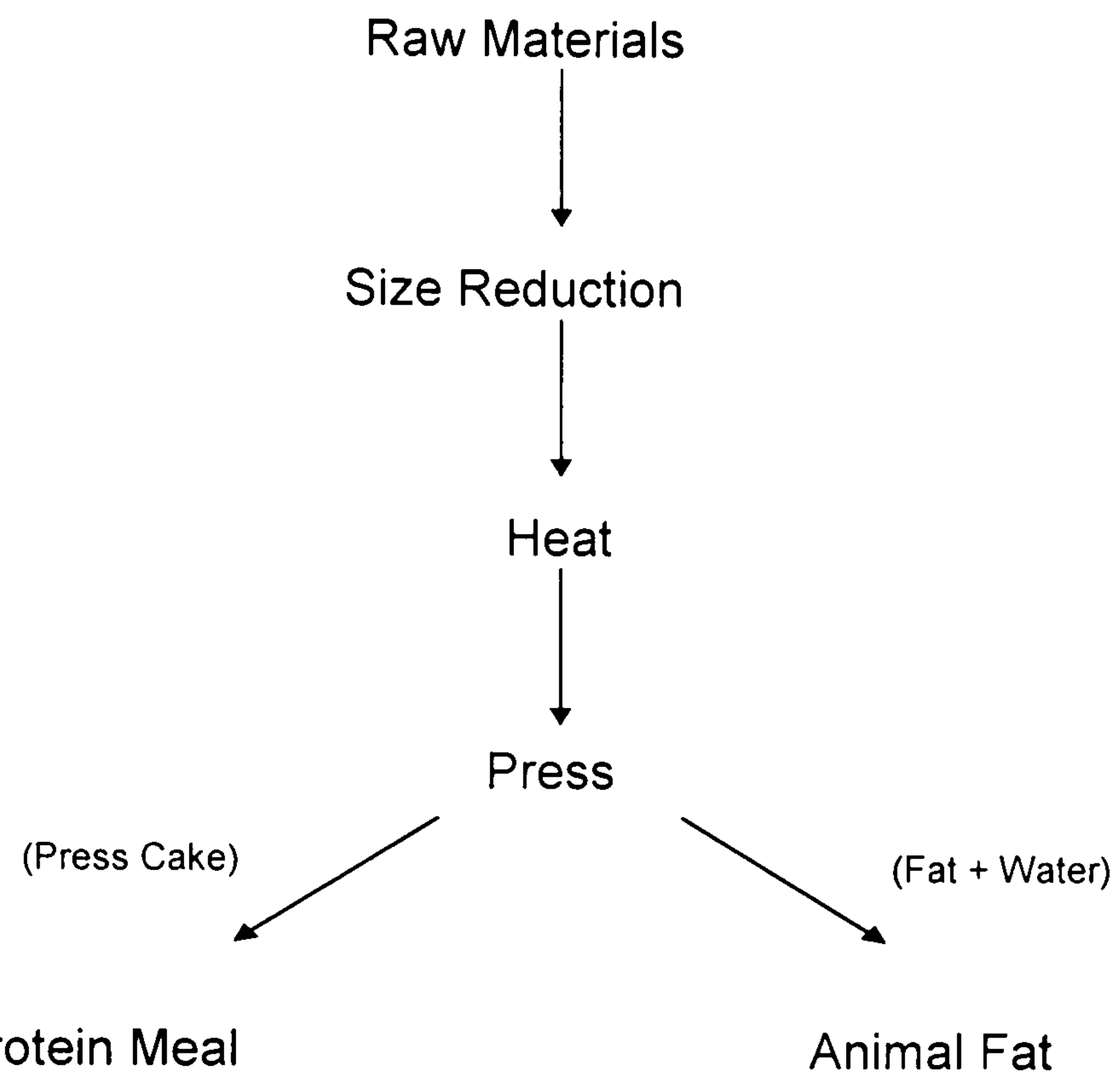


Figure 1.12 - Flow diagram of generalised processes of MBM production (Redrawn from Vermeulen et al. 2005).

In some European countries, the aim of rendering was to destroy dangerous pathogens, leaving it capable for uses in animal feed (Krenk 1991). In Germany, Belgium and Holland, legislation required that renderers were able to eliminate the spore-forming micro-organisms such as anthrax, and foot and mouth disease. In these countries, renderers were expected to process carcasses infected with those diseases and make the resultant meal suitable for use in animal feed. The heat treatment involves temperatures of around 130°C for a length of 20 minutes under 3 bar pressure (White *et al.* 2000). Other European countries, including the UK, notifiable diseases such as anthrax were dealt with on the farm by burial or incineration and renderers were not required to handle them. However, these other countries later adopted the legislation which made mandatory rendering systems capable of destroying dangerous pathogens a requirement (European Council Directive 90/667/EEC). In the UK, this Directive was implemented by the Animal By-Products Order of 1992. Essentially, the results indicated that the more heat-resistant strains of scrapie were not completely inactivated by conditions less rigorous than 133°C/3 bar pressure for a minimum of 20 minutes although all rendering processes investigated resulted in some degree of inactivation.

1.8.5 Biomolecule Survival in Degraded Bone

Although lipids and carbohydrates have been shown to survive in ancient bone for millions of years (Evershed *et al.* 1995; Logan *et al.* 1995), these two classes of biomolecules hold the least amount of information regarding properties of the organisms that produced them (Schweitzer 2004). All the genetic information needed to specify an organism is contained in molecules of DNA, which consists of a backbone of a five-carbon sugar covalently linked to a phosphate group and one of four nitrogenous bases. The specific order in which of these nitrogenous bases occur (adenine, guanine, cytosine and thymine) not only determines which proteins are produced and how they function, but also an evolutionary record of how species are related. However, DNA is more labile than most other classes of biomolecule and that which does survive in ancient bones is likely to be degraded or chemically altered (Paabo *et al.* 1989).

Proteins are possibly the most useful biomolcules to analyse in degraded material because they contain genetic information that records the evolutionary history of organisms, but are also more likely to survive in degraded fossil bone as well as thermally treated MBM. Proteins are thought more likely to survive because of their complex, multi-level structure, where inter- and intramolecular bonds that stabilize the molecule at each level must break for the protein to 'unfold' and expose backbone sites for peptide hydrolysis (Schweitzer 2004). This results in the internal residues of some proteins to be almost impervious to attack (Eglington and Logan 1991) and thus proteins are considered to have a much greater preservation potential than DNA. Condensation reactions in particular (Amadori rearrangments and/or Maillard reactions) can make the molecules increasingly insoluble and resistant to decay due to the additional formation of inter- and intramolecular cross-links (Schweitzer 2004). Protein diagenesis can also take several forms including the conversion of one amino acid to another, loss of functional groups condensation reactions, methylation and/or glycosylation. It was these formations that, until the applications of protein mass spectrometry, rendered such biomolecules difficult to analyse.

1.9 Analysis of Proteins and Peptides by Mass Spectrometry

Until the early 1990s, one of the most common methods of sequencing peptides was by a technique called Edman degradation, which relies on the identification of amino acids chemically cleaved in a stepwise fashion from the amino terminus of a peptide by reaction with phenylisothiocyanate and cleavage of the resulting phenylthiocarbamyl derivatives. However, this method failed when the peptide being analysed possessed an acetylated or otherwise blocked amino-terminus. Techniques using mass spectrometry (MS) overcome some of the problems associated with peptide sequencing by Edman degradation.

Mass spectrometry (MS) is an analytical technique in which molecules from within a test sample are converted to gaseous ions (i.e. become electrically charged) that are subsequently separated in a mass spectrometer according to their mass-to-charge ratio (m/z) and detected. Although early mass spectrometers required the sample to be in the gas phase (such as with Electron Ionisation (EI) and Chemical Ionisation (CI)), developments during the 1970s and 1980s in ionisation technologies allowed for the samples to be input as liquid solutions or solids (such as Plasma Desorption (PD), Fast Atom Bombardment (FAB) and Laser Desorption (LD)). Depending on the type of inlet and ionisation techniques used, the sample may already exist as ions in solution or it may be ionised in conjunction with its volatilisation or other methods in the ion source. 'Soft-ionisation' techniques are where the evaporation and ionisation of the molecular species into the gaseous phase are carried out without extensive fragmentation. Two of the most common ionisation methods currently used for the analysis of proteins and peptides are Matrix Assisted Laser Desorption/Ionisation (MALDI) first described by Karas and Hillenkamp (1988) and Tanaka *et al.* (1988), and Electrospray Ionisation (ESI) first described by Yamashite and Fenn (1984), both of which were used in the work described in this thesis.

Even though mass spectrometers can measure the m/z of intact proteins, such as in the analysis of OC by Ostrom *et al.* (2000), it is often peptides that are analysed in large-scale protein analyses (proteomics) because they are more soluble and the sensitivity of MS for proteins is much less than that for peptides. The endoprotease trypsin is one of the most commonly used enzymes because it is an aggressive and stable protease which very

specifically cleaves proteins on the carboxy-terminal side of arginine and lysine residues unless followed by a proline. This creates peptides in the preferred mass range for sequencing and with a basic residue at the carboxyl-terminus of the peptide ions which result in information-rich peptide-fragmentation spectra (Steen and Mann 2004). Although the enzymatic digest of a purified protein would result in a relatively simple analysis, it is often complex mixtures of proteins that are collectively digested and analysed by LC-MS. In order to improve the analysis of large numbers of peptides, several peptide separation techniques can be employed prior to MS analysis, the most common being liquid chromatography. The research decribed in this thesis used only reverse phase liquid chromatography on either a silica-based C18 solid phase extraction (SPE) column or a polystyrene divinylbenzene (PSDVB) monolith column, the latter of which has several advantages over conventional HPLC columns including greater porosity and permeability, and can be run at higher flow-rates so that less time is required per analysis (Lubda *et al.* 2001).

There are three main parts to a mass spectrometer; the ion source, the mass analyser and the ion detectors. There are four main types of mass analysers used in proteomics, particularly with MALDI and ESI; time-of-flight (TOF), quadrupole mass spectrometers, quadrupole-TOFs (qTOFs) and quadrupole ion traps. In order to obtain sequence information by analysing peptide fragment ions produced by collision induced dissociation (CID), two (or more) of these mass analysers are often placed in tandem. The most common tandem mass spectrometers in current use are the TOF-TOF, the tandem quadrupole and the quadrupole orthogonal TOF (Q-oTOF). The instruments used in the work described in this thesis were the Applied Biosystems 4700 Proteomics Analyser (MALDI-TOF-TOF) and the Applied Biosystems/MDS Sciex QSTAR Pulsar i (ESI-Q-oTOF), which will be described in more detail.

1.9.1 Matrix Assisted Laser Desorption/Ionisation Mass Spectrometry (MALDI-MS)

In MALDI-MS, analyte compounds embedded in a surplus of matrix are co-desorbed from the target plate upon laser excitation, and analysed by MS. The matrix is an organic compound, normally a weak organic acid such as a derivative of cinnamic acid or

dihydroxybenzoic acid that is mixed with the sample and spotted onto a target plate where sample and matrix co-crystallise on drying. As illustrated by Figure 1.13, focusing a laser onto the co-crystalised sample causes the excitation and vaporisation of the sample and the subsequent transition of solid-phase matrix and analyte ions into a plume of gas-phase matrix and analyte ions (Breuker *et al.* 1999; Knochenmuss and Zenobi 2003; Zenobi and Knochenmuss 1998). The Applied Biosystems 4700 mass spectrometer possesses capabilities for 'delayed extraction', which by the positioning and timing of additionally applied voltages allows a delay between ion formation and acceleration into the mass analyser to be controlled, improving the resolution of peaks in the spectra (Wiley and McLaren 1955). As the ions are accelerated by applying a high voltage pulse after a predetermined delay time, a potential gradient is applied by the use of a variable voltage and ground voltage that enables focusing of the ions, minimising the effect of the kinetic energy spread (Applied Biosystems 4700 Proteomics Analyzer Reference Manual, 2003).

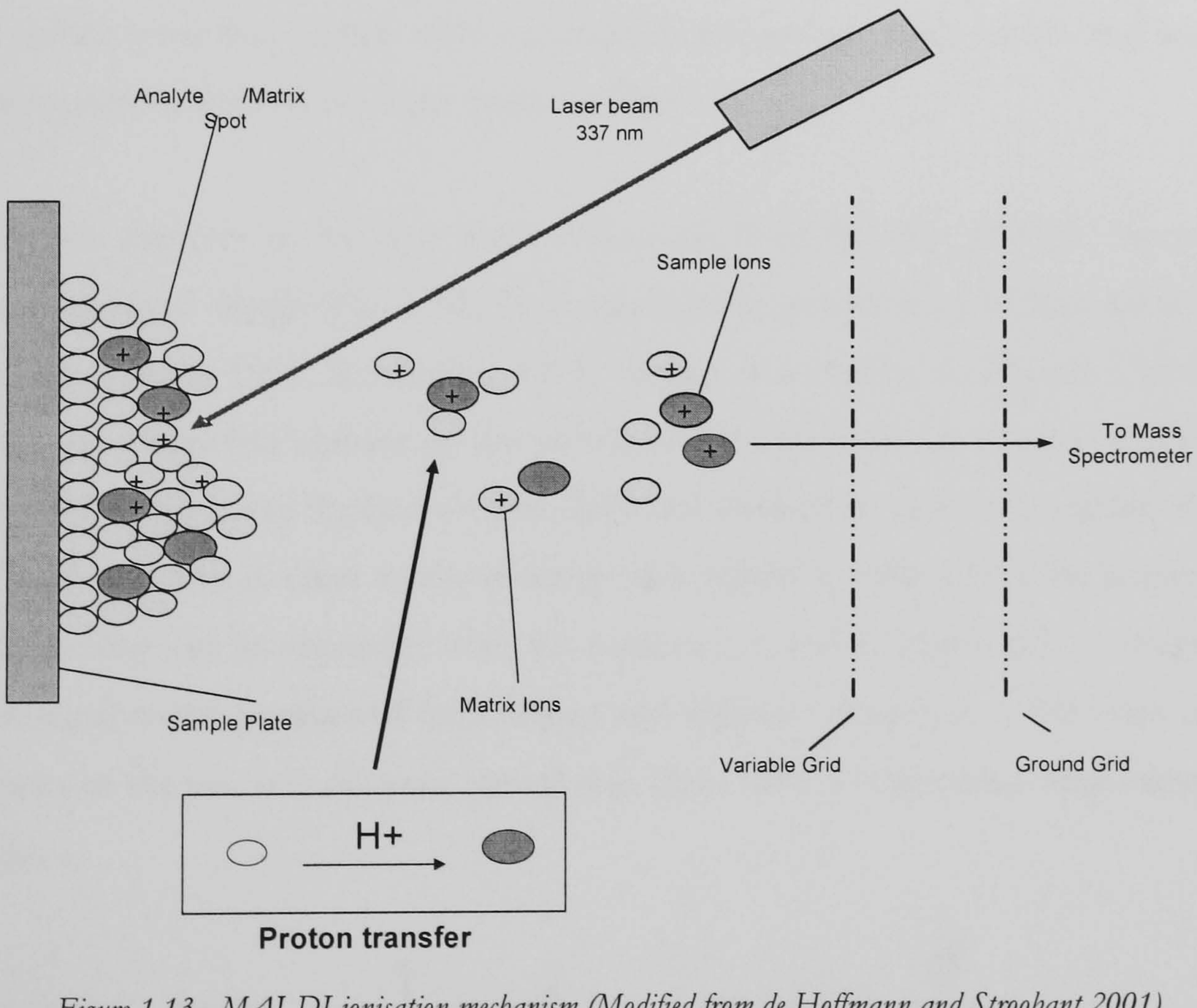


Figure 1.13 - MALDI ionisation mechanism (Modified from de Hoffmann and Stroobant 2001).

Dreisewerd (2003) describes three properties that the matrix should have to function adequately; to be able to isolate analyte molecules without aggregation, to absorb energy by electronic (or vibrational in the case of IR-MALDI) excitation, and to allow disintegration of

co-crystalised analytes without excessive heating and thus destruction/fragmentation of the analyte molecules. The matrix material (α-cyano-4-hydroxycinnamic acid throughout this thesis) is usually prepared in solutions of low pH which gives an abundance of protons available for the matrix molecules to become charged so that they can collide with and ionise analyte molecules that can subsequently be detected. Different models have been proposed to explain the desorption process of the matrix-sample material from the crystal surface; thermal desorption of individual molecules, surface layer-by-layer sublimation (Bencsura *et al.* 1997; Johnson 1994; Vertes *et al.* 1990; Vertes and Levine 1990), and/or volume ablation by either explosion or laser-induced pressure pulses (Dreisewerd 2003; Johnson 1994; Vertes *et al.* 1993; Vertes and Levine 1990; Zhigilei *et al.* 1997). Although there is no consensus as to how the sample molecules are ionised, several mechanisms for ion formation have been put forward (Karas *et al.* 2000; Knochenmuss and Zenobi 2003; Zenobi and Knochenmuss 1998). These are generally based around the concept of an expanding plume of primary ions (mostly matrix ions) that collide with and transfer protons to other matrix and analyte ions before being extracted into the mass analyser (Fig. 1.13).

A TOF mass analyser is the type most commonly coupled with MALDI because of its simple instrumental design (Fig. 1.14), high transmission properties and high-mass capability (Brown and Lennon 1995; Kaufmann 1995; Tanaka *et al.* 1988). As Figure 1.14 illustrates, TOF mass spectrometers operate on the principle that when a group of ions of differing m/z are subjected to the same applied electric field and allowed to drift in a region of constant electric field, the time it takes to travel across this region is related to their respective m/z. This relationship can be expressed with the equations 1 and 2, where K.E. = kinetic energy (which is equal to the product of their charge and applied voltage); m = the mass of the ion; v = velocity of the ion; d = distance travelled in flight tube, t = recorded time taken to travel flight tube.

$$1) \quad K.E. = \frac{1}{2} mv^2 \qquad 2) \quad v = \frac{d}{t}$$

Given that the applied energies are equal, the protonated molecules with higher mass are accelerated and travel slower and result in a higher recorded time of flight. Thus, as long as

the ions are accelerated with the same potential at a fixed time, they separate according to their m/z, the smallest protonated molecules hit the detector first and the largest hit the detector last (Fig. 1.14)(Brown and Lennon 1995; de Hoffmann and Stroobant 2001).

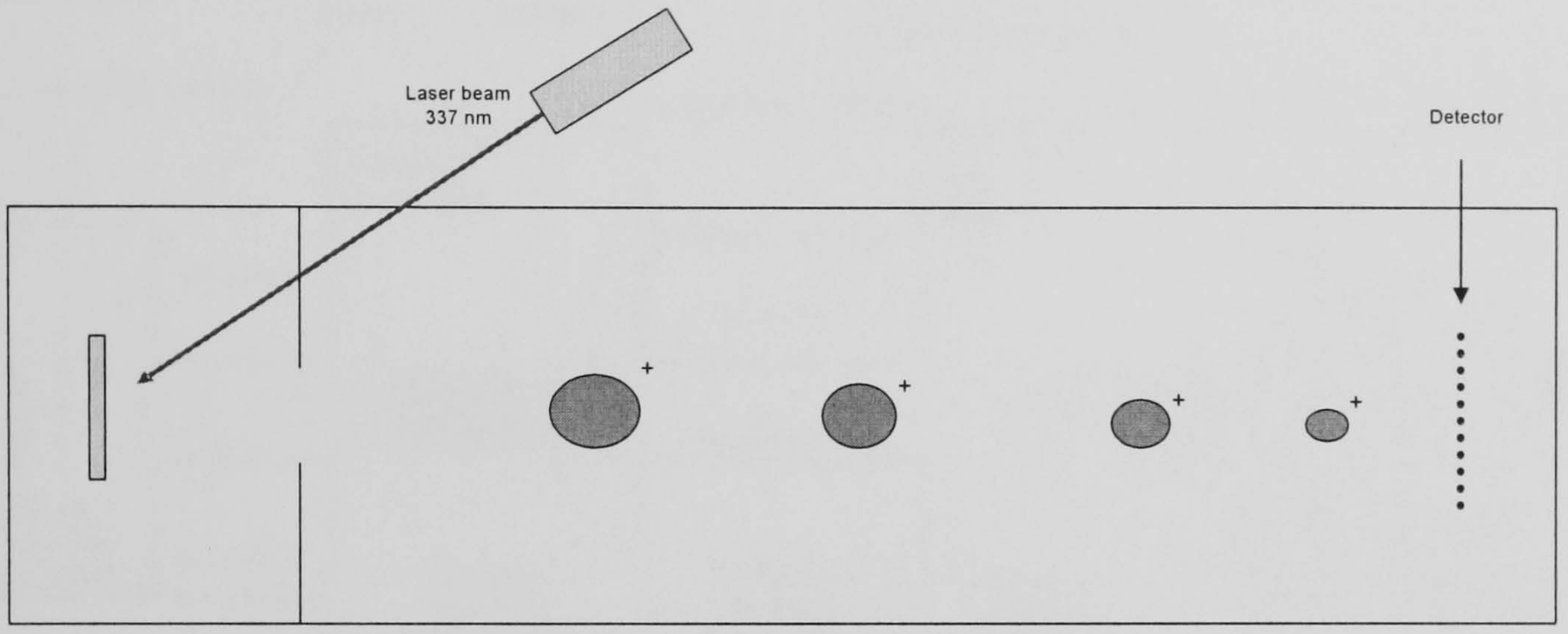


Figure 1.14 - Principle of 'Time-Of-Flight' (Redrawn from Mamyrin 2001).

Typically several hundred pulses of laser light are used, each of duration of a few nanoseconds, and are combined in the creation of each spectrum. During the transition of matrix and analyte ions into the gas-phase there are slight differences in the amount of energy that is acquired by similarly charged ions. This is due to the slight energy spread of the laser as well as the fact that when desorption occurs in a strong electrical field, energy is lost via collisions with other molecules within the plume, resulting in further dispersion of energy. These slight differences in energy of ions of the same m/z result in poorly resolved signals. To increase the resolution of TOF mass spectrometers, many scientists in the 1950s-60s tried to design ion sources with a smaller energy spread, but no substantial progress in this research was achieved (Mamyrin 2001). Instead, to compensate for the differences in energy obtained in the source of the instrument, the ions were decelerated by means of a reflectron ion mirror (Fig. 1.15) which redirects fragments that still bear a charge towards another detector. The device has a gradient electric field and the depth to which ions penetrate this field, before reversal of direction of travel, depends upon their energy. The more energetic ions that have slightly faster velocities penetrate further into the ion mirror and hence are delayed relative to the less energetic ions, thus focusing the flight times, while leaving neutral fragments unaffected (Mamyrin *et al.* 1973). This allows compensation for similarly charged ions having slightly different overall energies from ionisation and thus

improves the resolution (recorded as full width at half of maximum (FWHM)) of the peaks observed.

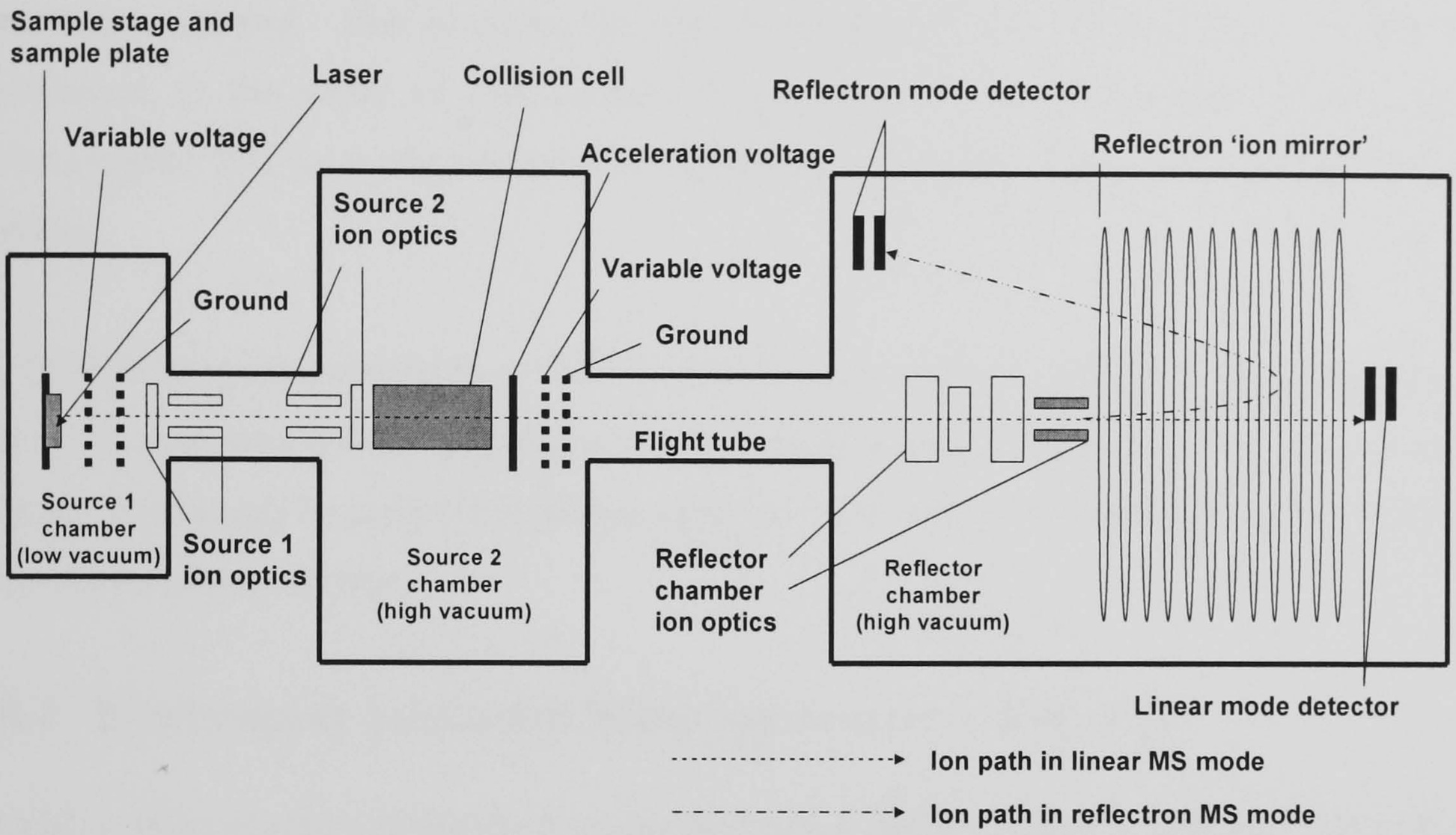


Figure 1.15 - Schematic of the Applied Biosystems 4700 Proteomics Analyser (Redrawn from Applied Biosystems 4700 Proteomics Analyzer Reference Guide, Applied Biosystems, 2003).

The final part of the mass spectrometer is the ion detector, which records the charge induced or current produced when an ion hits its surface and this electrical signal is sent to a time-to-digital converter which records the time of detection. As Figure 1.15 shows, MALDI-TOF mass spectrometers often have two detectors, the linear mode detector and the reflectron mode detector, which is positioned to detect ions that have passed through the reflectron. The linear mode detector can be used when studying large analytes because metastable fragmentation of ions which occurs post-source (also known as Post-Source Decay or PSD) is not apparent in the resulting spectra. This is because precursor and metastable decay ions, having been initially given a specific amount of kinetic energy before PSD, move with the same velocity in the flight tube (under vacuum) and arrive simultaneously at the detector (Kaufmann 1995). Linear systems combine excellent sensitivity with demonstrated upper mass limits in excess of 200,000 Da (Karas *et al.* 1989) and provide adequate mass resolution for many protein-analysis applications (Brown and Lennon 1995). Reflectron systems that use an ion mirror to correct for the initial ion velocity spread of MALDI-generated ions obtain much better resolution, but lose information from

not being able to transmit uncharged PSD fragments which occur after ion acceleration but prior to reflection in the ion mirror. The mode used often depends on the type of information desired. For example, the loss in sensitivity with a reflectron TOF-MS is detrimental to the study of some intact proteins, such as in the analysis of OC, (as archaeological OC is rarely observed in reflectron TOF-MS) where linear TOF-MS is required.

In TOF-TOF mass spectrometry, where two TOF mass analysers are placed in tandem (see Fig. 1.15), ions with specific m/z values are selected by a timed-ion selector and the selected ions are fragmented by interaction with residual gas molecules in the flight path (see Tandem Mass Spectrometry section).

1.9.2 Electrospray Ionisation Mass Spectrometry (ESI-MS)

In ESI, a dilute solution of analyte is passed through a capillary, often at very low flow rates (20-40 nL/min), and a high voltage is applied to the capillary, electrically charging the analyte solution (Fig. 1.16). The charged analyte solution in the nozzle becomes unstable as it is forced to hold more and more charge until it reaches a critical point at which it cannot hold any more electrical charge. At the tip of the nozzle, the liquid then blows apart into a cloud of tiny, highly charged droplets (Fig. 1.16), the aerosol being at least partially produced by a process involving the formation of what is known as a Taylor cone. When the solution that comprises the Taylor cone reaches the Rayleigh limit, the point at which Coulombic repulsion of the surface charge is equal to that of the surface tension of the solution (Taflin *et al.* 1989), droplets that contain an excess of positive or negative charge detach from its tip (Cech and Enke 2001). These highly charged droplets move towards a potential surface that is opposite in charge to their own, and shrink as solvent molecules evaporate from their surface. Because it is difficult for charge to evaporate, the distance between charges within the droplet decreases (Dole *et al.* 1968). When the 'Rayleigh limit' is again reached in each of these droplets, the process of 'Coulombic explosion', or fission, is thought to repeat itself either releasing ions from droplet surfaces (Iribarne and Thomson 1976) or continuing until droplets containing only a single ion are formed (Fig. 1.16) (Fenn 1993).

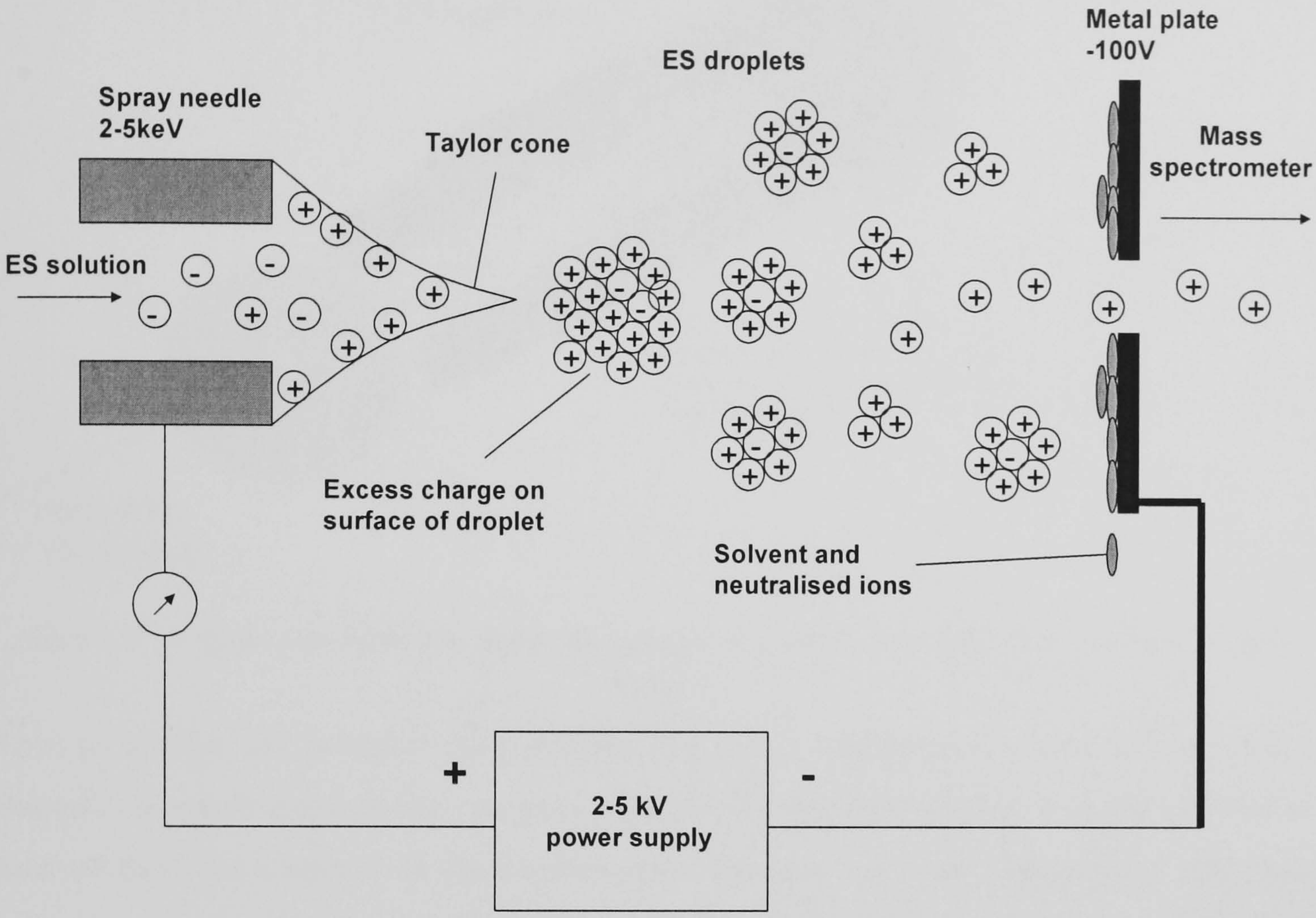


Figure 1.16 - Schematic of the electrospray ionisation process (Redrawn from Cech and Enke 2001).

The ions produced in the source of the instrument enter the mass analyser and are attracted to the detector or next stage of mass analysis by a potential difference. As well as TOF mass analysers, quadrupoles are often used for ESI instruments, or a combination of both. As shown by the schematic in Figure 1.17, the quadrupole consists of four parallel metal rods, where a direct current and an opposite alternating current is applied to each opposing rod pair. The direct current causes the ions to move in one particular direction whereas the alternating voltage causes the ions to spiral down the quadrupole. The applied voltages can be used to manipulate the flight path that the ions take through the quadrupole, which is dependent upon the ions' m/z (Dawson 1995; Miller and Denton 1986).

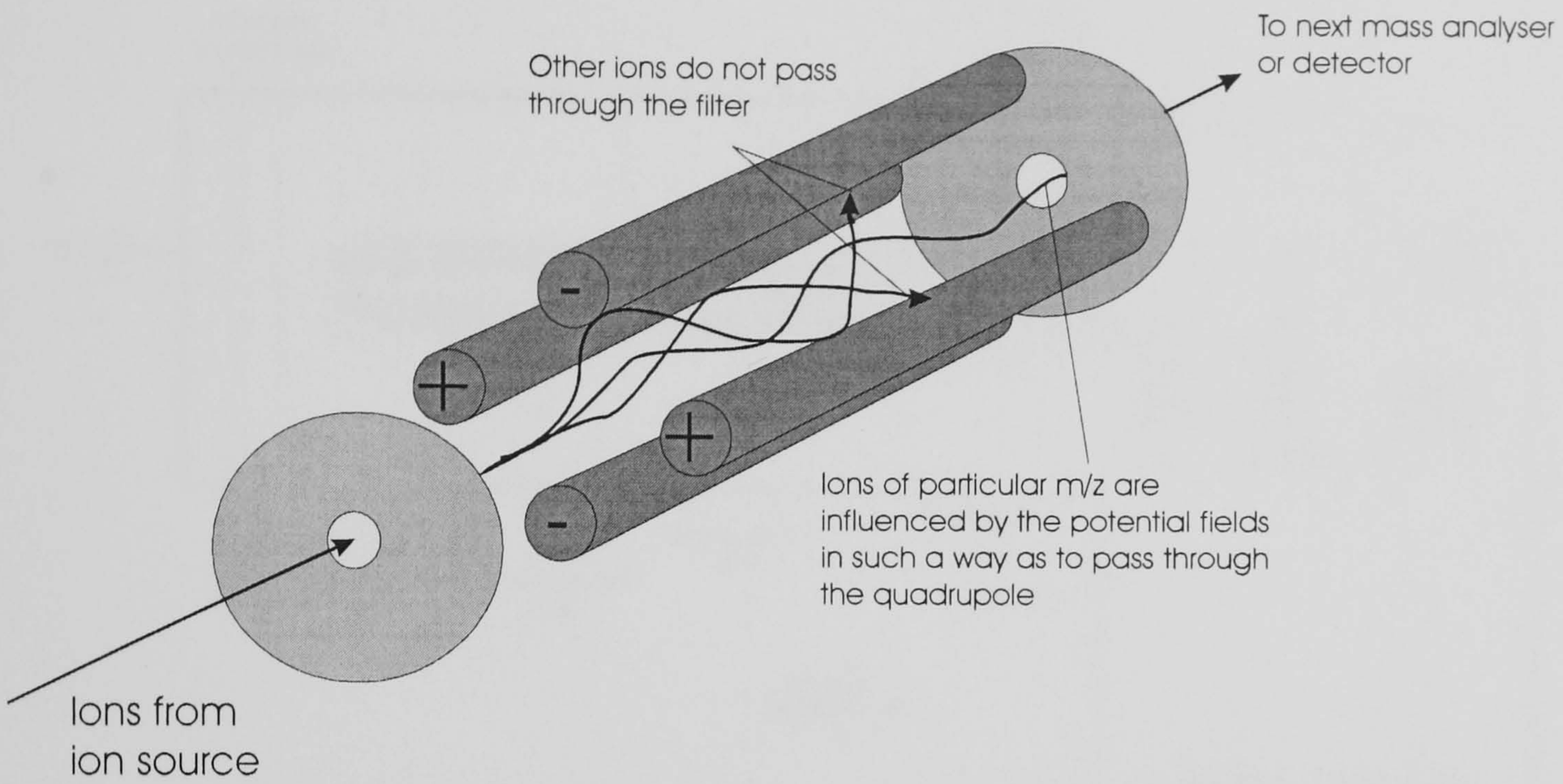


Figure 1.17 - Diagram of quadrupole mass analyser (Modified from de Hoffmann and Stroobant 2001; Miller and Denton 1986).

Thus by altering the potential fields applied (Fig. 1.17), only ions of a particular m/z pass through the quadrupoles at any one time, whereas all other ions do not. A stream of ions is focused onto the detector of the instrument to produce the mass spectrum as described earlier. In the Q-oTOF instrument (Fig. 1.18), the quadrupole mass analyser (Q1 in Fig. 1.18) simply guides the ions into the TOF in MS mode, and is used as a mass analyer only in MS/MS mode. The additional quadrupole (q0 in Fig. 1.18) is included to provide collisional cooling and focussing of the ions. With the addition of a collision cell hexapole (q2 in Fig. 1.18), this instrument (QStar) contains three sets of quadrupoles as well as a TOF mass analyser (see Fig. 1.18). The orthogonally arranged TOF is used to measure the 'time-of-flight' of ions in both MS and MS/MS modes.

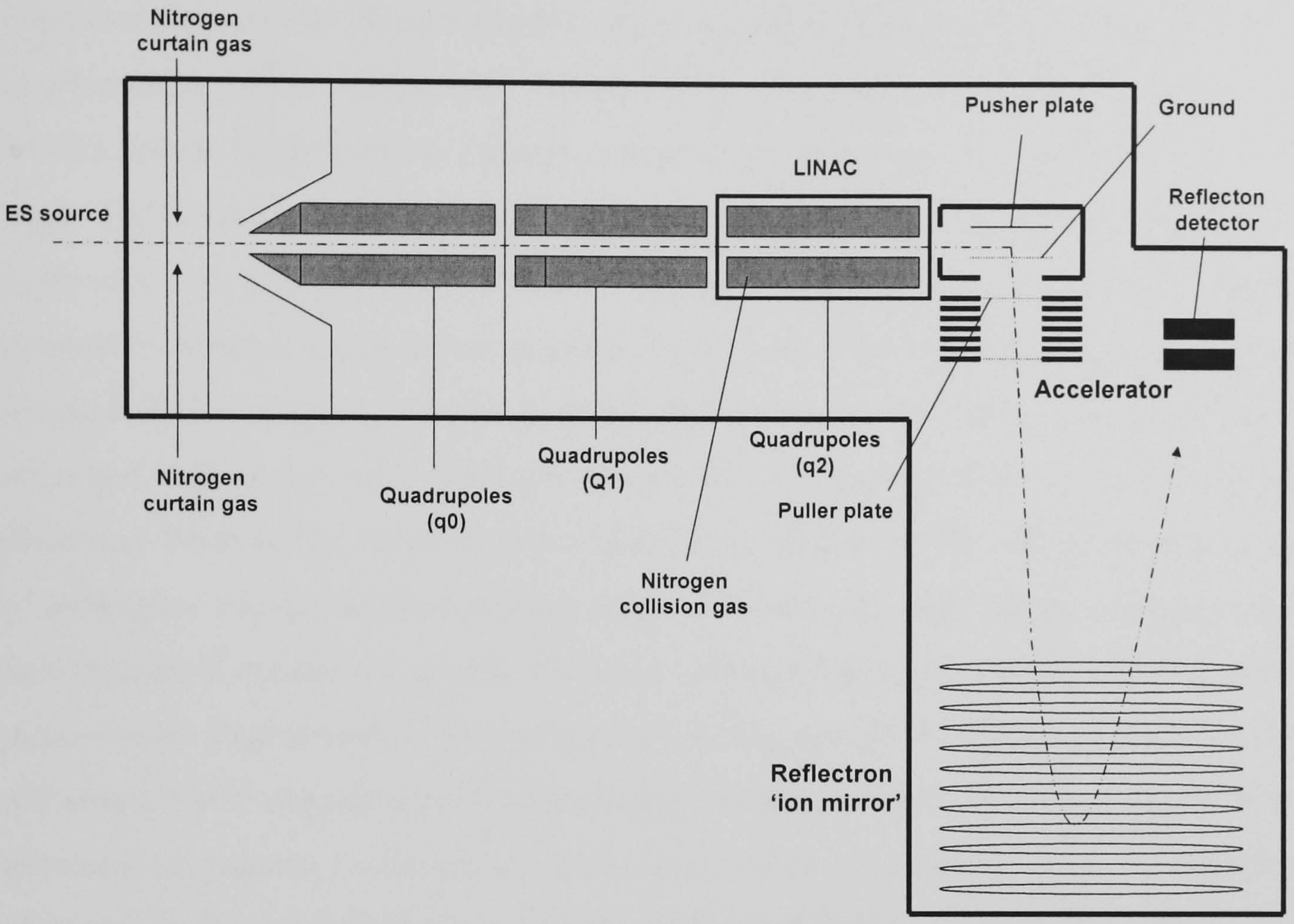


Figure 1.18 - Schematic of the Applied Biosystems API QSTAR Pulsar i Q-oTOF system (Modified from Instrument Manual published by Applied Biosystems).

A linear acceleration collision cell (LINAC) is used to obtain fragment ions for MS/MS analysis, in which the ions collide with gas molecules in the cell (nitrogen) and obtain the internal energy needed to fragment. The ions are then accelerated by a 'pusher' orthogonally into the TOF mass analyser.

1.9.3 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) is a way of measuring fragment ions, most commonly generated using CID, for sequence interpretation; protonated molecules may be fragmented by increasing their internal energies so that they obtain sufficient energy to break internal bonds. This additional energy is most commonly transfered by collisions with a collision gas (CID).

The CID process has been considered as two steps, the first corresponds to the excitation (or 'activation') of the ion by collision with another atom, and the second being the

subsequent dissociation (or unimolecular decomposition) of the activated ion (de Hoffmann and Stroobant 2001; Shukla and Futrell 2000). The most common methods for the activation of the target ions are through colliding with either low-energy (1-100 eV) or high-energy (keV) accelerated ions with gas molecules. Following the principles of the conservation of momentum, the kinetic energy of the moving ion cannot entirely be converted to internal energy of the colliding atom, and only a fraction of the kinetic energy is converted. Different precursor ion fragmentation pathways are believed to occur with low-energy and high-energy CID; for low-energy collisions, the excitation is considered mostly 'vibrational' because the collision interaction time is similar to the bond's vibration period, and undergoes charge-directed fragmentation, whereas for high-energy collisions the ion fragmentation is considered mostly electronic (shorter interaction times) and dominated by charge-remote fragmentation (not influenced by the site of protonation) with a relatively small amount of charge-directed fragmentation (Shukla and Futrell 2000). As the position of fragmentation requires protonation at the specific site in order to trigger charge-directed fragmentation (the protonation has a localised effect of altering bond strengths (Dongre *et al.* 1996; Somogyi *et al.* 1994)), the presence of ionised peptides protonated at various different sites is thought due to intramolecular proton transfer rather than being produced in the ionisation process. Dongre *et al.* (1996) modelled the low-energy dissociation pathways by a proton transfer model which assumes that protonation occurs at the basic site and that activation of the ion allows the proton to move to less basic sites along the peptide chain. The high-energy CID fragment ion spectra tend to contain much more information due to the greater amount of direct bond cleavage and fragmentation of side chains due to charge-remote fragmentation, which can result in the production of a more rich spectrum (de Hoffmann and Stroobant 2001). However, energy acquired from high-energy collisions can also be redistributed as vibrational energy, which may also result in bond cleavage.

In the most commonly used version of tandem MS, product ion scanning, following the fragmentation of selected ions in a collision cell, the fragment ions are introduced to a second stage of MS for analysis and directed towards a detector to obtain a representative mass spectrum. Whereas in MALDI-MS/MS, tryptic peptides are often observed as singly charged ions ($(M+H)^+$), in ES, tryptic peptides are often doubly or triply protonated ($(M+2H)^{2+}$) but can have more protons (a higher charge state) if they contain basic residues

such as proline, histidine, arginine and lysine, which can retain an additional proton. This can result in increased fragmentation and thus improved MS/MS spectra.

The nomenclature for labelling the fragment ions was proposed by Roepstorff and Fohlmann (1984) (later modified by Biemann 1988) which is based on the fact that cleavage can occur in either of the three types of bond in the peptide backbone: Cα-C, C-N, or N-Cα, further differentiated depending on which terminus retains the positive charge. When the amino-terminus retains the charge these are labelled a_n, b_n and c_n, whereas when the carboxy-terminus retains the charge they are labelled x_n, y_n and z_n respectively (where n is the number of amino acids). The difference in m/z of neighbouring fragment ions of the same type can allow for the determination of sequence (except between the isomers L and I, and the isobars Q and K). It is the y-type and b-type ions that are most frequently observed, both of which can lose water (18.011 Da) or ammonia (17.027 Da). The b-type ions can also lose CO (27.995 Da) to give the so-called a-type ion which tends to most commonly occur for the lower mass fragments containing a few of the amino-terminal residues. Other fragments that may occur are the so-called internal fragment ions that contain neither the carboxy- nor amino-terminus of the peptide, which can also lose water, ammonia, or CO, and the immonium ions, which are internal fragments containing individual amino acid residue ions and considered indicative of the amino acids present in the sequence (de Hoffmann and Stroobant 2001).

With the increase in applications of high-throughput proteomics, the vast quantities of resulting spectra, either as MS or MS/MS peak lists, are searched against protein sequence databases via search engines, such as Mascot. The most common searches are either Peptide Mass Fingerprints (PMFs) using mass spectra obtained from enzymatic digests, or MS/MS ion searches, using the MS/MS spectra of particular precursors. The MS/MS ion searches accept data in the form of peak lists containing mass and intensity values. Data from a single peptide may be searched, but the analytical approach generates more information when analysing an LC-MS/MS run containing data from multiple peptides from a mixture of proteins. Obtaining matches to a number of peptides from a single protein provides a very high level of confidence that the match is correct and that parts of the identified proteins are present in the sample. The two most common protein databases used are the NCBInr (a

comprehensive, non-redundant protein database) and Uniprot (formerly SwissProt, a high quality, curated protein database). The research in this thesis used a method of database searching called 'probability-based matching' using the Mascot search engine. This involves calculating the m/z values for all peptides derived from the proteins in the database and matching them to the observed fragments in a top-down fashion starting with the most intense b- and y- ions. The 'identification score' is calculated from the negative logarithm (multiplied by 10) of the probability that the number of fragment matches is random (Matrix Science http://www.matrixscience.com). However, the protein databases are limited to a particular species and protein sequences that differ greatly from those present are not identified via Mascot searches. Thus, *de novo* sequencing, which is the practice of manually interpreting the MS/MS spectra in order to determine sequence, is often carried out in addition to database searching.

1.9.4 Quality Control

It is important to note that poor reproducibility may arise between LC-MS analyses of the same sample due to LC-plate spotting issues. These may arise when matrix crystralises within or near to the spotting needle causing the needle to clog and thus not spot correctly. This could be avoided with appropriate pre-analysis instrument treatment and particularly with the use of fresh matrix every sample. However, one alternative may be to introduce a method of screening LC-MS analyses to eliminate runs that are below acceptable standards. This would include selecting criteria, such as a particular number of peaks in the first ten spots that must have a signal:noise ratio above a stated value. However, this could only be employed routinely on a particular known sample, such as 'collagen', and may not be applicable to samples of different protein composition.

The following 10 research chapters explore various applications of soft-ionisation mass spectrometry to the study of proteins in degraded material. Chapters 2 and 3 describe the use of MALDI-TOF instrumentation for the analysis of intact OC in ancient bone and MBM respectively. Chapters 5-10 focus on simple methods of peptide isolation and identification by de novo sequencing interpretation. Chapters 4, 9 and 10 involve the use of LC-MS approaches.

2 Comparing the survival of osteocalcin and mtDNA in archaeological bone from four European sites

Reprinted from *Journal of Archaeological Science, Volume 35, 27 November 2007, Pages 1756-1764,*
 Corrected Proof, Available online 16 January 2008.

I proposed the protein analysis of samples previously analysed for ancient DNA, designed these analyses, carried out these analyses and wrote up the research of this chapter.

ELSEVIER

Journal of Archaeological Science 35 (2008) 1756–1764

Journal of Archaeological SCIENCE

http://www.elsevier.com/locate/jas

Comparing the survival of osteocalcin and mtDNA in archaeological bone from four European sites

Michael Buckley [a,*], Cecilia Anderung [b], Kirsty Penkman [c], Brian J. Raney [d], Anders Götherström [e], Jane Thomas-Oates [c], Matthew J. Collins [f]

[a] *BioArch, S Block, Department of Biology, University of York, York, Yorkshire Y010 5YW, UK*
[b] *Department of Biology, University College London, NW1 2HE, UK*
[c] *BioArch, S Block, Department of Chemistry, University of York, York, Yorkshire Y010 5DD, UK*
[d] *Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA 95064, USA*
[e] *Department of Evolutionary Biology, Norbyvägen 18D, S-752 36 Uppsala*
[f] *BioArch, S Block, Department of Archaeology, University of York, York, Yorkshire Y01 7EP, UK*

Received 19 July 2007; received in revised form 14 November 2007; accepted 27 November 2007

Abstract

The small mineral-binding bone protein, osteocalcin, has been applied in a number of studies on ancient bone due to predictions of its long-term stability. However, the intact protein has not been shown to survive in ancient bone devoid of DNA, which is a much more phylogenetically informative biomolecule. In this investigation, the survival of osteocalcin is directly compared to the amplification of mtDNA in a set of 34 archaeological samples from four sites throughout Europe. We also present unpublished osteocalcin sequences of seven mammalian species in addition to the 19 published sequences to highlight phylogenetic limitations of this protein. The results indicate that the intact osteocalcin molecule survives less in archaeological samples than mtDNA and is more subject to the temperature of the archaeological site. Amino acid analyses show the persistence of the dominant protein collagen in samples that failed both osteocalcin and mtDNA analyses. The implications these findings present for biomolecular species identification in archaeological and palaeontological material are that, although proteins do survive beyond ancient DNA, osteocalcin does not appear to be the most ideal target.

Keywords: Osteocalcin; Protein mass spectrometry; ZooMS; Biomolecular archaeology; African elephant; Hedgehog; Pig-tailed macaque; Bushbaby; Opossum; Tree shrew; Lesser hedgehog; Lerna; Zauschwitz; Asine; Portalón; Collagen; Ancient DNA

1. Introduction

One decade ago, the hope that original DNA and protein would be widely persistent in geological samples fuelled biomolecular studies of fossil material. Ancient DNA (aDNA) led the way with spectacular claims of sequences being retrieved from tissues millions of years old, including Cretaceous dinosaur bones (Woodward et al., 1994). However, one by one, the claims for exceptional preservation were questioned and the field of aDNA research became focussed on analyses of predominately archaeological and permafrost material (for review, see Willerslev and Cooper, 2005). Following the notion that proteins survive for longer periods of time than DNA (Nielsen-Marsh, 2002; Curry, 1988), several proteins such as albumin and osteocalcin (OC), were investigated in fossil material as potential alternatives for obtaining phylogenetic information (Shoshani et al., 1985; Tuross and Stathoplos, 1993). Collagen, the most abundant biomolecule in bone (by weight), although widely used for radiocarbon dating and stable isotope analyses (Jahren et al., 1998; Macko et al., 1999; Richards et al., 2000), has received less attention, due to its highly repetitive amino acid sequence; this may change following claims for survival in Pleistocene mammals and a dinosaur (Asara et al., 2007; Schweitzer et al., 2002).

* Corresponding author. Tel.: +44 7813667779.
E-mail address: mbuckley82@gmail.com (M. Buckley).

0305-4403/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jas.2007.11.022

Osteocalcin was an initial focus of study because it is the most abundant non-collagenous protein (NCP) in bone and stabilised due to the presence of γ-carboxyglutamic acid (Gla) residues, allowing the protein to bind to bone mineral (Hauschka, 1977, 1980; Collins et al., 1998, 2000; Huq et al., 1990). Immunological and amino acid analysis suggested preservation of OC in fossil bones of extraordinary antiquity (Ulrich et al., 1987: Muyzer et al., 1992). Osteocalcin has a number of conserved domains, including the Gla rich calcium binding region within the central part of the primary sequence, and a highly conserved cluster of aromatic amino acids at its C-terminus. The species-specific sequence lies predominantly in the flexible N-terminus of the molecule, which is a putative collagen binding domain (Prigodich and Vesely, 1997) and the first to fragment under heating (at Asp_{14}-Pro_{15}) (McNulty et al., 2002). Initial apparent successes at immunologically detecting OC in fossil bone (Ulrich et al., 1987; Muyzer et al., 1992) were later met with scepticism on the grounds that immunological methods are difficult to apply to fossil samples and, when pushed to their limits, are prone to yielding false positives (Collins et al., 1991, 1992; Montgelard et al., 1997). Further studies looking at OC in archaeological bone suggested that the mineral apatite itself is not particularly stable and hence gives little protection (Ajie et al., 1992; Smith et al., 2005). Osteocalcin was an ideal target for the first forays into protein mass spectrometry due to its small size (<50 residues), abundance and relative ease of extraction (Ostrom et al., 2000, 2006; Nielsen-Marsh et al., 2002, 2005). However, despite earlier expectations, ancient samples shown to yield *intact* OC for phylogenetic purposes have also yielded the more informative fragments of aDNA (Nielsen-Marsh et al., 2002), and the only published attempt at obtaining OC from bone samples that did not contain amplifiable aDNA did not yield *intact* OC (Nielsen-Marsh et al., 2005). Further to this, recent investigations comparing the extent of OC survival with other measures of preservation (such as histological alteration, crystallinity and collagen yields (Smith et al., 2005)) show similar limitations as for aDNA survival (Gotherstrom et al., 2002). These observed similarities are probably because DNA is stabilised by mineral sorption (Gotherstrom et al., 2002), as is OC (Hauschka, 1986). However, no studies have directly compared DNA survival to OC survival in archaeological bone and investigations into using OC in palaeontological material still continue (Ostrom et al., 2006; Nielsen-Marsh et al., 2005; Schmidt-Schultz and Schultz, 2004).

This study aimed to directly compare the survival of aDNA with that of *intact* OC, which we use as a rapid species identification method (zooarchaeology by mass spectrometry (ZooMS)). The concept behind ZooMS is that the mass of the intact protonated OC molecule, determined by matrix assisted laser desorption ionisation time of flight mass spectrometry (MALDI TOF MS) is sufficiently diagnostic to allow species-specific protein identification (Table 1) and hence bone assignment (e.g. to discriminate cow from bison; Nielsen-Marsh et al., 2002). One application of this method was the re-identification of an osteologically-assigned 'human' sample, determined to be from a vegan from stable isotope analyses, as originating from horse (Jay and Richards, 2006). In order to explore the limited survival of OC further, the 'simple' intact OC mass measurement assay was compared with a more involved DNA analysis in cattle bones from four European archaeological sites.

2. Materials and methods

This study focuses on 34 samples of cattle bone from four European archaeological sites (Fig. 1). Ten were obtained from Zauschwitz, Saxony, a region in Germany between the cities of Dresden and Leipzig with an estimated mean annual temperature (MAT) of 8.5 °C, and an effective burial temperature (T_{eff}) (Smith et al., 2003) of 12.5 °C. Eleven samples were taken from the Portalón cave, 15 km east of Burgos in northern Spain (MAT 9.5 °C; T_{eff} also 9.5 °C, as cave samples lack seasonal fluctuations in temperature); four from Lerna on the western side of the Argolic Gulf, eastern Peloponnese, Greece; and nine from Asine, from the northern side of the Argolic Gulf (both estimated having MAT 18 °C, T_{eff} 21 °C). These were analysed for DNA at Uppsala (CA) and for OC and amino acid analysis in York (MB).

2.1. DNA extraction, sequencing, and analysis

Mitochondrial DNA was extracted and purified from pulverised archaeological cattle bone with hybridisation and magnetic bead separation following a modified method of Anderung et al. (2005) where approximately 70 mg of bone powder was incubated at 37 °C with 100 μg of Proteinase K in the presence of 2 M urea with constant agitation for 16 h. Samples were then centrifuged for 5 min at 300 × *g*, and the supernatant discarded. One millilitre of extraction buffer (0.2 M K_2HPO_4, pH 7.0) was added to the pellet and mixed and the DNA concentrated using centrifugation and ultrafiltration. DNA was extracted using biotinylated probes and amplified as described by Anderung et al. (2005) where a 272 base pair (bp) fragment in the control region was amplified in three overlapping fragments (139, 157 and 176 bp) in duplicate, with the addition of uracil *N*-glycosylase (UNG) to control for post-mortem alterations common in aDNA. PCR was carried out using 9 μl of extract, three units of HotStarTaq DNA polymerase (Qiagen, Valencia, CA), ×1 Qiagen PCR buffer (2.5 μl of a ×10 solution eventually made up to 25 μl), 2.5 mM $MgCl_2$, 200 μM each dNTP, and 0.2 μM each primer in a total volume of 25 μl following a profile as in Anderung et al. (2005). Sequencing was performed on a MegaBACE™ 96 capillary system using the DYEnamic™ Terminator Cycle sequence kit, after unincorporated primers and nucleotide were removed using ExoSAP-IT® (USB, Cleveland, OH). For every batch of four samples, two extraction blanks were included and each sample was extracted at least twice. Five of the eight recommended criteria to monitor contamination (Cooper and Poinar, 2000) were used: control amplifications, appropriate molecular behaviour, reproducibility, independent replication and biochemical preservation (amino acid analyses) as in Anderung et al. (2005).

Table 1
Amino acid sequences of all known mammalian osteocalcin sequences aligned to give maximum homology

Species	Accession Number	Osteocalcin Sequence
Bovine	P02820	YLDHWLGAPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGPV--
Steppe bison	P83489	YLDHGLGAPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGPV--
Sheep	Q1ZZ82	YLDPGLGAPAPYPDPLEPRREVCELNPDCDELADHIGFQEAYRRFYGPV--
Goat	P0C225	YLDPGLGAPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGIA--
Pig	Q8HYY9	YLDHGLGAPAPYPDPLEPRREVCELNPDCDELADHIGFQEAYRRFYGIA--
Horse	UCSC[a]	YLDHWLGAPAPYPDPLEPRREVCELNPDCDELADHIGFQEAYRRFYGTA--
African elephant	UCSC	YLDQVLGAPAPYPDPLELKKEVCELNPDCDELADHIGFHEAYRRFYGTV--
Dog	P81455	YLDSGLGAPVPYPDPLEPKREVCELNPNCDELADHIGFQEAYQRFYGPV--
Cat	P02821	YLAPGLGAPAPYPDPLEPKREICELNPDCDELADHIGFQDAYRRFYGTV--
Mouse	P04641	YL----GASVPSPDPLEPTREQCELNPACDELSDQYGLKTAYKRIYGITI-
Rat	P04640	YLNNGLGAPAPYPDPLEPHREVCELNPNCDELADHIGFQDAYKRIYGTTV-
Rabbit	P39056	QLINGQGAPAPYPDPLEPKREVCELNPDCDELADQVGLQDAYQRFYGPV--
Hedgehog	UCSC	YLAQGPGAPAPYPDPLEPKREVCELNADCDELADLIGFQEAYRRFYGTA--
Tammar wallaby	P0C226	YLYQTLGFPAPYPDPQENKREVCELNPDCDELADHIGFQEAYRRFYGTA--
Human	P02818	YLYQWLGAPVPYPDPLEPRREVCELNPDCDELADHIGFQEAYRRFYGPV--
Lowland gorilla	P84349	YLYQWLGAPVPYPDPLEPRREVCELNPDCDELADHIGFQEAYRRFYGPV--
Orangutan	P84350	YLYQWLGAPVPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGPV--
Chimpanzee	P84348	YLYQWLGAPVPYPDTLEPRREVCELNPDCDELADHIGFQEAYRRFYGPV--
Pig-tailed macaque	UCSC	YLYQWLGAPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGPV--
Rhesus macaque	A2D670	YLYQWLGAPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGPV--
Crab eating macaque	P02819	YLYQWLGAPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGPV--
Black-handed spider monkey	A2D4U1	YLYQWLGAPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGPV--
Bushbaby	UCSC	YLHHWLGASVPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGTA--
Opossum	UCSC	HLYNWQGLPAPYPDPLEQKREVCELNPDCDELADHIGFSEAYRRFYGTA--
Tree shrew	UCSC	YLDYGMGAPAPYPDPLEPKREMCELSADCDELADHIGFAEAYRRFYGTA--
Lesser hedgehog	UCSC	YLDQGLGVPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYQRFYGPV--

Either UniProt accession numbers are given, or 'UCSC' when the sequences were derived from the UCSC human (hg18) referenced 28-way whole-genome alignment (Karolchik et al., 2003). Conserved residues are highlighted in black and similar residues are highlighted in grey.
[a]Consistent with the sequence stated by Ostrom et al. (2006) and not the P83005 sequence as previously reported by Carstanjen et al. (2002).

2.2. Osteocalcin extraction and analysis

OC was isolated following a modification of the method of Nielsen-Marsh et al. (2002). In brief, samples of approximately 50 mg of bone powder were subject to demineralisation in 1 ml 0.5 M sodium EDTA (pH 8.2) for 24 h with constant mixing. Following demineralisation, the samples were centrifuged at 15,000 × *g* for 10 min, the supernatant collected and the pellet discarded. The 100 mg C18 solid phase extraction (SPE) column (Amersham Biosciences) used for OC purification was prepared with 1 bed volume of methanol, followed by 1 bed volume of 90% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA) (v/v) and subsequently by 1 bed volume of 5% ACN, 0.1% TFA (v/v). The demineralised supernatant from each sample was then applied to a prepared column and allowed to pass through under gravity. Once the meniscus of the sample reached the solid phase of the column, 2 bed volumes of 5% ACN, 0.1% TFA (v/v) were applied in order to wash the unbound peptides from the column. Following these washes, a stepped gradient was applied consisting of the following concentrations of ACN in 0.1% TFA (v/v) in order: 26%, 32%, 36%, 39%, and collected in 2 ml plastic microfuge tubes. These fractions were determined as the optimum for OC retrieval with the 36% fractions found to yield the most purified and concentrated OC. The 36% fractions were then dried overnight in a centrifugal evaporator and resuspended with 10 µl of 1% *n*-octyl-glucopyranoside in 50 mM Tris buffer pH 7.4. The resuspended fractions were diluted tenfold into 5% ACN, 0.1% TFA, and 0.5 µl spotted onto a MALDI target plate (Applied Biosystems). Then 0.5 µl of α-cyano-4-hydroxycinnamic acid matrix solution (1% in ACN/H_2O 1:1 v:v) was applied to the sample spot whilst still wet and allowed to air dry. The target plate was then loaded into the AB 4700 mass spectrometer and each sample was analysed by MALDI-TOF-MS in linear mode. Each spectrum was acquired using a mass/charge (*m/z*) range of 900–6000, where the expected *m/z* value for the bovine OC $M + H^+$ ion is 5721. The signal/noise (S/N)

Fig. 1. Map of Europe showing the location of the four archaeological sites from which samples were taken.

value of each OC $M + H^+$ peak was taken from Applied Biosystems's Data Explorer software and used as a means to describe the quality of the data used to identify species from the protonated molecule for OC.

As levels of recoverable OC fall, the intensity of the mass spectrometric signal for the protonated OC molecule decreases, while the intensity of the surrounding noise signal remains constant. The S/N value for the OC signal thus provides a crude estimate of the quality of data from recoverable OC. As well as sample origin and approximate age, the OC S/N values and the success or failure of the three aDNA amplicons were tabulated along with Asx D/L; Asx D/L is a common measure of the extent of deterioration of bone protein (see Poinar et al., 1996). No other estimate of the quality of bone preservation is reported (cf. Smith et al., 2005).

Contamination of OC in archaeological bone is considered less of a problem than in the case of aDNA because: (i) OC is only found in bone and allochthanous OC cannot therefore be "introduced" by microbial contamination or sample handling; (ii) OC has a unique identifiable mass (bovine OC $M + H^+$ has *m/z* 5721), and thus species identification depends on *m/z* measurement, in a similar manner to the authentication of expected DNA fragment masses in forensic testing (Petkovski et al., 2005) and aDNA research (Petkovski et al., 2006); and (iii) the OC is not being amplified before analysis, whereas small amounts of contaminating DNA can potentially become much more abundant than the target DNA after amplification. In addition to this, even with the sensitive instrumentation used here, OC is no longer detectable if its concentration falls by approximately four orders of magnitude from that in modern bone. DNA can still be detected even when concentrations fall by greater than eight orders of magnitude. In the light of these considerations, the most likely source of contamination is carry-over contamination of OC on the sample plate (MALDI target plates are usually washed thoroughly and reused), and so care was taken to ensure that archaeological OC preparations were not analysed on target positions used previously for modern bovine OC.

2.3. Amino acid analysis/racemisation

Amino acid racemisation analyses were also carried out on the 34 archaeological bone samples. Approximately 1 mg of bone powder was treated with 100 µl of 7 M HCl under N_2 at 110 °C for 18 h to demineralise the hydroxyapatite and release any peptide-bound amino acids by hydrolysis. Samples were then dried with a centrifugal evaporator and rehydrated

with 500 µl of 0.01 mM L-homo-arginine as an internal standard for RP-HPLC analysis. The amino acid compositions were analysed by RP-HPLC using fluorescence detection, following a modification of the method of Kaufman and Manley (1998). Following a tenfold dilution in 0.01 mM L-homo-arginine, a 2 µl sample was injected and mixed online with 2.2 µl of derivitising reagent (260 mM *n*-iso-butyryl L-cysteine (IBLC) and 170 mM *o*-phthaldialdehyde (OPA) in 1 M potassium borate buffer, adjusted to pH 10.4 with potassium hydroxide pellets). The amino acids were separated on a C18 HyperSil BDS column (5 mm × 250 mm) at 25 °C with a gradient elution of three solvents: sodium acetate buffer (solvent A; 23 mM sodium acetate trihydrate, 1.5 mM sodium azide, 1.3 mM EDTA, adjusted to pH 6.00 ± 0.01 with 10% acetic acid and sodium hydroxide), methanol (solvent C) and acetonitrile (solvent D). The L and D isomers of ten amino acids were routinely analysed. During preparative hydrolysis both asparagine and glutamine undergo rapid irreversible deamidation to aspartic acid and glutamic acid, respectively (Hill, 1965). It is therefore not possible to distinguish between the acidic amino acids and their amide derivatives; they are therefore reported together as Asx and Glx respectively.

3. Results

3.1. Comparative survival of biomolecules in archaeological bone

As shown in Table 2, of the 34 bone samples, 23 gave successful results under the stringent aDNA procedure and 11 did not, whereas only 15 passed for OC screening and 19 failed. Of the 19 that failed OC screening, five gave successful aDNA amplifications and of the 11 samples that were not successful at the aDNA amplifications, only one succeeded for OC screening. Although 15 of the 34 samples show the presence of OC, six of these gave approximately tenfold lower signal:noise values than modern bone, seven gave peaks in the order of fivefold lower signal:noise, and only two gave signals approaching half that of modern bone. Despite using a different method of OC estimation, our findings that even in the best

Table 2
Origin, approximate age, Asx D/L values, Asx concentration, OC S/N values (a high S/N value indicates good OC recovery, whereas S/N = 0 indicates that OC was not observed) and success/failure of aDNA screening

Sample name	Origin	Age	Site type	Asx D/L	Asx conc. (nmol/mg)	S/N	DNA success (3 fragments)
Modern	Yorkshire	Modern	–	0.070	141.744	96.73	Y
AS1	Asine	1550–2000 BC	Open air	0.147	24.435	0	Y
AS11[a]	Asine	1550–2000 BC	Open air	0.112	96.081	0	Y
AS12[a]	Asine	1550–2000 BC	Open air	0.121	70.702	0	Y
AS3	Asine	1550–2000 BC	Open air	0.116	21.168	0	Y
AS4	Asine	1550–2000 BC	Open air	0.123	99.181	0	N
AS5	Asine	1550–2000 BC	Open air	0.134	86.493	0	N
AS6	Asine	1550–2000 BC	Open air	0.117	72.909	0	Y
AS8	Asine	1550–2000 BC	Open air	0.124	75.243	0	N
AS9	Asine	1550–2000 BC	Open air	0.124	93.952	0	N
DD11	Saxony	4900–4400 BC	Open air	0.147	20.720	0	N
DD16	Zauschwitz-Muschelgrube	3500–2900 BC	Open air	0.102	77.598	9.43	Y
DD17	Zauschwitz-Muschelgrube	3500–2900 BC	Open air	0.104	63.480	14.12	Y
DD18	Zauschwitz-Muschelgrube	3500–2900 BC	Open air	0.105	77.127	9.39	Y
DD19	Zauschwitz-Muschelgrube	3500–2900 BC	Open air	0.102	58.423	20.28	Y
DD20[a]	Zauschwitz-Muschelgrube	3500–2900 BC	Open air	0.102	51.282	5.99	Y
DD3	Saxony	5500–4900 BC	Open air	0.167	42.235	0	N
DD35	Saxony	4900–4400 BC	Open air	0.124	49.905	21.86	N
DD68	Zauschwitz-Neue Grube	1200–600 BC	Open air	0.105	73.843	20.45	Y
DD69	Zauschwitz-Neue Grube	1200–600 BC	Open air	0.113	43.548	21.86	Y
LE1	Lerna	Preclassical	Open air	0.112	8.051	0	N
LE2	Lerna	Preclassical	Open air	0.140	12.189	0	N
LE3	Lerna	Preclassical	Open air	0.116	55.672	0	Y
LE5	Lerna	Preclassical	Open air	0.132	16.974	0	N
M45	Portalón	Bronze Age	Cave	0.091	136.413	20.45	Y
M52	Portalón	Bronze Age	Cave	0.094	111.058	6.74	Y
M53	Portalón	Bronze Age	Cave	0.091	128.614	7.13	Y
M54	Portalón	Bronze Age	Cave	0.093	155.732	0	Y
S12	Portalón	Bronze Age	Cave	0.104	50.550	0	Y
S13	Portalón	Bronze Age	Cave	0.305	8.105	0	N
S2	Portalón	Bronze Age	Cave	0.107	52.443	0	Y
S23	Portalón	Bronze Age	Cave	0.090	140.098	24.22	Y
S3	Portalón	Bronze Age	Cave	0.117	26.792	42.23	Y
S4	Portalón	Bronze Age	Cave	0.111	94.017	8.4	Y
S7	Portalón	Bronze Age	Cave	0.112	32.437	28.03	Y

[a] Replicate aDNA amplification was carried out at another institution.

preserved bone, signals were only half that of modern bone, are similar to the immunological estimates of Smith et al. (2005). While there was some correspondence between Asx D/L and both aDNA amplification and OC screening results (e.g. samples with Asx D/L >0.140 do not contain either aDNA or OC), many fewer samples yielded a protonated molecule for OC than successfully yielded aDNA (Fig. 2).

3.2. Amino acid concentrations and biomolecule survival

To investigate the relationships between amino acid concentration and the survival of amplifiable DNA and/or the presence of the OC $M + H^+$ ion, the concentrations of selected amino acids were plotted (Fig. 3). The absence of OC is independent of the amino acid concentration (Fig. 3B,D); it is absent even in samples with similar amino acid concentrations to modern material (see Fig. 3A). However, the failure of aDNA amplification (Fig. 3C,D) does appear to have some correlation with lower amino acid yields. In the absence of aDNA (Fig. 3C,D) and OC (Fig. 3B,D), the relative amino acid compositions are identical to those of modern and well-preserved archaeological samples (Fig. 3A).

4. Discussion

4.1. Survival of OC vs DNA

The data presented here indicate that we are unable to detect the *intact* OC molecule (i.e. the protonated molecule, $M + H^+$ at *m/z* 5721) in several samples that amplified all three selected aDNA amplicons in duplicate. All of the samples from both Greek open sites (T_{eff} 21 °C), nine from Asine and four from Lerna, failed OC screening, but only four of the Asine samples and three of the Lerna samples failed aDNA sequencing. In contrast, the majority of samples from the German open sites (T_{eff} 12.5 °C) and the Spanish Portalón cave site (T_{eff} 9.5 °C) succeeded in yielding both OC and DNA. Although limited by sample size, a preliminary analysis of the factors that affect burial diagenesis, such as temperature and site type, does appear to indicate fewer bones with detectable OC in the sites of warmer climate. It is interesting to note that the success of aDNA retrieval appears to be less affected by burial temperature than OC retrieval.

4.2. Screening Bone Samples Using Asx Racemisation and Amino Acid Concentration Ratios

The Asx D/L data presented show some correlation with OC and aDNA survival, where no sample of Asx D/L >0.140 is successful for either molecule. The survival of aDNA appears to bear some relationship to amino acid concentrations, in which the samples that failed aDNA amplification appear to have lower amino acid concentrations than those samples that were successful (although this apparent correlation is not statistically significant at a 95% confidence level: p-value = 0.053, Mann–Whitney 2-tailed test). The survival of OC does not correlate with amino acid concentration (p-value = 0.088, Mann–Whitney 2-tailed test), as several

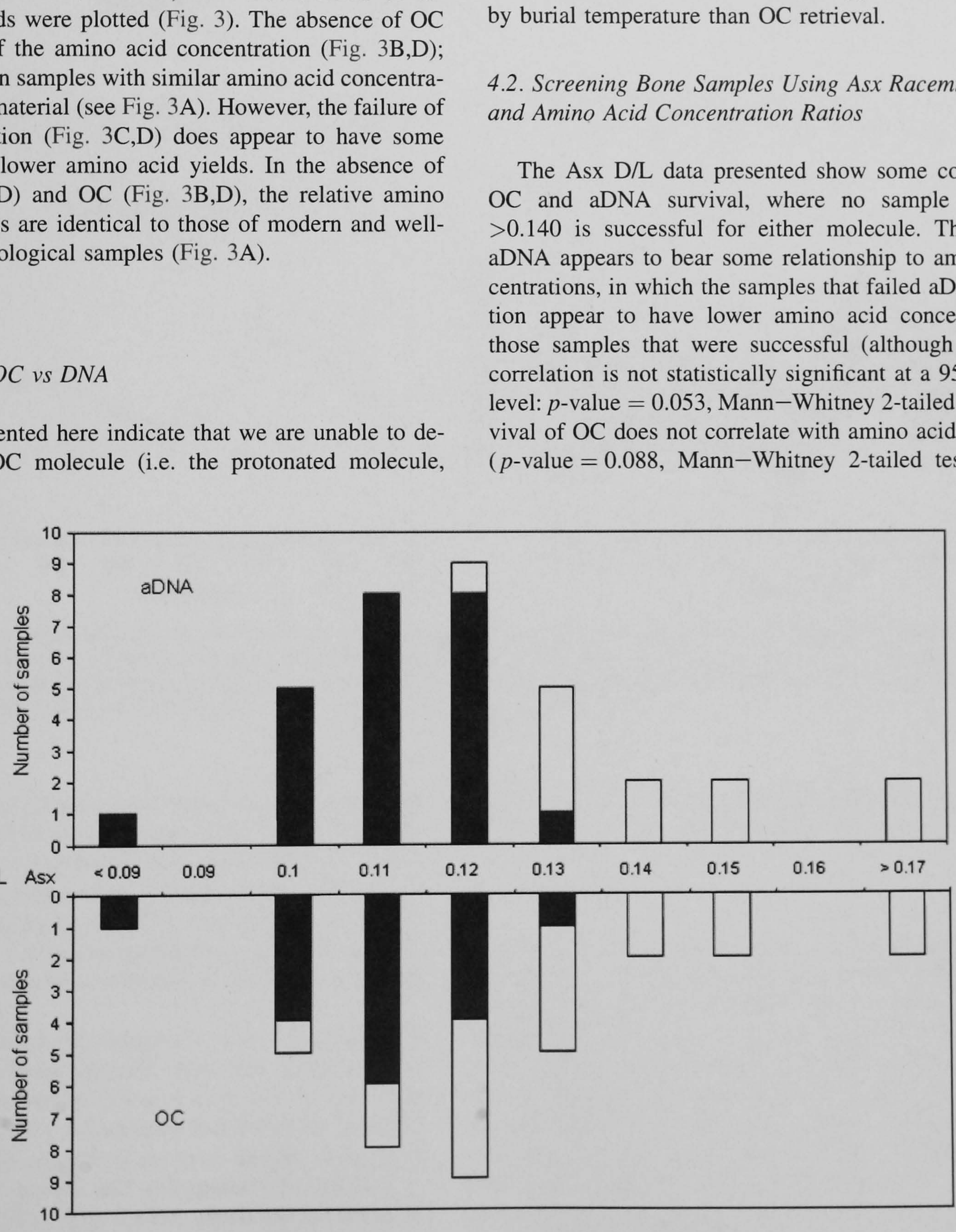


Fig. 2. Comparative ability to recover aDNA and OC as a function of D/L Asx (filled bars indicate the success of aDNA amplification (top) and OC detection (bottom), hollow bars indicate the failure of aDNA amplification (top) and OC detection (bottom)).

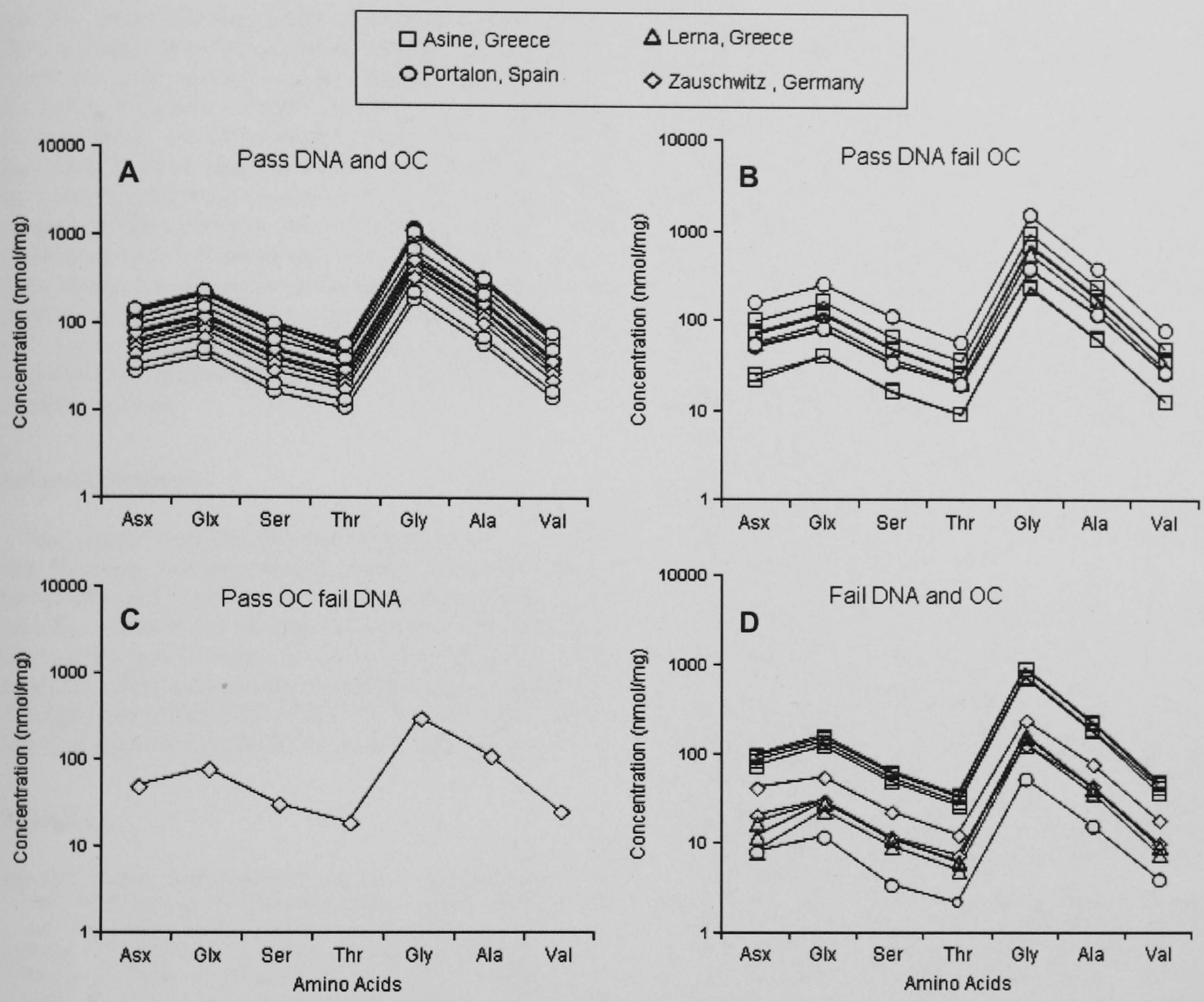


Fig. 3. Amino acid concentrations of sub-fossil cattle bones: (A) samples that passed both aDNA screening and OC screening; (B) samples that passed aDNA screening but failed OC screening; (C) samples that passed OC screening but failed aDNA screening; and (D) samples that failed both aDNA and OC screening. The modern standard is represented by a strong unbroken line without any symbols (in A). Saxony samples are represented by a diamond, Lerna samples by a triangle, Portalón samples by a circle and Asine samples by a square. The *y*-axis is on a logarithmic scale.

samples that have amino acid concentrations as high as those of modern samples failed OC screening (Fig. 3B,D). Although using Asx D/L values and amino acid concentrations to screen for OC preservation may not be useful in samples that are well preserved, they could be beneficial in screening out the poorly preserved samples. For example, samples with high Asx D/L values and low amino acid concentrations are unlikely to yield either OC or aDNA.

Due to the dominance of collagen in bone, the amino acid profiles of modern bone samples effectively exhibit amino acid ratios characteristic of collagen (e.g. Gly to Asp ratio 5.5) (Elster et al., 1991). All samples that failed for both OC detection and aDNA amplification have similar amino acid compositions to the modern and well-preserved samples (i.e., those containing aDNA) (Fig. 3A,D), which are indicative of the persistence of bone collagen. Although the long-term survival of bone collagen is not unexpected, mineral-bound NCPs such as OC have long been considered to survive for much longer (Nielsen-Marsh, 2002; Masters, 1987).

5. Conclusions

We have attempted to survey a substantial number of archaeological samples and the results indicate that OC does not have the potential for use as a routine species-specific biomarker (i.e. ZooMS) partly due to limitations in information content (Table 1) but principally, and somewhat unexpectedly, due to poor survival. One advantage of analysing protein, as opposed to aDNA, is the avoidance of contamination problems that arise with the amplification required for aDNA analysis. The high number of negative results for OC screening and the mass values obtained for those positive results (*m/z* value of OC $M + H^+$ peaks) indicate the absence of OC contamination within this study. We have shown that samples with high

Asx D/L values and low amino acid concentrations are unlikely to yield either OC or aDNA. Furthermore, amino acid composition analyses in this study indicate the persistence of the dominant protein collagen in archaeological bones that fail both aDNA and OC screening. Thus we do not propose that aDNA is more likely to survive in archaeological bone than proteins, but instead want to point out that ancient protein research should investigate the potential of collagen, which certainly survives beyond the scope of current aDNA research. Some studies are starting to investigate collagen's survival in detail on exceptionally well-preserved palaeontological samples (Asara et al., 2007) and our own research is investigating its potential for species identification and phylogenetic reconstruction analyses.

Acknowledgements

The authors would like to thank Rengert Elburg for samples from Germany, José Miguel Carratero and Ana Isabel Ortega for samples from Spain and Christina Riesberg, Gulög Nordquist and Jan Storå for samples from Greece. We would also like to thank Abi Bowman for ancient DNA replications. The authors gratefully acknowledge the NERC for provision of a studentship grant NER/S/J/2004/13017 for MB, NE/C511148/1 to MJC and grant NE/C513842/1 for amino acid analyses.

References

Ajie, H.O., Kaplan, I.R., Hauschka, P.V., Kirner, D., Slota, P.J., Taylor, R.E., 1992. Radiocarbon dating of bone osteocalcin - isolating and characterizing a noncollagen protein. Radiocarbon 34, 296–305.

Anderung, C., Bouwman, A., Persson, P., Carretero, J.M., Ortega, A.I., Elburg, R., Smith, C., Arsuaga, J.L., Ellegren, H., Gotherstrom, A., 2005. Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. Proceedings of the National Academy of Sciences of the United States of America 102, 8431–8435.

Asara, J.M., Schweitzer, M.H., Freimark, L.M., Phillips, M., Cantley, L.C., 2007. Protein sequences from mastodon and tyrannosaurus rex revealed by mass spectrometry. Science 316, 280–285.

Carstanjen, B., Wattiez, R., Armory, H., Lepage, O.M., Remy, B., 2002. Isolation and characterization of equine osteocalcin. Annales de Medecine Veterinaire 146, 31–38.

Collins, M.J., Muyzer, G., Westbroek, P., Curry, G.B., Sandberg, P.A., Xu, S.J., MacKinnon, D.I., 1991. Preservation of fossil biopolymeric structures: conclusive immunological evidence. Geochimica et Cosmochimica Acta 55, 2253–2257.

Collins, M.J., Westbroek, P., Muyzer, G., deLeeuw, J.W., 1992. Experimental evidence for condensation reactions between sugars and proteins in carbonate skeletons. Geochimica et Cosmochimica Acta 56, 1539–1544.

Collins, M.J., Child, A.M., van Duin, A.T.C., Vermeer, C., 1998. Ancient osteocalcin; the most stable bone protein? Ancient Biomolecules 2, 223–238.

Collins, M.J., Gernaey, A.M., Nielsen-Marsh, C.M., Vermeer, C., Westbroek, P., 2000. Slow rates of degradation of osteocalcin: green light for fossil bone protein? Geology 28, 1139–1142.

Cooper, A., Poinar, H.N., 2000. Ancient DNA: do it right or not at all. Science 289, 1139.

Curry, G.B., 1988. Amino acids and proteins from fossils. In: Eglinton, G., Curry, G.B. (Eds.), Molecular Evolution and the Fossil Record. Paleontological Society, Lancaster, PA, pp. 20–33.

Elster, H., Gil-Av, E., Weiner, S., 1991. Amino acid racemization of fossil bone. Journal of Archaeological Science 18, 605–617.

Gotherstrom, A., Collins, M.J., Angerbjorn, A., Liden, K., 2002. Bone preservation and DNA amplification. Archaeometry 44, 395–404.

Hauschka, P.V., 1977. Quantitative determination of γ-carboxyglutamic acid in proteins. Analytical Biochemistry 80, 212–223.

Hauschka, P.V., 1980. Osteocalcin: a specific protein of bone with potential for fossil dating. In: Hare, P.E., Hoering, T.C., King Jr., K. (Eds.), Biogeochemistry of Amino Acids. Wiley, New York, pp. 75–82.

Hauschka, P.V., 1986. Osteocalcin: the vitamin K-dependent Ca^{2-}-binding protein of bone matrix. Haemostasis 16, 258–272.

Hill, R.L., 1965. Hydrolysis of proteins. Advances in Protein Chemistry 20, 37–107.

Huq, N.L., Tseng, A., Chapman, G.E., 1990. Partial amino acid sequence of osteocalcin from an extinct species of ratite bird. Biochemistry International 21, 491–496.

Jahren, A.H., Todd, L.C., Amundson, R.G., 1998. Stable isotope dietary analysis of bison bone samples from the Hudson-Meng bonebed: effects of paleotopography. Journal of Archaeological Science 25, 465–475.

Jay, M., Richards, M.P., 2006. Diet in the Iron Age cemetery population at Wetwang Slack, East Yorkshire, UK: carbon and nitrogen stable isotope evidence. Journal of Archaeological Science 33, 653–662.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., Weber, R.J., Haussler, D., Kent, W.J., 2003. The UCSC Genome Browser Database. Nucleic Acids Research 31, 51–54.

Kaufman, D.S., Manley, W.F., 1998. A new procedure for determining enantiomeric DL amino acid ratios in fossils using reverse phase liquid chromatography. Quaternary Science Reviews 17, 987–1000.

Macko, S.A., Lubec, G., Teschler-Nicola, M., Andrusevich, V., Engel, M.H., 1999. The Ice Man's diet as reflected by the stable nitrogen and carbon isotopic composition of his hair. FASEB Journal 13, 559–562.

Masters, P.M., 1987. Preferential preservation of non-collagenous protein during bone diagenesis: implications for chronometric and stable isotopic measurements. Geochimica et Cosmochimica Acta 51, 3209–3214.

McNulty, T., Calkins, A., Ostrom, P., Gandhi, H., Gotffried, M., Martin, L., Gage, D.A., 2002. Stable isotope values of bone organic matter: artificial diagenesis experiments and palaeoecology of Natural Trap Cave. Wyoming, Palaios 17, 36–49.

Montgelard, C., Buchy, M.C., Gautret, P., Dauphin, Y., 1997. Biogeochemical characterization of ichthyosaur bones from Holzmaden (Germany, Lias). Buletin de la Société Géologique de France 168, 759–766.

Muyzer, G., Sandberg, P., Knapen, M.H.J., Vermeer, C., Collins, M.J., Westbroek, P., 1992. Preservation of the bone protein osteocalcin in dinosaurs. Geology 20, 871–874.

Nielsen-Marsh, C., 2002. Biomolecules in fossil remains. The Biochemist, 12–14.

Nielsen-Marsh, C.M., Ostrom, P.H., Gandhi, H., Shapiro, B., Cooper, A., Hauschka, P.V., Collins, M.J., 2002. Sequence preservation of osteocalcin protein and mitochondrial DNA in bison bones older than 55 ka. Geology 30, 1099–1102.

Nielsen-Marsh, C.M., Richards, M.P., Hauschka, P.V., Thomas-Oates, J.E., Trinkaus, E., Pettitt, P.B., Karavanic, I., Poinar, H., Collins, M.J., 2005. Osteocalcin protein sequences of Neanderthals and modern primates. Proceedings of the National Academy of Sciences of the United States of America 102, 4409–4413.

Ostrom, P.H., Schall, M., Gandhi, H., Shen, T.L., Hauschka, P.V., Strahler, J.R., Gage, D.A., 2000. New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. Geochimica et Cosmochimica Acta 64, 1043–1050.

Ostrom, P.H., Gandhi, H., Strahler, J.R., Walker, A.K., Andrews, P.C., Leykam, J., Stafford, T.W., Kelly, R.L., Walker, D.N., Buckley, M., Humpula, J., 2006. Unraveling the sequence and structure of the protein osteocalcin from a 42 ka fossil horse. Geochimica et Cosmochimica Acta 70, 2034–2044.

Petkovski, E., Keyser-Tracqui, C., Hienne, R., Ludes, B., 2005. SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology. Journal of Forensic Sciences 50, 535–541.

Petkovski, E., Keyser-Tracqui, C., Crubézy, E., Hienne, R., Ludes, B., 2006. MALDI-TOF MS analysis of Y-SNPs in ancient samples. Progress in Forensic Genetics 11 (1288), 25–27.

Poinar, H.N., Hoss, M., Bada, J.L., Paabo, S., 1996. Amino acid racemization and the preservation of ancient DNA. Science 272, 864–866.

Prigodich, R.V., Vesely, M.R., 1997. Characterization of the complex between bovine osteocalcin and type I collagen. Archives of Biochemistry and Biophysics 345, 339–341.

Richards, M.P., Pettitt, P.B., Trinkaus, E., Smith, F.H., Paunovic, M., Karavanic, I., 2000. Neanderthal diet at Vindija and Neanderthal predation: the evidence from stable isotopes. Proceedings of the National Academy of Sciences of the United States of America 97, 7663–7666.

Schmidt-Schultz, T.H., Schultz, M., 2004. Bone protects proteins over thousands of years: Extraction, analysis, and interpretation of extracellular matrix proteins in archeological skeletal remains. American Journal of Physical Anthropology 123, 30–39.

Schweitzer, M.H., Hill, C.L., Asara, J.M., Lane, W.S., Pincus, S.H., 2002. Identification of immunoreactive material in mammoth fossils. Journal of Molecular Evolution 55, 696–705.

Shoshani, J., Lowenstein, J.M., Walz, D.A., Goodman, M., 1985. Proboscidean origins of mastodont and woolly mammoth demonstrated immunologically. Paleobiology 11, 429–437.

Smith, C.I., Chamberlain, A.T., Riley, M.S., Stringer, C., Collins, M.J., 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. Journal of Human Evolution 45, 203–217.

Smith, C.I., Craig, O.E., Prigodich, R.V., Nielsen-Marsh, C.M., Jans, M.M.E., Vermeer, C., Collins, M.J., 2005. Diagenesis and survival of osteocalcin in archaeological bone. Journal of Archaeological Science 32, 105–113.

Tuross, N., Stathoplos, L., 1993. Ancient proteins in fossil bones. In: Abelson, J.N., Simon, M.I., Zimmer, E.A., White, T.J., Cann, R.L., Wilson, A.C. (Eds.), Molecular Evolution: Producing the Biochemical Data, Part 1. Academic Press, San Diego, CA, pp. 121–129.

Ulrich, M.M.W., Perizonius, W.R.K., Spoor, C.F., Sandberg, P., Vermeer, C., 1987. Extraction of osteocalcin from fossil bones and teeth. Biochemical and Biophysical Research Communications 149, 712–719.

Willerslev, E., Cooper, A., 2005. Ancient DNA. Proceedings of the Royal Society of London Series B Biological Sciences 272, 3–16.

Woodward, S.R., Weyand, N.J., Bunnell, M., 1994. DNA sequence from Cretaceous period bone fragments. Science 266, 1229–1232.

3 Osteocalcin survival in high-temperature rendered Meat and Bone Meal

3.1 Abstract

The small mineral-binding bone protein, osteocalcin, has been applied in a number of studies on ancient bone due to predictions of its long-term stability, and offers limited species information due to its variable amino-terminus. However, such methods of species identification are currently sought for industrial applications, particularly in the food and agriculture industries. Increases in rendering temperatures for Meat and Bone Meal intended to destroy potentially harmful proteins (prions) result in the increased degradation of DNA, the most informative biomolecule used for species identification. Although fragmentary DNA does appear to survive in some of these higher rendering temperatures, contamination from laboratory reagents is problematic. The aims of this research were to investigate osteocalcin as a possible target for speciation in MBM, because it exhibits sequence variation between key economic species (such as cattle, sheep and pigs) and is believed to be thermally stable. The survival of osteocalcin was investigated in cattle and chicken MBM rendered at four temperatures; 133°C, 137°C, 141°C and 145°C. Temperatures greater than 133°C appeared to destroy intact osteocalcin, probably as a result of increased hydrolysis of constituent proteins into peptides. Although it was possible to retrieve the carboxy-terminal fragment of the protein, this peptide lacks some of the sequence variation contained in the amino-terminus, being able to distinguish between ruminant (cattle and sheep), pig and chicken samples.

3.2 Introduction

There has been much interest in the use of proteins as species-specific markers in processed foods (Muldoon *et al.* 2004), including MBM (Kim *et al.* 2004; Kim *et al.* 2005), due to their greater resistance against degradation than DNA (Curry 1988; Nielsen-Marsh 2002). Some initial studies have focussed on particular soft tissue proteins such as troponin, due to their proposed thermostability and abundance in foods derived from meat (Chen and Hsieh 2002). However, on the grounds that mineralised tissues are more resistant to decay than non-

mineralised tissues (DeNiro and Weiner 1988; Sakae *et al.* 1995; Weiner *et al.* 1989), proteins found within bone form the ideal targets for use as markers.

Osteocalcin would appear to be an ideal species-specific marker in MBM because it is one of the most abundant proteins in bone and stabilised due to the presence of Gla residues, allowing the protein to bind to bone mineral (Collins *et al.* 1998; Collins *et al.* 2000; Hauschka 1977; Hauschka 1980; Huq *et al.* 1990). OC has a number of conserved domains, including the Gla-rich calcium-binding region within the central part of the sequence, and a highly conserved cluster of aromatic amino acids at its carboxy-terminal region. It has a variable sequence that lies predominantly in the flexible amino-terminal region of the molecule, which is a putative collagen-binding domain (Prigodich and Vesely 1997) but is the first to fragment under heating (at Asp_{14}-Pro_{15}) (McNulty *et al.* 2002). Osteocalcin was an ideal target in archaeological studies for the first forays into protein mass spectrometry due to its small size (<50 residues), abundance and relative ease of extraction (Nielsen-Marsh *et al.* 2005; Ostrom *et al.* 2006; Ostrom *et al.* 2000). However, despite earlier expectations, the limited studies so far undertaken have not been particularly encouraging in respect to survival, as recent investigations comparing the extent of OC survival in comparison to other measures of bone preservation, such as the extent of histological alteration, crystallinity change and reduction in collagen yields (Smith *et al.* 2005), showed similar limitations as for aDNA survival (Gotherstrom *et al.* 2002) (which is supported herein by the results presented in Chapter 2). These observed similarities are probably because DNA is also stabilised by mineral sorption (Gotherstrom *et al.* 2002), as is OC (Hauschka 1986).

The aim of this chapter was to establish whether or not the described method of OC analysis by mass spectrometry is suitable for speciating rendered MBM samples. The four species that make up the MBM standards used throughout this thesis; cattle (*Bos taurus*), sheep (*Ovis aries*), pig (*Sus scrofa*), and chicken (*Gallus gallus*), all have published OC sequences, and at least three amino acid differences between the sequences for the different species (Table 3.1). Excluding the two carboxy-terminal residues of pig OC, most of the amino acid variation shown in the OC from these four species (and most known mammalian OC sequences) lie in the amino-terminal 19 residues.

Table 3.1 - Table of four osteocalcin sequences showing conserved residues as dots and sequence similarity is indicated by greyscale shading.Calculated average m/z values for decarboxylated OC M+H⁺ ions are also presented.

Species	Complete Osteocalcin Sequence	*m/z* OC M+H⁺
Cattle	YLDHWLGAPAPYPDPLEPKREVCELNPDCDELADHIGFQEAYRRFYGPV	5721
Sheep	...PG.....R.....	5607
Pig	....G.....R.....IA	5580
Chicken	HYAQDS.VAGAP.N...AQ.....S.....Q.....	5411

Table 3.1 indicates the OC sequence variation amongst the four species that make up the MBM samples. It is clear that the amino-terminal region is the most variable region and that the mid-region and the carboxy-terminal region of the molecule is highly conserved.

3.3 Experimental

3.3.1 Preparation of MBM Samples

It was considered that a bank containing feed samples with known characteristics (such as composition, compound feed origin, etc) stored in optimum conditions would be of great value in the development and validation of analytical methods for analysing and classifying feeds. Two types of sample sets were thus created: those from a 'real process' of production (ie., from industrial feed mills and rendering plants) and those deliberately produced under laboratory or pilot-plant conditions. Many of the sample sets of standards used within the STRATFEED project were of the latter type and listed by Garrido-Varo *et al.* (2005). Towards the end of the STRATFEED project, a different protocol was drawn to ensure the quality, origin, and processing details would be fully documented and this was supported by European Fat Processors and Renderers Association (EFPRA). These samples were produced in a pilot plant owned by Prosper de Mulder (UK) where batches of cattle, sheep, pig and chicken meals were ground through a 50 mm plate grinder and heated in an oven at 100°C. The cooker was vented for 10 min and the temperature raised to one of the focus test temperatures (133°C, 137°C, 141°C or 145°C), under 3 bar pressure. After 20 min, the pressure was released, and when a moisture level of 5-10% was reached, the MBM was discharged from the cooker, pressed and ground. This resulted in the creation of a set of 16

standard samples (MBM of four species rendered at four temperatures) called Set B-EFPRA (Garrido-Varo *et al.* 2005).

3.3.2 EU Standardised Sedimentation of MBM Samples (carried out at the Veterinary Laboratories Agency)

The MBM sample was ground in a mill (mesh size 2 mm) to produce particles of approximate size 1.5-1.7 mm. The ground sample was then added to a conical settlement beaker in 50 mL tetrachloroethylene (TCE) and the bone particles allowed to separate from the remaining tissues and organic compounds for 1 min stirring repeatedly. The sediment was left to stand for 5 min and then separated by removing the solvent, leaving the sediment at the bottom of the beaker. The sediment was taken out and dried.

3.3.3 Defatting Samples of Bone and MBM

Pieces of unheated cattle, chicken and duck bone were lipid-extracted in chloroform/methanol (83/17%) for 2 h, the soluble lipids removed and the bone shards allowed to dry before being ground to a fine powder with a liquid nitrogen grinder (Spexmill 6750) using a 5 min precooling phase followed by three repetitions of 2 min grinding and 2 min cooling steps. Because the MBM samples did not pellet with centrifugation in chloroform/methanol, the MBM samples (supplied as a powder) were lipid-extracted with 2 mL 100% hexane (being constantly mixed for 15 min). The samples were then centrifuged at 13,000 $\times g$ and the soluble lipids removed and discarded. A further 2 mL 100% hexane was then added to the pellet, vortexed and constantly mixed for 2 h. The lipid-extracted samples were then centrifuged once more and the supernatants discarded.

3.3.4 Ultrafiltration to Remove Gelatine

To remove high molecular-weight gelatine from one aliquot of the samples prior to OC isolation in order to reduce the potential suppression effects over the isolation of OC, ultrafiltration was carried out on the chicken MBM rendered at 133°C, 137°C, 141°C and 145°C; chicken was selected for further analyses because it was the only species with a complete published genomic information. 2 mL EDTA solubilised proteins were added to

Millipore 4 mL 10 kDa molecular weight cut-off (MWCO) ultrafiltration units and centrifuged at 5000 $\times g$ for 30 min. 4 mL 50 mM ammonium bicarbonate was added and centrifuged and this step repeated. The filtrates were then applied to the C18 SPE cartridges for OC isolation as previously described (section 2.2.3).

3.3.5 Isolation of Osteocalcin

OC was also isolated from aliquots of defatted MBM samples without the use of ultrafiltration following the method described in section 2.2.3, producing OC-containing C18 fractions. For the analysis of spiked OC in MBM, OC was extracted from duck bone following EDTA demineralisation and isolated by SPE and analysed by MALDI-MS. Duck was chosen as an avian species with OC having a mass distinct from that of chicken (duck OC $M+H^+$ known from Buckley 2004). The OC was resuspended in 1 mL 1% OG 50 mM Tris-HCl and combined with the EDTA-soluble fraction of the chicken MBM sample rendered at 145°C. The spiked sample was then reapplied to a C18 SPE column and extraction for OC was carried out.

3.3.6 SPE-MALDI-MS Analysis of Osteocalcin

The resuspended C18 fractions were diluted tenfold into 5% acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA) and 0.5 µL spotted onto an Applied Biosystems 192 well MALDI target plate. 0.5 µL of α-cyano hydroxycinnamic acid (Fluka) solution (1% w/v in diluent = 50% ACN, 50% H_2O, 0.1% TFA, and this solution diluted fourfold in diluent) was applied to the sample spot whilst still wet and allowed to air dry. The target plate was then loaded into an Applied Biosystems 4700 mass spectrometer and one aliquot of each of the samples was directly analysed by MALDI-TOF-MS using the linear detector with an m/z range 900-6000. The spectra were then noise filtered in Applied Biosytems Data Explorer (Noise Removal Std Dev = 2) on all spectra resulting from MBM samples. The m/z values of peaks observed in samples were searched against possible non-specific cleavage products of chicken OC (UniProt Accession No. P02822) and chicken collagen (I) (UniProt Accession No. P02457 for α1 chain and P02467 for α2 chain) using the FindPept tool (http://www.expasy.ch/tools/findpept.html). Another aliquot of each of the samples was

subject to LC-MALDI in order to further investigate peptides present in the expected OC fractions.

3.3.7 LC-MALDI-MS Analysis of Osteocalcin

3 μL of the diluted sample is applied to an Ultimate nanoLC (LC Packings) using a Dionex 0.2 mm x 50 mm PSDVB monolith column with a 1-50% (solvent B in A) continuous gradient in 20 min (solvent A = 2% ACN, 98% H_2O, 0.1% HFBA; solvent B = 100% ACN, 0.1% HFBA) at a flow rate of 3.0 μL/min. Six second fractions were plated onto a 180 spot MALDI target plate simultaneously with matrix solution (6 mg/mL α-cyano hydroxycinnamic acid (Sigma), 26 μL/mL of 5 mg/mL ammonium citrate (Sigma) in 60% ACN/40% H_2O) using a Dionex Probot sample spotter. Following the gradient, the concentration of solvent B (in A) was raised to 85% in preparation for the following sample. The MALDI analysis was carried out using both linear detector mode (m/z range 900-6000 for OC $M+H^+$ screening) and reflectron detector mode (m/z range 900-4400 for peptide analysis). Following reflectron detector mode, the 15 MS peaks of greatest S/N (above 40) were selected for product ion tandem MS (MS/MS) analysis. The resulting product ion spectra were then collectively converted to peak lists by Applied Biosystems' 4000 Explorer version 3.6 with a S/N cut-off of 15, and searched against the UniProt database (31st Jan 2007) by Mascot version 2.2. 'No enzyme' was selected and the variable modifications of oxidation of methionine, hydroxylation of proline and lysine, and the deamidation of glutamine and asparagine were included. Mass tolerances of 0.5 Da were applied to both precursor and fragment ions. 254,609 protein sequences were searched against with a 99% significance ($p = 0.01$) threshold. Significant peptide matches (Mascot 'bold red' matches; matches presented in bold represent their first match, and those presented in red represent the top ranking peptide) were selected for in each protein match and Mascot's 'standard scoring' was applied to each search.

3.4 Results

3.4.1 Presence of Intact OC in Defatted MBM after Rendering at 133°C, 137°C, 141°C, and 145°C

Cattle and chicken MBM samples treated at 133°C, 137°C, 141°C, and 145°C were defatted with hexane and demineralised with EDTA. For controls, samples of powdered bone from the four species; cattle, sheep, pig and chicken, were also defatted and demineralised for analysis. Following isolation, resuspension with detergent and dilution, the samples were analysed by MALDI-MS in linear detector mode (Figs 3.1 & 3.2).

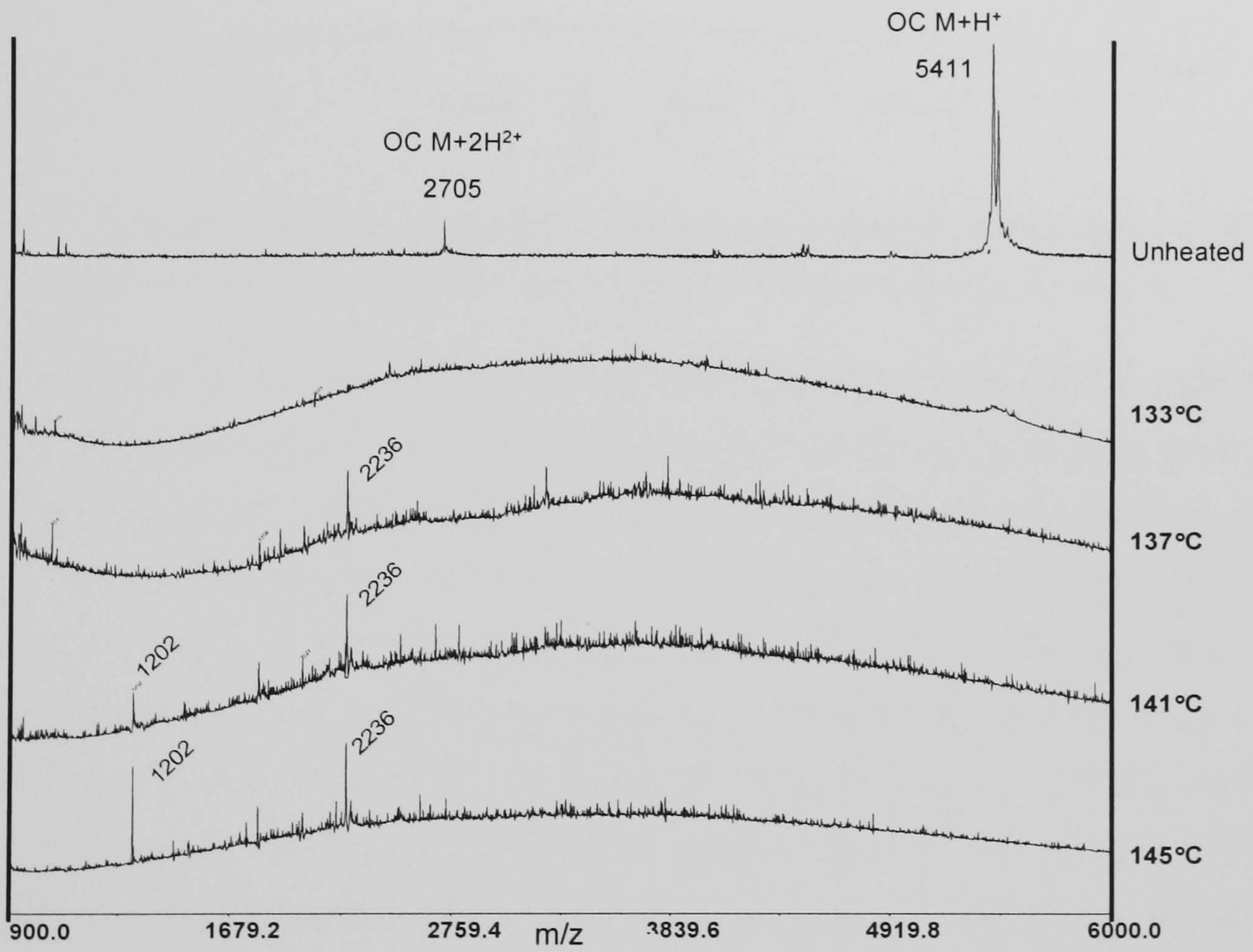


Figure 3.1 – MALDI-MS linear mode spectra of four chicken MBM samples rendered at 133-145°C extracted for OC in comparison to OC extracted from chicken bone. Average m/z values for each distinct peak are labeled.

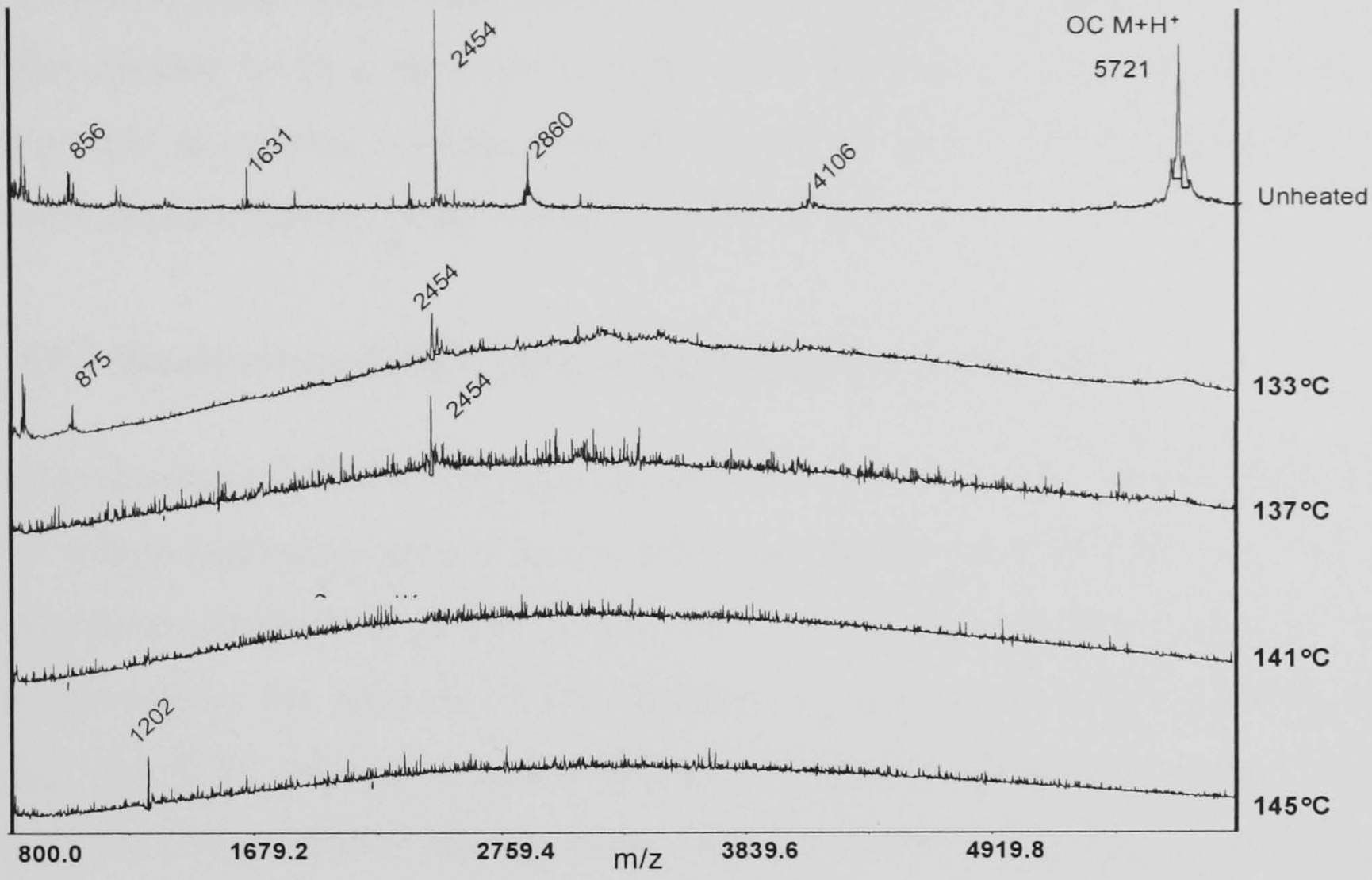


Figure 3.2 – MALDI-MS linear mode spectra of four cattle MBM samples rendered at 133-145°C extracted for OC in comparison to OC extracted from cattle bone. Average m/z values for each distinct peak are labeled.

In the spectrum of OC extracted from chicken bone (Fig. 3.1), the OC $M+H^+$ peak, which was clearly observed as a highly resolved peak at *m/z* 5411 and the only other peak present was the doubly charged OC molecule ($M+2H^{2+}$) peak at *m/z* 2705. In the spectrum of OC extracted from cattle bone (Fig. 3.2), the OC $M+H^+$ peak was clearly observed as a highly resolved peak at *m/z* 5721 and several other peaks of *m/z* 1631, 2454, 2860 and 4106, which all reflect hydrolysis fragments of OC (see Appendix 2.1); the peak at *m/z* 4106 matches the same OC fragment peak (position 15-49 of cattle OC) that McNulty *et al.* (2002) observed to increase with heating. In the unheated MBM samples, the OC $M+H^+$ peak was only represented by a poorly resolved peak in the chicken and cattle OC fractions produced from the MBM samples rendered at the lowest temperature (133°C). In the higher temperature rendered chicken MBM samples there were two peaks of *m/z* 1202 and 2236. Neither peak *m/z* value matches any of the possible calculated OC fragments but both can be attributed to both alpha chains of chicken collagen (I) (Appendix 2.1). In the higher temperature rendered cattle MBM samples there were three peaks of *m/z* 875, 1202 and 2454 where the *m/z* 875 peak could match cattle OC fragment AYRRFY, the *m/z* peak 2454 could match APAPYPDPLEPKREVCELNPDC, and the peak at *m/z* 1202 was not matched (Appendix

2.2). However all peaks could also derive from cattle collagen (I) fragmentation (Appendix 2.2). There appears to be a decrease in signal noise with the increase in MBM rendering temperature but the spectra resulting from OC extractions of the MBM samples have higher and noisier baselines than the bone samples (Figs 3.1 & 3.2).

3.4.2 OC Analysis in MBM following Gelatine Removal

Because the increasing MBM rendering temperature should cause an increase in the amount of degraded collagen (gelatine) present in the MBM sample, the possible effects of this on the chromatography and retrieval of OC needed to be considered. In order to account for the possible decrease in the amount of OC observed in any of the samples due to gelatine 'swamping' the C18 column, several additional purification procedures were carried out alternatively on chicken MBM (chosen as the only one of the four species to have complete genome sequences);

(i) spiking the highest rendering temperature MBM sample with intact OC from duck with OC of different mass in order to determine whether this OC could be detected even in the presence of the expected excess soluble collagen (from the increased gelatinisation of bone and/or skin tissue).

(ii) sedimentation of bone chips from MBM using TCE

(iii) removing the gelatine by ultrafiltration and second dimension LC

3.4.3 OC Extraction of MBM following Spiking MBM Sample with Intact OC from Unheated Bone

The resuspended C18 fractions of isolated duck OC and 145°C-rendered MBM were diluted tenfold into 5% ACN/0.1% TFA and 0.5 μL spotted onto a MALDI target plate. 0.5 μL of α-cyano-4-hydroxycinnamic acid matrix solution (1% in ACN/H_2O 1:1 v:v) was applied to the sample spot whilst still wet and allowed to air dry. The OC fraction of the spiked sample, as well as the fraction of duck OC, was analysed by MALDI-MS in linear mode.

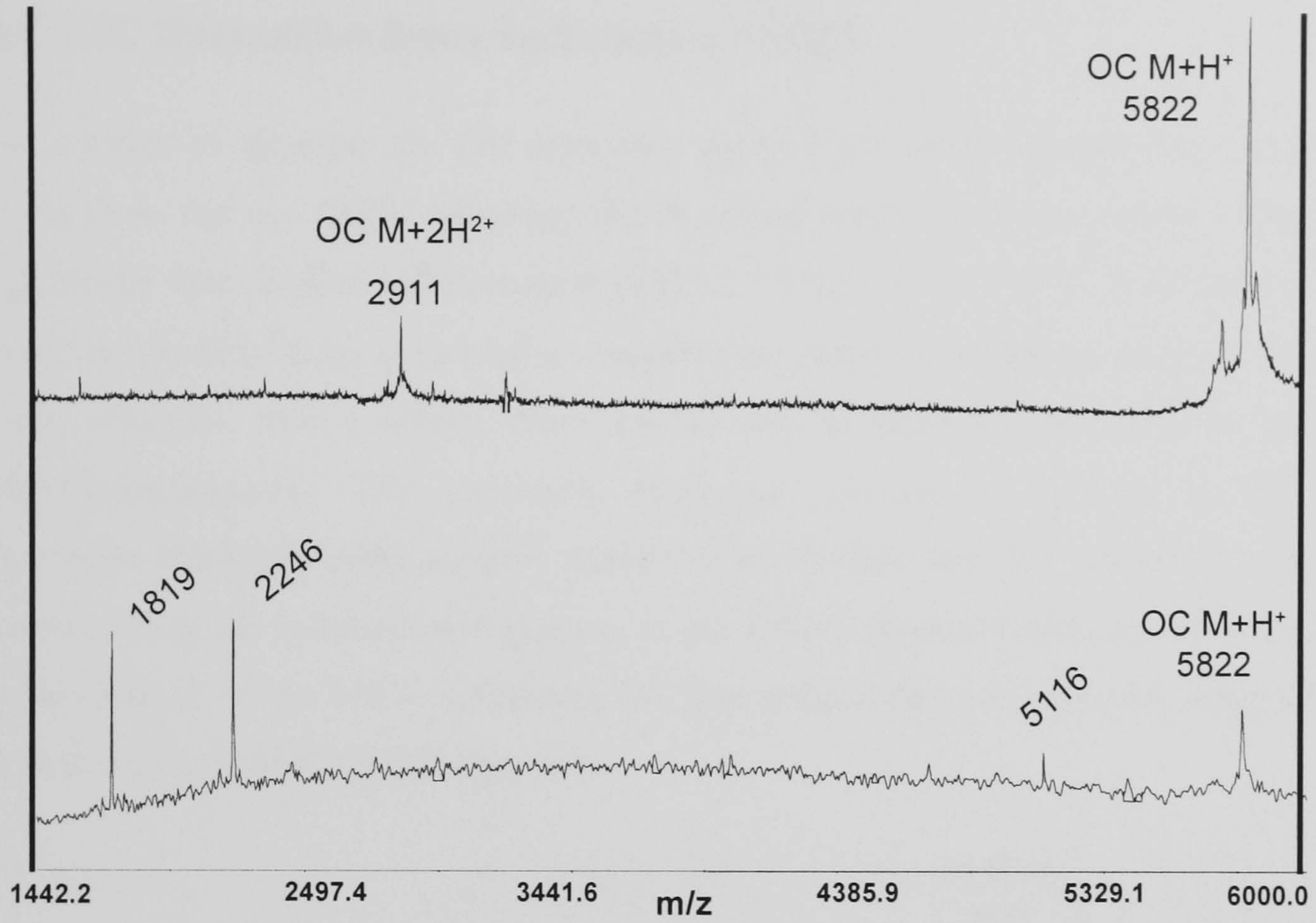


Figure 3.3 – MALDI-MS linear mode spectra of OC from duck bone (top) and chicken MBM (rendered at 145°C) spiked with duck OC prior to SPE (bottom). Average m/z values for each distinct peak are labeled.

In the analysis of OC from the unheated duck bone sample, the $M+H^+$ ion was observed at *m/z* 5822 and the OC $M+2H^{2+}$ ion was observed with *m/z* 2911. In the sample where intact OC was added to preparations of the highest temperature rendered (145°C) and thus potentially most gelatinised chicken MBM sample and purified using C18 SPE, the expected $M+H^+$ ion (*m/z* 5822) of the added OC was observed (Fig. 3.3) although with a somewhat reduced S/N (as reported by Applied Biosystems Data Explorer). Three peaks at *m/z* 1819, 2246 and 5116 were observed in the spiked MBM sample that were not observed in the spectrum for the duck OC. Of these three peaks, only the peak at *m/z* 2246 matches a hydrolysis fragment of chicken OC (Appendix 2.3), but both peaks could also be attributed to either alpha chain of chicken collagen (I) as determined by *m/z* values that matches calculated hydrolysis fragments (Appendix 2.3). However, the identity of these peptides could not be confirmed using product ion tandem MS analyses.

3.4.4 OC Extraction from Sedimented MBM

In an attempt to improve the OC extraction method for MBM samples, bone chips were purified from the raw MBM following the described method of sedimentation (see section 3.3.2) for the four chicken MBM samples (133°C, 137°C, 141°C, 145°C). Roberts *et al.* (2002) showed that boiling bone at increasing temperatures removes increasing amounts of protein (mostly collagen), thus a similar effect should also be expected for MBM of increasing rendering-temperatures. The potentially increasing amounts of gelatine in the higher temperature rendered MBM samples which could interfere with the detection of OC was removed during the sedimentation process, as the soluble fractions were discarded. Following demineralisation of the MBM sediments, OC was isolated from the samples using C18 SPE and analysed using MALDI-MS (Fig. 3.4).

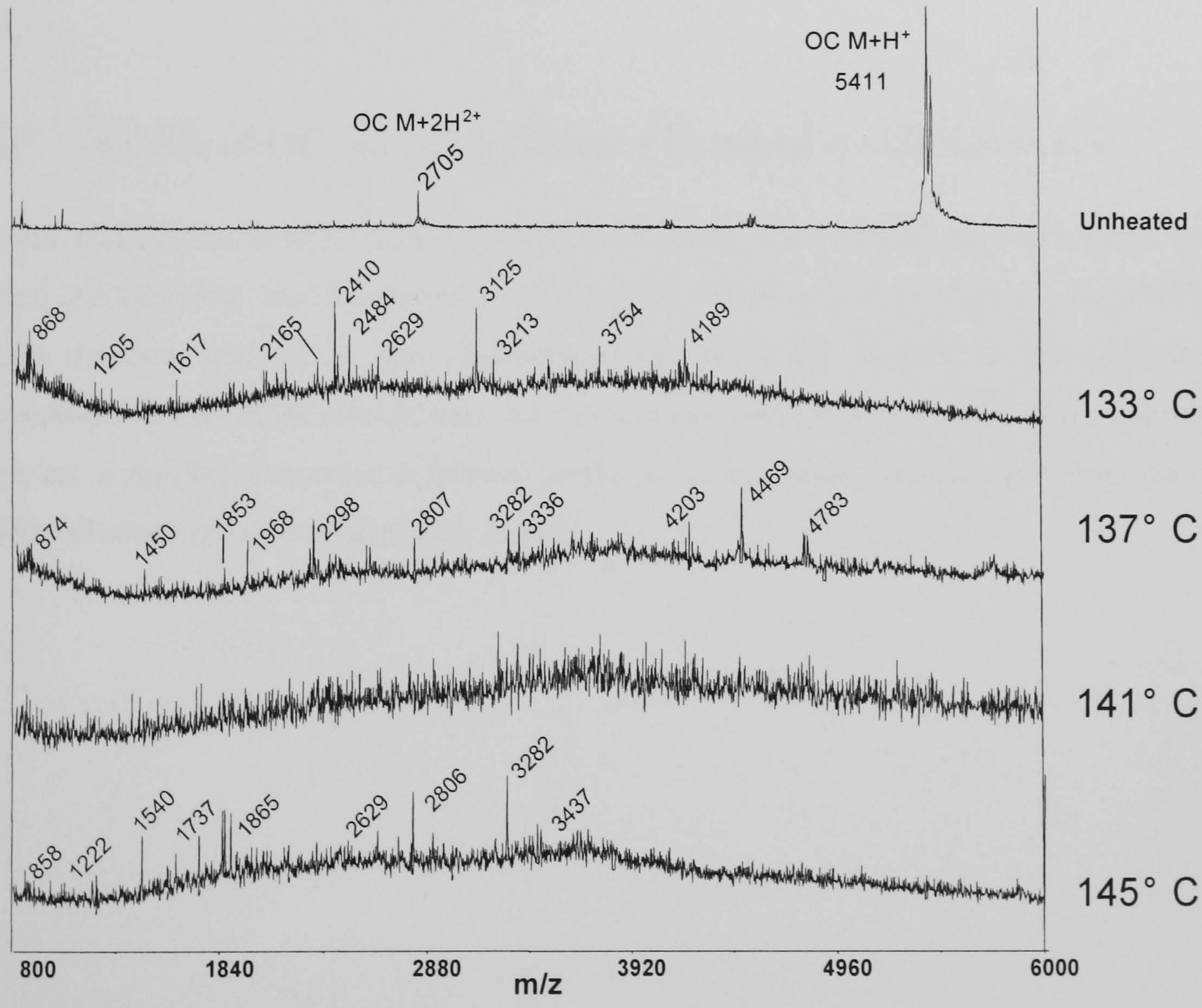


Figure 3.4 – MALDI-MS linear mode spectra of four chicken MBM sediment samples rendered at 133-145°C extracted for OC in comparison to OC extracted from chicken bone. Average m/z values for each distinct peak are labeled.

The spectra of the four MBM samples after sedimentation each had apparently lower baselines and fewer resolved peaks than their unsedimented counterparts (compare Fig. 3.4 to Fig. 3.1). Although there was only a slight signal for OC in the spectrum of the lowest temperature rendered chicken MBM sample (without sedimentation; Fig. 3.1), the spectrum of the sedimented chicken MBM sample of the same rendering temperature showed no signal for OC (Fig. 3.4). Several peaks were observed in each of the chicken MBM sediment samples (except the MBM rendered at 137°C) of which only those at m/z 2410, 3754 and 4189 in the 133°C sample, m/z 2807 and 3282 in the 141°C sample and m/z 1222, 3283 and 3437 in the 145°C sample match OC fragment masses (see Appendix 2.4). However, all peak m/z values for all MBM samples presented in Fig. 3.4 could be matched to chicken collagen (I) fragment masses (Appendix 2.4). The earlier observation of decreasing noise with increasing temperature-rendering (e.g. Figures 3.1 & 3.2) was not apparent with the sediment samples.

3.4.5 Analysis of OC following Gelatine Removal by Ultrafiltration

In order to further investigate methods of OC isolation from the MBM samples, C18 SPE was carried out following ultrafiltration of the samples. An increase in the volume of the EDTA-soluble fraction retained by the ultrafiltration unit with the increase in the rendering-temperature of the MBM sample was observed during preparation, possibly indicating that there was a significant increase in gelatine produced with increased temperature rendering of MBM and removed by the ultrafiltration units.

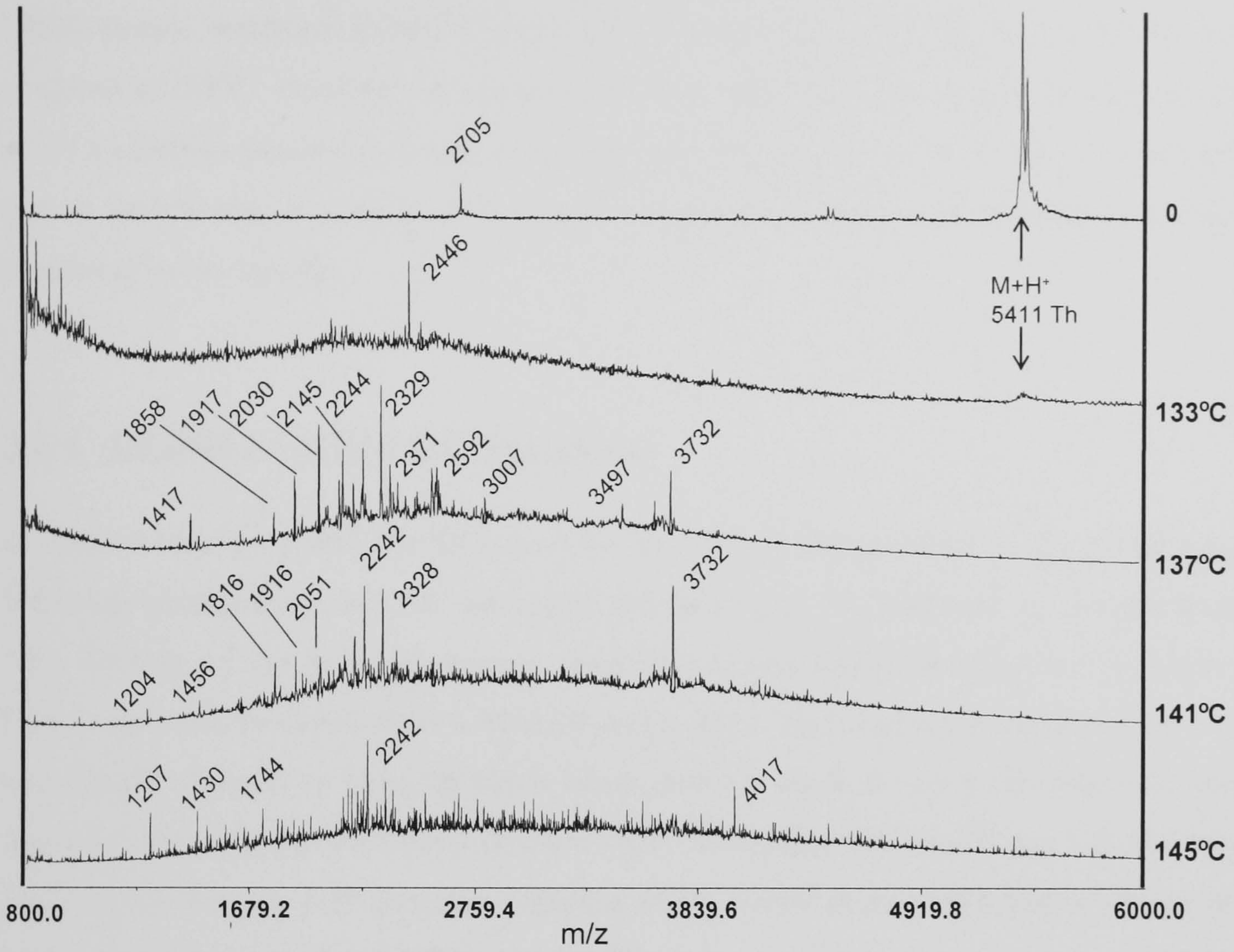


Figure 3.5 – MALDI-MS linear mode spectra of four chicken MBM samples (rendered at 133-145°C) following ultrafiltration of EDTA-soluble fraction and isolation of OC by SPE in comparison to OC from chicken bone. Average m/z values for each distinct peak are labeled.

The spectra from the fractions following ultrafiltration and isolation of OC by SPE (Fig. 3.5) contained many more peaks than the spectra of the same samples analysed for OC without ultrafiltration (Fig. 3.1). A peak at the the expected m/z for the $M+H^+$ of chicken OC was observed in the 133°C-rendered MBM sample but not in any of the samples treated above 133°C (Fig. 3.5) following a similar trend to that observed with the non-ultrafiltered samples (Fig. 3.1 & 3.2). However, the OC $M+H^+$ peak appears more resolved in the samples following ultrafiltration. Although adequate product ion spectra could not be obtained for the many peaks of lower molecular weight (m/z <5000) observed in each of the rendered MBM samples following ultrafiltration, their m/z values were searched against possible non-specific cleavage products of chicken OC and chicken collagen (I)(see Appendix 2.5). The only possible matches to chicken OC hydrolysis products were the peaks at m/z 2145, 2329, 2371 and 2592 in the MBM sample rendered at 137°C, peaks at m/z 2328 and 2242 in the

MBM sample rendered at 141°C and peaks at m/z 1207 and 2422 in the MBM sample rendered at 145°C. However, all peaks could be matched to either chicken collagen α1 (I) or α2 (I) hydrolysis products, as well as the potential matches to OC. It is thus more plausible that at least some, if not all of the peaks observed are hydrolysis fragments of collagen remaining in the sample.

3.4.6 LC-MS of MBM OC Fractions

In order to further purify the OC molecule and identify the peptides in the MBM samples following ultrafiltration and OC extraction, peptides from 3 μL aliquots of the resuspended 36% fraction of the four temperature rendered chicken MBM samples were separated by HPLC and directly plated onto a MALDI target plate. Each sample fractionated by LC-MS was initially analysed by MALDI-MS in linear detector mode to see if the intact OC can be observed following this further purification step. The chicken OC $M+H^+$ ion was observed in fractions 65-68 in the 133°C-rendered chicken MBM sample at m/z 5411, but not in any of the higher temperature rendered MBM samples (Fig. 3.6).

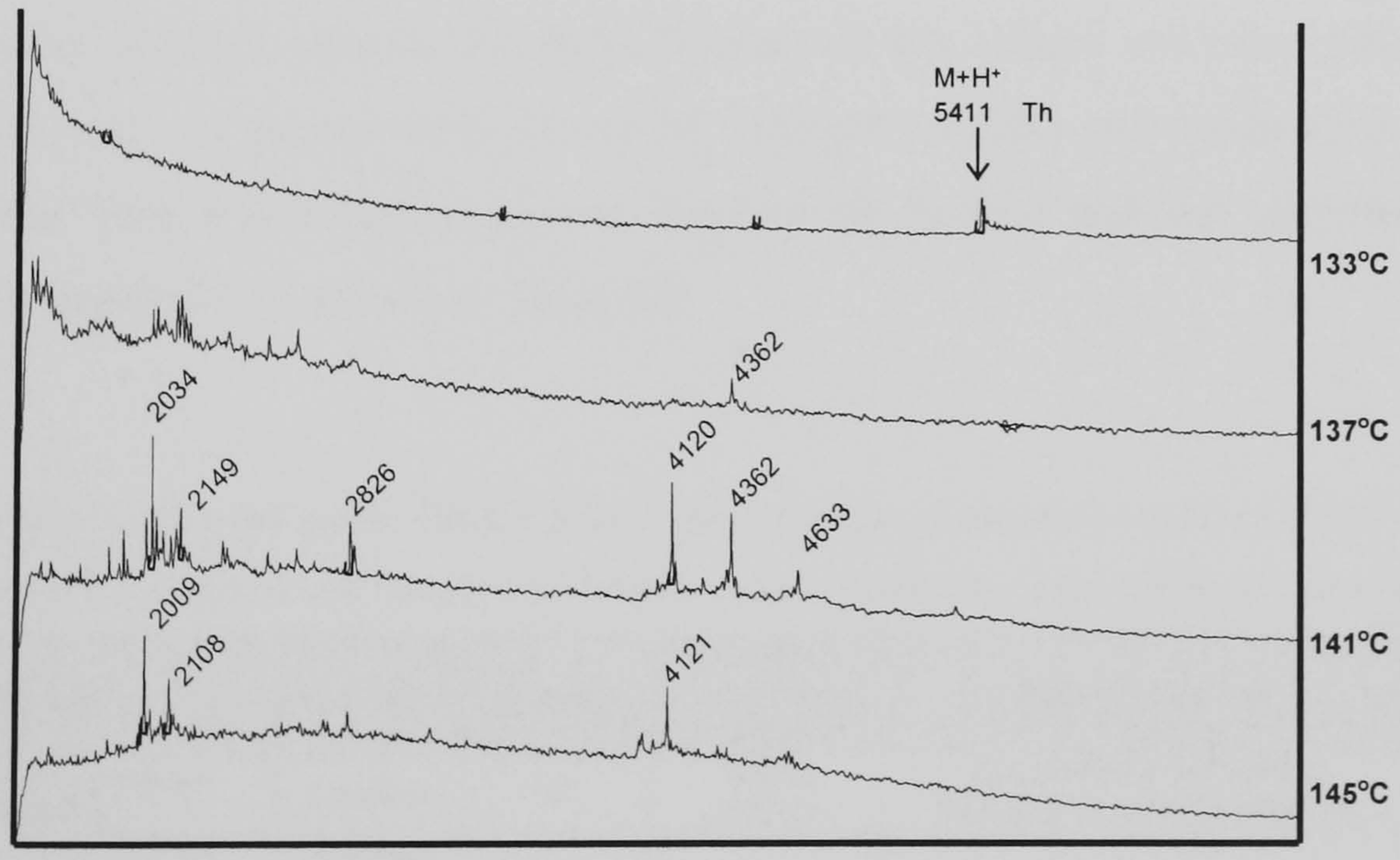


Figure 3.6 – MALDI- MS linear mode spectra of 'fraction 67' (t_R = 17 min 42 s) from four samples extracted for osteocalcin by C18 SPE and then further purified by nanoLC; chicken MBM rendered at 133-145°C. Average m/z values for each distinct peak are labeled

Intact OC was not observed in the 137°C-rendered chicken MBM sediment sample following LC-MS analysis (Fig. 3.6), but a large signal at m/z 4362.5 was observed that was not previously observed prior to LC-MS. This peak was absent in reflectron mode spectra (data not shown), presumably due to post-source decay. In the linear mode spectrum obtained for the chicken MBM sediment sample rendered at 141°C, there were six signals at m/z 2034, 2149, 2626, 4120, 4362, and 4633 and for the chicken MBM sediment sample rendered at 145°C there were four peaks observed with m/z 2009, 2108, 4121 and 4617 (Fig. 3.6). Of these peaks, only that of m/z 2149 matches an OC fragment mass (for the sequence EAQREVCELSPDCDELADQ) (see Appendix 2.6), whereas all 10 peaks match chicken collagen (I) fragment masses (Appendix 2.6). However, it was not possible to support these matches because these peaks were not observed in reflectron mode and thus no adequate product ion tandem mass spectra could be obtained.

3.4.7 UniProt Search Results for MBM Peptides

Although the additional purification step using HPLC further supported the observation that intact OC cannot be retrieved from higher temperature rendered (>133°C) MBM by the methods described, further information regarding the peptides that are present could be obtained from MS/MS analysis. LC-MALDI analysis was carried out using reflectron mode with 15 MS peaks of greatest S/N (above 40) targeted for CID and product ion tandem MS analysis. The data was searched against UniProt by Mascot and the matches above the significance threshold are shown in Table 3.2.

Table 3.2 – Mascot search results against UniProt database showing peptide sequence matches of MS/MS data. Standard scoring is applied with a peptide ion score cut-off >65. Underline indicates deamidation. Monoisotopic m/z values are labeled.

MBM Temp. (°C)	Protein	Species	Ion Score	Observed Peptide Mass	Peptide Sequence	Other MBM Samples That Peak m/z Also Observed In (°C)
137	ATP synthase protein 8 (P14093)	Chicken	83	1843.99 2127.12	KITTTKPTPWTWPWT PANKITTTKPTPWTWPWT	141 x
137	L-lactate dehydrogenase chain A	19 reptile-mammal species	77	1362.73	TLWGIQKELQF	141 & 145
141	(P02605) Myosin light chain 3, skeletal muscle isoform	Chicken	126	1566.85 2023.06	AKKITFEEFLPML AKKITFEEFLPMLQAAAN	x x
141	(P00563) Creatine kinase M Type and matches B Type	10 fish-mammal species	83	1712.03 2329.23	DHFLFDKPVSPLLLA DHFLFDKPVSPLLLASGMARD	137 x

In the 133°C MBM sample, not a single match was observed. In the 137°C MBM sample, the most confident protein match was with two peptides to an ATP synthase protein (specific to chicken), the second most confident protein match was l-lactate dehydrogenase A chain with a single peptide match (peaks at a similar m/z were also observed in the 141°C and 145°C samples). In the 141°C sample, there were two identifications; the most confident was myosin light chain 3 (a skeletal muscle isoform) with a single peptide match (with a second shorter fragment of the same peptide also matched). The second identification was to creatine kinase, (matching both muscle (M) and brain (B) types) with a single peptide match (and a fragment thereof). There were no significant peptide matches in the 145°C sample. Interestingly, there was a peak that was matched to the carboxy-terminal peptide of OC in the three samples of highest temperature, 137°C, 141°C, and 145°C, but with scores of 49, 64 and 48 (below the selected ion score threshold) respectively (see Appendix 2.8). This matched the peptide of the chicken OC sequence QIGFQEAYRRFYGPV (observed monoisotopic m/z 1813.7) but assumes that the amino-terminal glutamine residue was modified to pyro-glutamic acid, a cyclisation that is known to occur to glutamine residues under heating (Archibald 1945) and pressure (Schneider *et al.* 2003). A peak of similar m/z (monoisotopic m/z 1814.2) was present in the 133°C sample but not matched to this OC fragment. Another interesting match (but with a score below the threshold) was the identification of a peptide of collagen α2 (I) chain from chicken, of the sequence PGPVGPVGPAGAFGPR (observed monoisotopic m/z 1432.8) but this was only matched in the 137°C and 141°C samples (see Appendix 2.7 for data files and Appendix 2.8 for search results).

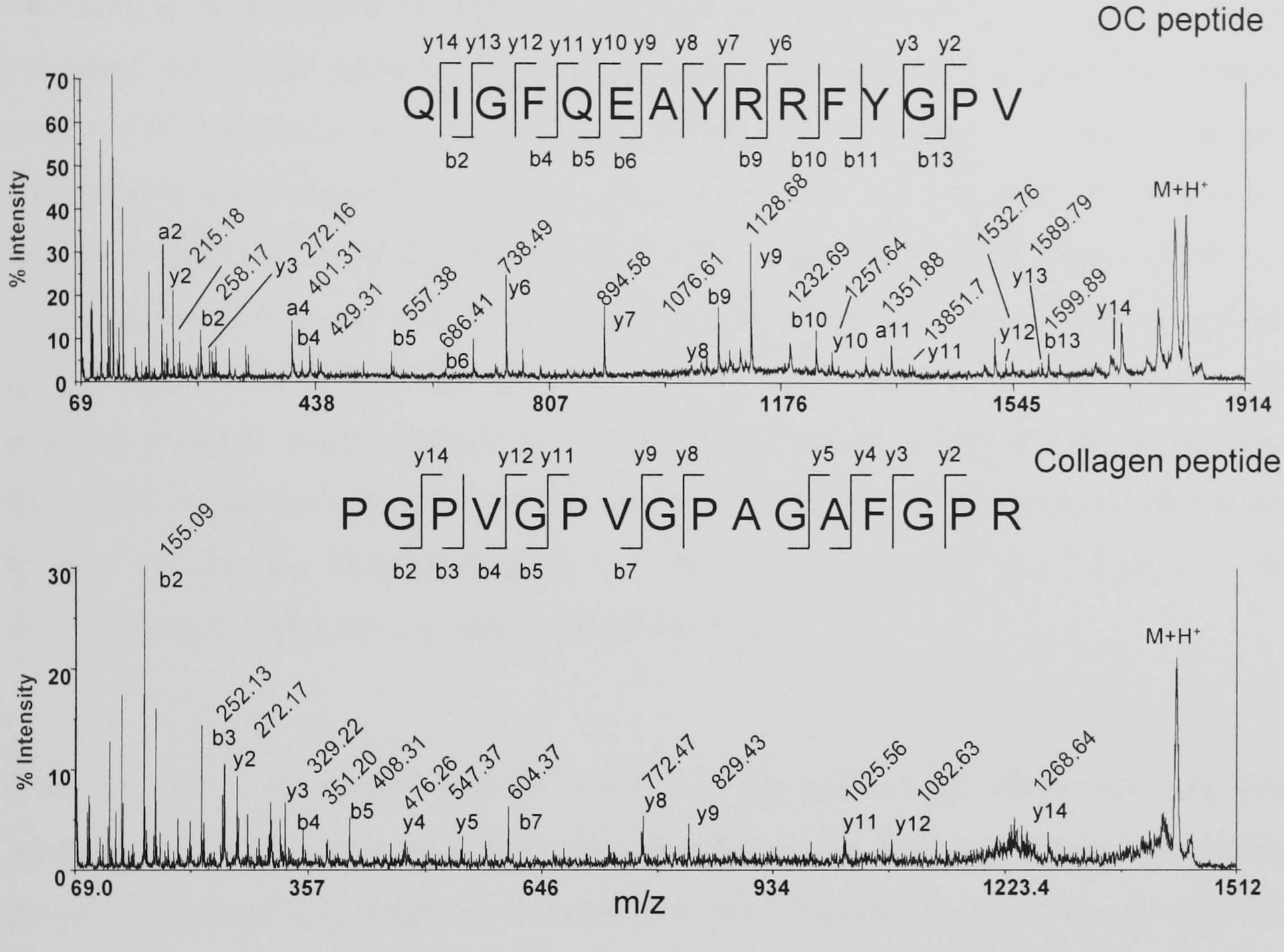


Figure 3.7 - *Manual interpretation for de novo sequencing of product ion spectra of the OC peptide and collagen peptide.*

Clearly the spectrum of the OC peptide is of better quality than the collagen peptide with more b and y ions matching the proposed sequence. The collagen peptide also appears to exhibit lower fragment ions carboxyterminal to the proline residues than aminoterminal to the proline residues, a well-known observation in CID spectra of proline-containing peptides (Breci *et al.* 2003; Vaisar and Urban 1996). The OC peptide is clearly identified with a large proportion of the expected b and y ions.

3.5 Discussion

Attempts at OC extraction from the chicken MBM rendered at 133°C, 137°C, 141°C and 145°C only resulted in the detection of an intact OC $M+H^+$ peak in the spectrum of the

lowest temperature (133°C) sample that was not present in any of the higher temperature rendered samples despite several purification steps. A pattern of OC fragmentation by hydrolysis at particularly labile peptide bonds, like that of McNulty *et al.* (2002), could not be confirmed due to the presence of collagen peptides and inability to confirm the identity of suspect OC fragments with MS/MS data. Despite the attempts to remove gelatine by ultrafiltration and further fractionation using HPLC, OC was not observed in any of the samples rendered at temperatures higher than 133°C, supporting the proposal that OC is not present in the sample and that it not being observed is not due to interferences from other components of the MBM. One possible explanation for the lack of intact OC in MBM rendered at higher temperatures is that it is being hydrolysed at temperatures slightly greater than 133°C under these rendering conditions. Interestingly, the only fragment of OC identified in these samples was from fragmentation at the Asp_{14}-Gln_{15} bond. The bond most readily cleaved in dilute acid is the Asp-Xaa bond (Smith 2002).

With the failure to observe the OC $M+H^+$ in the sedimented MBM from the lowest temperature MBM sample, it is likely that the OC protein does not survive via remaining bound to the mineral as might be expected. It is plausible that the alteration observed in the bone apatite (Wess *et al.* unpublished data) due to the presence of high pressure water leads to partial destruction of the bone mineral/protein composite and liberation of the OC into solution where it decomposes; rate of OC survival is highly dependent upon the integrity of the mineral structure (Smith *et al.* 2005). This decomposition appears to be such that hydrolysis occurs at temperatures >133°C. Although the possible presence of the carboxy-terminal fragment of OC as shown by the LC-MS analysis of the samples rendered at 137°C, 141°C and 145°C indicates its potential for survival, the sequence is highly conserved throughout most mammalian species, the only exception being that the carboxy-terminal two amino acids of pig OC are different. Hence, if shown to be reproducibly observed, despite these identifications all being slightly below the generally accepted Mascot threshold for significance, there may be potential for this peptide to distinguish between chicken, pig and ruminant, but not between the relevant ruminants, such as sheep and cattle. The Mascot search results for the MS/MS data unexpectedly showed no confident matches to collagen hydrolysis products and only one sample contained one unconfidently identified peptide (Appendix 2.8), despite the expected

abundance of gelatine in the sample. This could indicate that any hydrolysed collagen peptides not bound within the colloid of high molecular weight gelatine and thus removed by ultrafiltration are more hydrophilic than the OC fraction and were washed from the SPE cartridge before the collection of OC occured.

The only peptide matched in all four samples was from the protein l-lactate dehydrogenase (detailed data presented in Appendix 2.8), although in samples 133°C, 141°C and 145°C the matches were not considered confident with scores of only 48, 70 and 56 respectively. l-lactate dehydrogenase is an enzyme that catalyses the interconversion of lactate and pyruvate and is widespread in skeletal muscle (as well as kidney, liver and myocardium). The enzyme is produced in elevated concentrations in the blood when such tissues are injured, and its observation in MBM is likely due to a response immediately prior and during slaughter. However, this match was based on a single peptide which exhibits an identical amino acid sequence throughout at least 19 species ranging from mammals to reptiles, and is therefore not suitable for species identification purposes. The thermostability of these muscle-tissue proteins is supported by their identifications in high temperature rendered MBM samples.

3.6 Concluding Remarks

These results have shown that, despite various attempts to isolate it from other proteinaceous material, *intact* OC cannot be retrieved from high-temperature rendered (>133°C) MBM using the same methods as previously described for isolating archaeological OC and identifying it by protein mass spectrometry. Matches to some peptides of potentially thermostable proteins that are present in the OC fractions were made with confident identifications, such as l-lactate dehydrogenase and myosin in samples rendered at 137°C and 141°C respectively. However, none of these identified proteins exhibit sufficient amounts of variation to be able to distinguish between species of interest to the food and agricultural industry. The only peptide that may offer distinction between at least the ruminants (cattle and sheep) and pig is the carboxy-terminal peptide of OC, but this was not a confident match using Mascot. However, this peptide could not distinguish between chicken and ruminant OC and thus may also be highly conserved amongst other mammalian (except pig) species.

4 The characterisation of Meat and Bone Meal samples by LC-MS and amino acid racemisation

4.1 Abstract

LC-MS methods are used to investigate the potentially complex mixture of bone proteins in MBM. Bone samples from the four species of economic interest; cattle, sheep, pig and chicken, are demineralised and the acid-insoluble residue gelatinised, digested and analysed by LC MALDI-MS/MS for the purpose of identifying constituent proteins. LC-MALDI-MS/MS analyses are then carried out on cattle and chicken raw and sedimented MBM samples in order to investigate the most thermostable proteins surviving at these rendering conditions (beyond the rendering limits of most other published biomolecule studies) and shows the almost exclusive dominance of the two chains of bone collagen (I). This chapter also characterises the STRATFEED project's 'Set B-EFPRA' standards, which comprised of 16 samples (four bone meals – cattle, sheep, pig and chicken meals - each rendered at four different temperatures 133°C, 137°C, 141°C and 145°C) by amino acid composition and racemisation analyse,. Amino acid racemisation was used to investigate the extent of thermal alteration within the set of MBM standards, pre- and post-sedimentation and carried out on 'total', 'acid-soluble' and 'acid-insoluble' fractions. Contrary to expectations that the sedimented fraction would consist almost entirely of 'protected' collagen from within the bone mineral, rather than a mixture of protected mineralised collagen and unprotected non-mineralised collagen, the extent of amino acid degradation in the MBM sediment fraction was as great as that observed in the 'raw' (not sedimented) MBM fraction. The extent of racemization is also much larger in the 'acid-insoluble' fraction than expected, indicating extensive denaturation and renaturation of the dominant protein collagen. Thirdly, the extent of degradation in the lowest temperature rendering of the sheep and pig sample sets was observed to be as great as in the highest temperature rendered samples, indicating possible contamination within the set of standards.

4.2 Introduction

The reasons for investigating proteins found within mineralised tissues, such as bone, in MBM samples have been discussed in section 3.2. The organic fraction of bone is made up of ~90% collagen (I), which is also found in skin (and hence likely to be a common protein in food products in its degraded form as gelatine) the repeating Gly-Xaa-Yaa amino acid

motif (where Xaa and Yaa can be any amino acid but often proline and hydroxyproline) has deterred some studies from its use as a source of species-specific markers; some studies have used the identification of gelatine peptides to indicate the presence of animal protein in MBM (Ocana *et al.* 2004). Recent interest in proteomics for investigations into the illegal inclusions of 'contaminant' proteins has been fueled by the completion of multiple genome projects and ignited by the common need of biologists to rapidly and comprehensively evaluate complex samples of proteins on a global level. Most methods of extracting collagen (I) from bone, such as for radiocarbon dating and stable isotope analyses, use the basic purification step of acid demineralisation, where the insoluble protein residue is isolated from the soluble proteins and peptides (Brown *et al.* 1988; Longin 1971). The acid-insoluble residue was often described as 'collagen' in most early archaeological science literature, and the peptides solubilised by heating (gelatinisation) in most preparation procedures were initially thought to be solely derived from collagen (I). However, in recent years, a growing number of NCPs such as decorin, were found to bind to collagen (I) (Weber *et al.* 1996).

The aims of this chapter were to establish which peptides and how much of the collagen sequence can be identified by LC-MS from the 'acid-insoluble' fraction, what other proteins are present, and whether these sequences can be used as markers for species identification. To do this, LC-MALDI-MS was used on the tryptic digests of gelatinised bone collagen from bone samples as well as two unsedimented MBM (rendered at 145°C) and two sedimented MBM (rendered at 145°C) samples.

4.2.1 Amino Acid Racemisation

Of all the amino acids studied in terms of racemisation kinetics, the racemisation of Asx (see Chapter 2) is most widely used in archaeology due to its fast rates (van Duin and Collins 1998). The rationale for this investigative study was to establish the extent of protein alteration in MBM of different species, treated at the range of temperatures commonly used in processing plants. If dominated by collagen (I) as predicted, following the mechanisms of Asx racemisation as proposed by van Duin & Collins (1998) where the structural contstraints on the molecule prevent Asx racemisation, the extent of Asx racemisation in the 'acid-insoluble' fraction could be expected to be minimal where only the collagen telopeptides and

NCPs would racemise, the 'acid-soluble' fraction could be expected to show greatest amount of Asx racemisation because it is representative of the 'vulnerable' collagen, and the 'total' fraction could be expected to show intermediate levels of Asx racemisation. The comparison between sedimented and unsedimented MBM should highlight the overall differences between MBM in general, and the bone chips it contains. These six fractions were analysed by Asx racemisation analysis for the 16 MBM samples that make up 'Set B-EFPRA'.

4.2.2 The STRATFEED Project

The STRATFEED project (section 1.2.1) was dedicated to the investigation of new methodologies to detect the illegal addition of mammalian tissues in feedingstuffs. These methods included PCR (Brambilla *et al.* 2004; Chiappini *et al.* 2005), NIRS (Baeten *et al.* 2001; Behnam *et al.* 2002) and NIRM (Baeten *et al.* 2001). In addition, work was done on the development of ELISA tests for detecting terrestrial animal proteins in MBM samples that had been rendered at high temperatures (Chen *et al.* 2004) which focused on the thermostable muscle protein troponin. However, despite the large amount of work carried out, the standards themselves had not been analysed to establish if whether their amino acid compositions, in particular the extent of Asx racemisation, are consistent with their gradually increasing rendering temperatures (Liardon and Hurrel 1983). For further research using these samples, it was considered important to characterise them using amino acid analyses and protein analyses.

4.3 Experimental

4.3.1 Acid Treatment of MBM and Bone Samples

The preparation of the 16 STRATFEED MBM samples, the 16 STRATFEED sediment MBM samples, and the four specimens of bone (from cattle, pig, sheep, and chicken) was as previously described in sections 3.3.1 & 3.3.2. The acid-insoluble pellet obtained on centrifugation of the samples resulting from 4 h demineralisation with 0.6 M HCl at 4°C was rinsed 5 times with 2 mL distilled and deionised water (ddH_2O) until the pH reached neutral. The rinsed pellet was then freeze dried overnight before weighing to determine the dry weight. 1 mg of the acid-insoluble pellet was then resuspended in 80 µL 50 mM ammonium

bicarbonate and gelatinised for 3 hr at 65°C (for method development see Appendix 3.1). Following gelatinisation, the sample was centrifuged for 15 min at 13,000 $\times g$ and an aliquot of the supernatant removed for tryptic digestion.

4.3.2 Tryptic Digest and LC-MALDI-MS Analysis of MBM Acid-Insoluble Proteins

1 µL of 1µg/µL trypsin was added to 9 µL of sample and incubated at 37°C for 18 h. A 3 µL aliquot was analysed by LC-MALDI (section 3.3.5), with database searches carried out using Mascot's MudPIT scoring as well as standard scoring. With standard scoring, the protein score is the sum of the ion scores of all the non-duplicate peptides; where there are duplicate peptides, the highest scoring peptide is used. For MudPIT scoring, the score for each peptide is not the absolute score but the amount that it is beyond its threshold. Therefore, peptides with a score below the threshold do not count towards the protein score. The average of the thresholds used is added to the score. For each peptide, the threshold is the homology threshold if it exists; otherwise it is the identity threshold. MudPIT scoring thus reduces the amount of high scoring protein matches that consist entirely of poorly matched peptides. By default, MudPIT scoring is generally considered more appropriate for searches with greater than 1000 spectra (Matrix Science).

4.3.3 Amino Acid Composition and Racemisation Analysis

Amino acid composition and racemisation analyses were carried out on unprocessed MBM (a possible mixture of tissues), on the bone chips obtained on sedimentation of MBM, and on acid-soluble and acid insoluble fractions of both (Fig. 4.1).

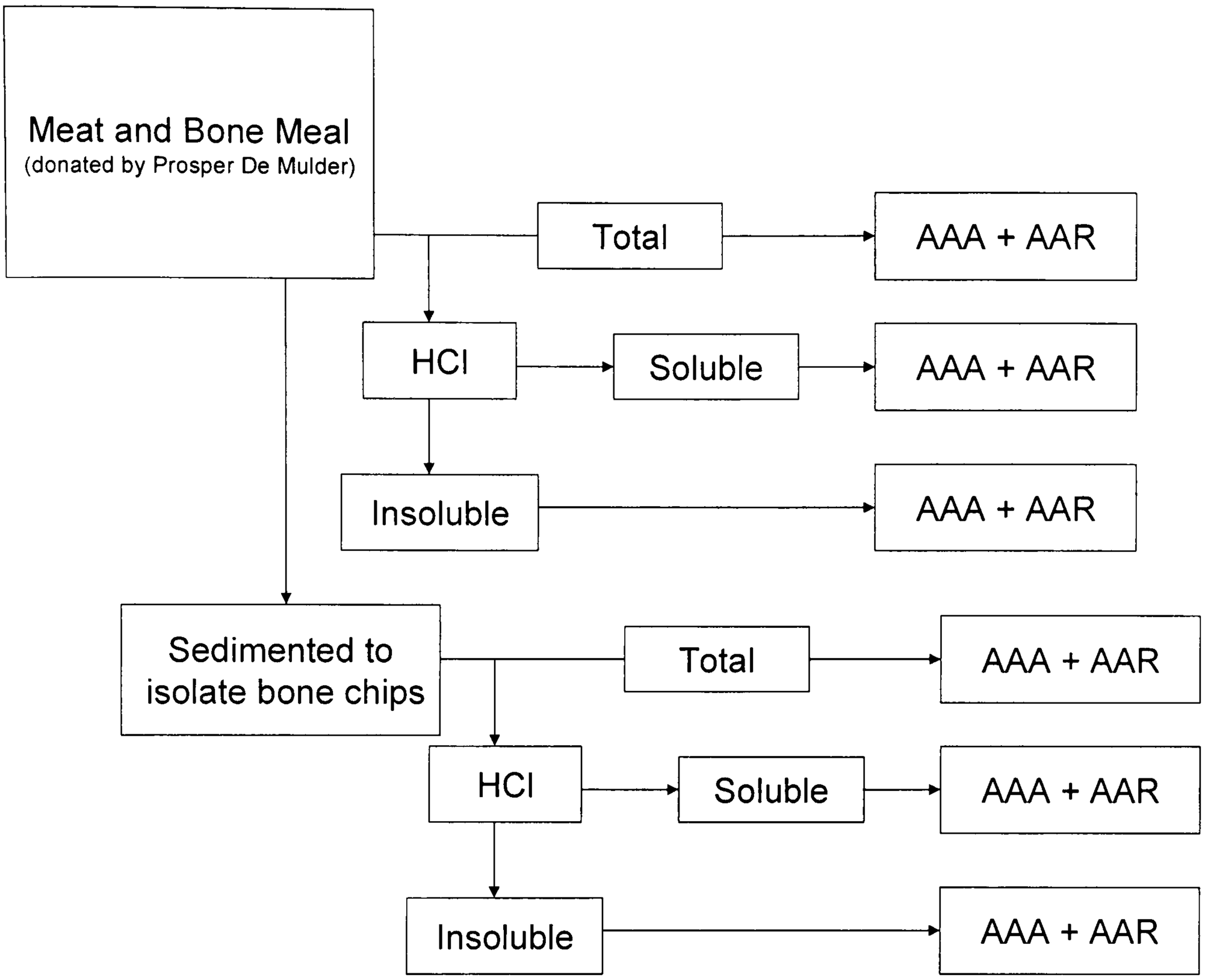


Figure 4.1 - Flow chart showing the preparation of six fractions for each sample of MBM

Amino acid composition and racemisation analysis was carried out as follows. Approximately 1 mg sample (bone and MBM) powder ('total' fraction) was treated with 100 μL 7 M HCl under N_2 at 110°C for 18 h to demineralise the hydroxyapatite and hydrolyse peptide bonds. Samples were then dried with a centrifugal evaporator. In addition to this, approximately 5 mg of bone mineral was demineralised with 200 μL 0.6 M HCl for 4 h, and following centrifugation at 13,000 *x g*, the 200 μL supernatant dried with a centrifugal evaporator. The sample ('acid-soluble' fraction) was then treated with 100 μL 6 M HCl under N_2 at 110°C for 18 h to release any peptide-bound amino acids by hydrolysis. The acid-insoluble residue from microcentrifugation was washed five times with ddH_2O and lyophilised. 1 mg lyophilised residue was measured and then treated with 100 μL 6 M HCl under N_2 at 110°C for 18 h to release any peptide-bound amino acids by hydrolysis ('acid-insoluble' fraction). All three fractions for each sample were then separated by HPLC and

analysed as described in section 2.2.4. Principal component analyses using a non-normalised classical classification type were created in XLstat.

4.4 Results

4.4.1 LC-MALDI MS analysis of Acid Insoluble Proteins in Bone

Gelatinised cattle, sheep, pig and chicken 'collagen' residues from bone were digested with trypsin and analysed by LC-MALDI-MS and MS/MS, and the data (Appendix 3.2) collectively searched using Mascot against the UniProt protein database. The search results were exported into Microsoft Excel using a significance threshold of $p<0.01$ (99% confidence) and requiring each identification to have at least one top-ranking significant peptide (as determined by Mascot 'bold red' assignments indicating the best protein match for that particular peptide). Table 4.1 shows a detailed protein summary showing protein identifications, accession numbers, number of peptide matches per protein, and total ion scores for all unique proteins. Replicates of cattle bone were analysed (labelled Samples 1, 2 and 3) and the coverage maps for the top-matching collagen (I) sequences presented (Fig. 4.2). A summary of the collagen coverages and numbers of other peptides (NCPs and unidentified peptides) from cattle bone replicates is presented along with results from pig, sheep and chicken bone collagen analyses in Table 4.2. Further results of the Mascot searches including calculated and expected peptide molecular weight values, expected m/z values, start and end positions in the matched protein, number of missed cleavages, peptide scores, peptide homology values, peptide identity values, peptide expect values, amino acid sequence information, PTM information and locations, unassigned peptides, protein length, percentage coverages, and exponentially modified Protein Abundance Index (emPAI) values for each identification are presented in Appendix 3.3 for both 'standard scoring' and 'MudPit scoring' searches.

Table 4.1 – Mascot summary of protein matches from digested cattle (in triplicate, labeled Sample 1, 2 and 3), sheep, pig, and chicken bone acid-insolluble 'collagen' analysed by LC-MALDI using standard scoring.

Sample	Protein	Organism	Acc. No	No. peptides	Total Ion Score
Cattle Bone (Sample 1)	Collagen α2(I)	Cattle	P02465	66	2886
	Collagen α1(I)	Cattle	P02453	71	2564
	Collagen α1(XI)	Mouse	Q61245	5	133
	A2HS	Cattle	P12763	5	108
	SPARC	5 mammal	*	2	87
Cattle Bone (Sample 2)	Collagen α2(I)	Cattle	P02465	82	3544
	Collagen α1(I)	Cattle	P02453	82	3093
	A2HS	Cattle	P12763	4	148
	Biglycan	7 mammal	*	2	125
	Osteocalcin	Cattle	P02820	1	111
	SPARC	5 mammal	*	2	94
	Prothrombin	Cattle	P00735	1	65
Cattle Bone (Sample 3)	Collagen α2(I)	Cattle	P02465	53	1740
	Collagen α1(I)	Cattle	P02453	56	1368
	Collagen α1(II)	Human	P02458	7	193
Sheep Bone	Collagen α2(I)	Cattle	P02465	57	3311
	Collagen α1(I)	Cattle	P02453	73	2600
	A2HS	Sheep	P29721	4	183
	Biglycan	7 mammal	*	3	134
	Collagen α2(V)	Human	P05997	2	102
	Collagen α1(II)	Human	P02458	5	196
	Serum albumin	Sheep	P14639	2	70
	Collagen α1(VII)	Human	Q02388	2	66
	SPARC	Mouse	P07214	2	61
Pig Bone	Collagen α1(I)	Cattle	P02453	63	2010
	Collagen α2(I)	Cattle	P02465	30	1189
	Collagen α1(III)	Cattle	P04258	6	158
	Collagen α1(II)	Human	P02458	7	136
	Collagen α1(XI)	Mouse	Q61245	7	179
	Collagen α1(V)	Human	P20908	5	109
	SPARC	5 mammal	*	2	95
	Serum albumin	Pig	P08835	2	71
	Phosphate carrier protein	Cattle	P12234	1	65
Chicken Bone	Collagen α2(I)	Chicken	P02467	33	2374
	Collagen α1(I)	Chicken	P02457	44	2253
	Collagen α1(II)	Human	P02458	8	179
	Collagen α1(III)	Human	P02461	5	126
	Haemoglobin subunit α-D	6 avian	*	1	125
	Haemoglobin subunit beta	11 avian	*	2	105
	Apolipoprotein A-I	Chicken	P08250	2	104
	Collagen α2 (V)	Human	P05997	3	70
	Haemoglobin subunit α-A	10 avian	*	1	62

It is clear from Table 4.1 that many more peptides matches are to collagen (I) in the LC-MS analyses of bone samples from all four species than to any other protein. Additional identification of collagen types II, III, V, VII, and XI vary throughout the four species and resulted in much fewer matches than those to collagen (I). However these identifications could be misidentifications of collagen (I) sequences by the Mascot search engine due to their similarities, where difficulties in inferring protein identities given a particular set of matched peptides are known to arise (at least in the case of higher eukaryote organisms) because of sequence redundancy, particularly in different proteins having a high degree of sequence homology, as in the case with protein families (Nesvizhskii and Aebersold 2005), such as these collagens. In addition to collagen, the mammalian bone samples all exhibit trace amounts of various NCPs, including OC, SPARC, A2HS and serum albumin, biglycan and prothrombin (brief descriptions of which are presented in Appendix 3.4). The presence of these NCPs is consistent with their reported associations with bone collagen (I) and bone mineral (see description in Appendix 3.4 for associated references). Interestingly, the only avian sample (chicken) did not result in any matches to any of these NCPs, with the only non-collagenous proteins identified being to the blood protein haemoglobin. This could reflect differences in tissue structure, where the greater relative abundance of haemoglobin was masking the presence of the other NCPs in the MALDI-MS analysis.

Table 4.2 - Summary of the LC-MALDI analyses describing collagen sequence coverage (omitting procollagen sequence), top species identifications and the number of unmatched peptides.

Sample	**% Coverage (α1/ α2)**	**Top Hits**	**No. Unidentifed peptides**
Modern Cattle Bone Sample 1	56/63	Cattle/Cattle	88
Modern Cattle Bone Sample 2	60/68	Cattle/Cattle	191
Modern Cattle Bone Sample 3	39/48	Cattle/Cattle	240
Modern Sheep Bone	56/63	Cattle/Cattle	107
Modern Pig Bone	47/31	Human/Cattle	138
Modern Chicken Bone	50/44	Chicken/Chicken	102

The list of protein identifications (Table 4.1) shows that the two most dominant protein matches were to the two alpha chains of collagen (I). Interestingly, the less abundant α2 (I) chain is often (except in the chicken and pig bone samples) observed to have greater

sequence coverages than the twice as abundant α1 (I) chain (see section 1.6.5). The observation that fewer α2 (I) chain peptides were matched to pig collagen than its α1 (I) chain counterpart could be due to the lack of a more closely related species' sequence in the public database than the cattle sequence (sheep sequences likely closely match cattle sequences based on the high coverage of cattle collagen peptides observed in the LC-MS results), as this chain was identified in section 1.6.5 as being much more variable. Because the annotated genomes for pig and sheep were not published, pig and sheep MBM samples were not further considered in the LC-MS analyses for this chapter. In order to investigate reproducibility, tryptic digests of the acid-insoluble residues from three different aliquots of the same cattle bone powder sample were analysed by LC-MALDI-MS/MS in consecutive analyses. In the collagen (I) alpha chain sequence coverages of the replicate analyses, even in samples which appear to have similar overall percentage coverages (see Table 4.2), the actual peptides matched appear to differ somewhat, particularly with α2 (I) chain peptides (see Fig. 4.2 Cattle Bone Sample 1 and Cattle Bone Sample 2). One of the three sample replicates (Cattle Bone Sample 3) shows a surprisingly poor coverage of both alpha chain sequences (Fig. 4.2), no NCP identifications (Table 4.1) and a relatively large number of unidentified peptides (Table 4.2). Although run at approximately the same time (consecutive runs), this could still be due to instrumental variation or technical problems (eg, HPLC probot needle becoming clogged whilst using an increasingly crystalised matrix over time). Alternatively, the variation observed in the LC-MS analyses of collagen (I) could be due to the within-sample heterogeneity of extractable collagen molecules where a single collagen extract contains collagen molecules of varying ages (and thus varying extent of cross-linking) (Jackson and Bentley 1960) that can potentially reduce the comparability between samples.

The comparison of standard scoring with MudPIT scoring (Appendix 3.3) highlighted that there were more protein identifications in the standard scoring results, many of which could be expected to be present in such tissues, although there were additional protein matches that are less realistic. Although MudPIT scoring could be considered appropriate for the bone samples because they generate a lot of spectra (~1000-1500 MS/MS spectra each), the proteins present in the MBM samples are much more degraded, resulting in many fewer MS/MS spectra of good quality. Because the protein score in MudPIT scoring Mascot searches does not include the ion scores from peptides that are below the threshold,

standard scoring is more suitable to the analyses of MBM samples that are more likely to contain more poor MS/MS spectra than non-rendered samples. This is particularly important where the mixture of proteins in MBM likely contains several small proteins such as OC and myoglobin, which may only be identified by a few peptides (contrary to standard scoring, in MudPIT scoring, a few matched peptides may not result in a protein match if the majority are below the threshold). For these reasons, the results obtained from using standard scoring are presented throughout this chapter (Table 4.3 and Table 4.4).

Cattle Sample 1 α1 (I)

```
   1 MFSFVDLRLL LLLAATALLT HGQEEGQEEG QEEDIPPVTC VQNGLRYHDR
  51 DVWKPVPCQI CVCDNGNVLC DDVICDELKD CPNAKVPTDE CCPVCPEGQE
 101 SPTDQETTGV EGPKGDTGPR GPRGPAGPPG RDGIPGQPGL PGPPGPPGPP
 151 GPPGLGGNFA PQLSYGYDEK STGISVPGPM GPSGPRGLPG PPGAPGPQGF
 201 QGPPGEPGEP GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG
 251 ARGLPGTAGL PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM
 301 GPRGLPGERG RPGAPGPAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA
 351 KGEGGPQGPR GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG
 401 APGIAGAPGF PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP
 451 GPTGIQGPPG PAGEEGKRGA RGEPGPAGLP GPPGERGGPG SRGFPGADGV
 501 AGPKGPAGER GAPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG
 551 KTGPPGPAGQ DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP
 601 GPPGAVGPAG KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP
 651 PGEAGKPGEQ GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG
 701 APGNDGAKGD AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK
 751 GADGAPGKDG VRGLTGPIGP PGPAGAPGDK GEAGPSGPAG PTGARGAPGD
 801 RGEPGPPGPA GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPAGPPG
 851 PIGNVGAPGP KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA
 901 GKEGSKGPRG ETGPAGRPGE VGPPGPPGPA GEKGAPGADG PAGAPGTPGP
 951 QGIAGQRGVV GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG
1001 PPGLAGPPGE SGREGAPGAE GSPGRDGSPG AKGDRGETGP AGPPGAPGAP
1051 GAPGPVGPAG KSGDRGETGP AGPAGPIGPV GARGPAGPQG PRGDKGETGE
1101 QGDRGIKGHR GFSGLQGPPG PPGSPGEQGP SGASGPAGPR GPPGSAGSPG
1151 KDGLNGLPGP IGPPGPRGRT GDAGPAGPPG PPGPPGPPGP PSGGYDLSFL
1201 PQPPQEKAHD GGRYYRADDA NVVRDRDLEV DTTLKSLSQQ IENIRSPEGS
1251 RKNPARTCRD LKMCHSDWKS GEYWIDPNQG CNLDAIKVFC NMETGETCVY
1301 PTQPSVAQKN WYISKNPKEK RHVWYGESMT GGFQFEYGGQ GSDPADVAIQ
1351 LTFLRLMSTE ASQNITYHCK NSVAYMDQQT GNLKKALLLQ GSNEIEIRAE
1401 GNSRFTYSVT YDGCTSHTGA WGKTVIEYKT TKTSRLPIID VAPLDVGAPD
1451 QEFGFDVGPA CFL
```

Cattle Sample 1 α2 (I)

```
   1 MLSFVDTRTL LLLAVTSCLA TCQSLQEATA RKGPSGDRGP RGERGPPGPP
  51 GRDGDDGIPG PPGPPGPPGP PGLGGNFAAQ FDAKGGGPGP MGLMGPRGPP
 101 GASGAPGPQG FQGPPGEPGE PGQTGPAGAR GPPGPPGKAG EDGHPGKPGR
 151 PGERGVVGPQ GARGFPGTPG LPGFKGIRGH NGLDGLKGQP GAPGVKGEPG
 201 APGENGTPGQ TGARGLPGER GRVGAPGPAG ARGSDGSVGP VGPAGPIGSA
 251 GPPGFPGAPG PKGELGPVGN PGPAGPAGPR GEVGLPGLSG PVGPPGNPGA
 301 NGLPGAKGAA GLPGVAGAPG LPGPRGIPGP VGAAGATGAR GLVGEPGPAG
 351 SKGESGNKGE PGAVGQPGPP GPSGEEGKRG STGEIGPAGP PGPPGLRGNP
 401 GSRGLPGADG RAGVMGPAGS RGATGPAGVR GPNGDSGRPG EPGLMGPRGF
 451 PGSPGNIGPA GKEGPVGLPG IDGRPGPIGP AGARGEPGNI GFPGPKGPSG
 501 DPGKAGEKGH AGLAGARGAP GPDGNNGAQG PPGLQGVQGG KGEQGPAGPP
 551 GFQGLPGPAG TAGEAGKPGE RGIPGEFGLP GPAGARGERG PPGESGAAGP
 601 TGPIGSRGPS GPPGPDGNKG EPGVVGAPGT AGPSGPSGLP GERGAAGIPG
 651 GKGEKGETGL RGDIGSPGRD GARGAPGAIG APGPAGANGD RGEAGPAGPA
 701 GPAGPRGSPG ERGEVGPAGP NGFAGPAGAA GQPGAKGERG TKGPKGENGP
 751 VGPTGPVGAA GPSGPNGPPG PAGSRGDGGP PGATGFPGAA GRTGPPGPSG
 801 ISGPPGPPGP AGKEGLRGPR GDQGPVGRSG ETGASGPPGF VGEKGPSGEP
 851 GTAGPPGTPG PQGLLGAPGF LGLPGSRGER GLPGVAGSVG EPGPLGIAGP
 901 PGARGPPGNV GNPGVNGAPG EAGRDGNPGN DGPPGRDGQP GHKGERGYPG
 951 NAGPVGAAGA PGPQGPVGPV GKHGNRGEPG PAGAVGPAGA VGPRGPSGPQ
1001 GIRGDKGEPG DKGPRGLPGL KGHNGLQGLP GLAGHHGDQG APGAVGPAGP
1051 RGPAGPSGPA GKDGRIGQPG AVGPAGIRGS QGSQGPAGPP GPPGPPGPPG
1101 PSGGGYEFGF DGDFYRADQP RSPTSLRPKD YEVDATLKSL NNQIETLLTP
1151 EGSRKNPART CRDLRLSHPE WSSGYYWIDP NQGCTMDAIK VYCDFSTGET
1201 CIRAQPEDIP VKNWYRNSKA KKHVWVGETI NGGTQFEYNV EGVTTKEMAT
1251 QLAFMRLLAN HASQNITYHC KNSIAYMDEE TGNLKKAVIL QGSNDVELVA
1301 EGNSRFTYTV LVDGCSKKTN EWQKTIIEYK TNKPSRLPIL DIAPLDIGGA
1351 DQEIRLNIGP VCFK
```

Cattle Sample 2 α1 (I)

```
   1 MFSFVDLRLL LLLAATALLT HGQEEGQEEG QEEDIPPVTC VQNGLRYHDR
  51 DVWKPVPCQI CVCDNGNVLC DDVICDELKD CPNAKVPTDE CCPVCPEGQE
 101 SPTDQETTGV EGPKGDTGPR GPRGPAGPPG RDGIPGQPGL PGPPGPPGPP
 151 GPPGLGGNFA PQLSYGYDEK STGISVPGPM GPSGPRGLPG PPGAPGPQGF
 201 QGPPGEPGEP GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG
 251 ARGLPGTAGL PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM
 301 GPRGLPGERG RPGAPGPAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA
 351 KGEGGPQGPR GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG
 401 APGIAGAPGF PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP
 451 GPTGIQGPPG PAGEEGKRGA RGEPGPAGLP GPPGERGGPG SRGFPGADGV
 501 AGPKGPAGER GAPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG
 551 KTGPPGPAGQ DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP
 601 GPPGAVGPAG KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP
 651 PGEAGKPGEQ GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG
 701 APGNDGAKGD AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK
 751 GADGAPGKDG VRGLTGPIGP PGPAGAPGDK GEAGPSGPAG PTGARGAPGD
 801 RGEPGPPGPA GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPAGPPG
 851 PIGNVGAPGP KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA
 901 GKEGSKGPRG ETGPAGRPGE VGPPGPPGPA GEKGAPGADG PAGAPGTPGP
 951 QGIAGQRGVV GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG
1001 PPGLAGPPGE SGREGAPGAE GSPGRDGSPG AKGDRGETGP AGPPGAPGAP
1051 GAPGPVGPAG KSGDRGETGP AGPAGPIGPV GARGPAGPQG PRGDKGETGE
1101 QGDRGIKGHR GFSGLQGPPG PPGSPGEQGP SGASGPAGPR GPPGSAGSPG
1151 KDGLNGLPGP IGPPGPRGRT GDAGPAGPPG PPGPPGPPGP PSGGYDLSFL
1201 PQPPQEKAHD GGRYYRADDA NVVRDRDLEV DTTLKSLSQQ IENIRSPEGS
1251 RKNPARTCRD LKMCHSDWKS GEYWIDPNQG CNLDAIKVFC NMETGETCVY
1301 PTQPSVAQKN WYISKNPKEK RHVWYGESMT GGFQFEYGGQ GSDPADVAIQ
1351 LTFLRLMSTE ASQNITYHCK NSVAYMDQQT GNLKKALLLQ GSNEIEIRAE
1401 GNSRFTYSVT YDGCTSHTGA WGKTVIEYKT TKTSRLPIID VAPLDVGAPD
1451 QEFGFDVGPA CFL
```

Cattle Sample 2 α2 (I)

```
   1 MLSFVDTRTL LLLAVTSCLA TCQSLQEATA RKGPSGDRGP RGERGPPGPP
  51 GRDGDDGIPG PPGPPGPPGP PGLGGNFAAQ FDAKGGGPGP MGLMGPRGPP
 101 GASGAPGPQG FQGPPGEPGE PGQTGPAGAR GPPGPPGKAG EDGHPGKPGR
 151 PGERGVVGPQ GARGFPGTPG LPGFKGIRGH NGLDGLKGQP GAPGVKGEPG
 201 APGENGTPGQ TGARGLPGER GRVGAPGPAG ARGSDGSVGP VGPAGPIGSA
 251 GPPGFPGAPG PKGELGPVGN PGPAGPAGPR GEVGLPGLSG PVGPPGNPGA
 301 NGLPGAKGAA GLPGVAGAPG LPGPRGIPGP VGAAGATGAR GLVGEPGPAG
 351 SKGESGNKGE PGAVGQPGPP GPSGEEGKRG STGEIGPAGP PGPPGLRGNP
 401 GSRGLPGADG RAGVMGPAGS RGATGPAGVR GPNGDSGRPG EPGLMGPRGF
 451 PGSPGNIGPA GKEGPVGLPG IDGRPGPIGP AGARGEPGNI GFPGPKGPSG
 501 DPGKAGEKGH AGLAGARGAP GPDGNNGAQG PPGLQGVQGG KGEQGPAGPP
 551 GFQGLPGPAG TAGEAGKPGE RGIPGEFGLP GPAGARGERG PPGESGAAGP
 601 TGPIGSRGPS GPPGPDGNKG EPGVVGAPGT AGPSGPSGLP GERGAAGIPG
 651 GKGEKGETGL RGDIGSPGRD GARGAPGAIG APGPAGANGD RGEAGPAGPA
 701 GPAGPRGSPG ERGEVGPAGP NGFAGPAGAA GQPGAKGERG TKGPKGENGP
 751 VGPTGPVGAA GPSGPNGPPG PAGSRGDGGP PGATGFPGAA GRTGPPGPSG
 801 ISGPPGPPGP AGKEGLRGPR GDQGPVGRSG ETGASGPPGF VGEKGPSGEP
 851 GTAGPPGTPG PQGLLGAPGF LGLPGSRGER GLPGVAGSVG EPGPLGIAGP
 901 PGARGPPGNV GNPGVNGAPG EAGRDGNPGN DGPPGRDGQP GHKGERGYPG
 951 NAGPVGAAGA PGPQGPVGPV GKHGNRGEPG PAGAVGPAGA VGPRGPSGPQ
1001 GIRGDKGEPG DKGPRGLPGL KGHNGLQGLP GLAGHHGDQG APGAVGPAGP
1051 RGPAGPSGPA GKDGRIGQPG AVGPAGIRGS QGSQGPAGPP GPPGPPGPPG
1101 PSGGGYEFGF DGDFYRADQP RSPTSLRPKD YEVDATLKSL NNQIETLLTP
1151 EGSRKNPART CRDLRLSHPE WSSGYYWIDP NQGCTMDAIK VYCDFSTGET
1201 CIRAQPEDIP VKNWYRNSKA KKHVWVGETI NGGTQFEYNV EGVTTKEMAT
1251 QLAFMRLLAN HASQNITYHC KNSIAYMDEE TGNLKKAVIL QGSNDVELVA
1301 EGNSRFTYTV LVDGCSKKTN EWQKTIIEYK TNKPSRLPIL DIAPLDIGGA
1351 DQEIRLNIGP VCFK
```

Cattle Sample 3 α1 (I)

```
   1 MFSFVDLRLL LLLAATALLT HGQEEGQEEG QEEDIPPVTC VQNGLRYHDR
  51 DVWKPVPCQI CVCDNGNVLC DDVICDELKD CPNAKVPTDE CCPVCPEGQE
 101 SPTDQETTGV EGPKGDTGPR GPRGPAGPPG RDGIPGQPGL PGPPGPPGPP
 151 GPPGLGGNFA PQLSYGYDEK STGISVPGPM GPSGPRGLPG PPGAPGPQGF
 201 QGPPGEPGEP GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG
 251 ARGLPGTAGL PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM
 301 GPRGLPGERG RPGAPGPAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA
 351 KGEGGPQGPR GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG
 401 APGIAGAPGF PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP
 451 GPTGIQGPPG PAGEEGKRGA RGEPGPAGLP GPPGERGGPG SRGFPGADGV
 501 AGPKGPAGER GAPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG
 551 KTGPPGPAGQ DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP
 601 GPPGAVGPAG KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP
 651 PGEAGKPGEQ GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG
 701 APGNDGAKGD AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK
 751 GADGAPGKDG VRGLTGPIGP PGPAGAPGDK GEAGPSGPAG PTGARGAPGD
 801 RGEPGPPGPA GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPAGPPG
 851 PIGNVGAPGP KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA
 901 GKEGSKGPRG ETGPAGRPGE VGPPGPPGPA GEKGAPGADG PAGAPGTPGP
 951 QGIAGQRGVV GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG
1001 PPGLAGPPGE SGREGAPGAE GSPGRDGSPG AKGDRGETGP AGPPGAPGAP
1051 GAPGPVGPAG KSGDRGETGP AGPAGPIGPV GARGPAGPQG PRGDKGETGE
1101 QGDRGIKGHR GFSGLQGPPG PPGSPGEQGP SGASGPAGPR GPPGSAGSPG
1151 KDGLNGLPGP IGPPGPRGRT GDAGPAGPPG PPGPPGPPGP PSGGYDLSFL
1201 PQPPQEKAHD GGRYYRADDA NVVRDRDLEV DTTLKSLSQQ IENIRSPEGS
1251 RKNPARTCRD LKMCHSDWKS GEYWIDPNQG CNLDAIKVFC NMETGETCVY
1301 PTQPSVAQKN WYISKNPKEK RHVWYGESMT GGFQFEYGGQ GSDPADVAIQ
1351 LTFLRLMSTE ASQNITYHCK NSVAYMDQQT GNLKKALLLQ GSNEIEIRAE
1401 GNSRFTYSVT YDGCTSHTGA WGKTVIEYKT TKTSRLPIID VAPLDVGAPD
1451 QEFGFDVGPA CFL
```

Cattle Sample 3 α2 (I)

```
   1 MLSFVDTRTL LLLAVTSCLA TCQSLQEATA RKGPSGDRGP RGERGPPGPP
  51 GRDGDDGIPG PPGPPGPPGP PGLGGNFAAQ FDAKGGGPGP MGLMGPRGPP
 101 GASGAPGPQG FQGPPGEPGE PGQTGPAGAR GPPGPPGKAG EDGHPGKPGR
 151 PGERGVVGPQ GARGFPGTPG LPGFKGIRGH NGLDGLKGQP GAPGVKGEPG
 201 APGENGTPGQ TGARGLPGER GRVGAPGPAG ARGSDGSVGP VGPAGPIGSA
 251 GPPGFPGAPG PKGELGPVGN PGPAGPAGPR GEVGLPGLSG PVGPPGNPGA
 301 NGLPGAKGAA GLPGVAGAPG LPGPRGIPGP VGAAGATGAR GLVGEPGPAG
 351 SKGESGNKGE PGAVGQPGPP GPSGEEGKRG STGEIGPAGP PGPPGLRGNP
 401 GSRGLPGADG RAGVMGPAGS RGATGPAGVR GPNGDSGRPG EPGLMGPRGF
 451 PGSPGNIGPA GKEGPVGLPG IDGRPGPIGP AGARGEPGNI GFPGPKGPSG
 501 DPGKAGEKGH AGLAGARGAP GPDGNNGAQG PPGLQGVQGG KGEQGPAGPP
 551 GFQGLPGPAG TAGEAGKPGE RGIPGEFGLP GPAGARGERG PPGESGAAGP
 601 TGPIGSRGPS GPPGPDGNKG EPGVVGAPGT AGPSGPSGLP GERGAAGIPG
 651 GKGEKGETGL RGDIGSPGRD GARGAPGAIG APGPAGANGD RGEAGPAGPA
 701 GPAGPRGSPG ERGEVGPAGP NGFAGPAGAA GQPGAKGERG TKGPKGENGP
 751 VGPTGPVGAA GPSGPNGPPG PAGSRGDGGP PGATGFPGAA GRTGPPGPSG
 801 ISGPPGPPGP AGKEGLRGPR GDQGPVGRSG ETGASGPPGF VGEKGPSGEP
 851 GTAGPPGTPG PQGLLGAPGF LGLPGSRGER GLPGVAGSVG EPGPLGIAGP
 901 PGARGPPGNV GNPGVNGAPG EAGRDGNPGN DGPPGRDGQP GHKGERGYPG
 951 NAGPVGAAGA PGPQGPVGPV GKHGNRGEPG PAGAVGPAGA VGPRGPSGPQ
1001 GIRGDKGEPG DKGPRGLPGL KGHNGLQGLP GLAGHHGDQG APGAVGPAGP
1051 RGPAGPSGPA GKDGRIGQPG AVGPAGIRGS QGSQGPAGPP GPPGPPGPPG
1101 PSGGGYEFGF DGDFYRADQP RSPTSLRPKD YEVDATLKSL NNQIETLLTP
1151 EGSRKNPART CRDLRLSHPE WSSGYYWIDP NQGCTMDAIK VYCDFSTGET
1201 CIRAQPEDIP VKNWYRNSKA KKHVWVGETI NGGTQFEYNV EGVTTKEMAT
1251 QLAFMRLLAN HASQNITYHC KNSIAYMDEE TGNLKKAVIL QGSNDVELVA
1301 EGNSRFTYTV LVDGCSKKTN EWQKTIIEYK TNKPSRLPIL DIAPLDIGGA
1351 DQEIRLNIGP VCFK
```

Figure 4.2 - Sequence coverage maps of collagen a1 (I) and a2 (I) chains from three separate analyses of unheated cattle bone gelatine digests (% coverages are listed in Table 4.1). Grey shading indicates procollagen sequence that is not part of the tropocollagen molecule, red text indicates peptide matches in LC-MALDI analyses.

4.4.2 Acid-Insoluble Proteins in MBM

In order to compare the proteins surviving in MBM to those surviving in its sedimented components, samples of cattle 145°C-rendered MBM, chicken 145°C-rendered MBM, cattle sedimented 145°C-rendered sedimented MBM and chicken 145°C-rendered sedimented MBM 'acid-insoluble' residues were gelatinised, digested with trypsin and analysed by LC-MALDI-MS. The Mascot search results (from MS/MS only) were exported into Microsoft Excel using a significance threshold of $p<0.01$ (99% confidence) and each identification to require a bold red peptide (indicating the best protein match for that particular peptide). Tables 4.3 and 4.4 list a protein summary for MBM and MBM sediment respectively; showing protein identifications, accession numbers, number of peptide queries per protein and total ion scores for all unique protein matches (with a total ion score greater than 65). Further results of the Mascot searches are presented in Appendix 3.3. Collagen (I) sequence coverages of MBM and MBM sediment samples are shown in Figures 4.3 and 4.4 respectively.

Table 4.3 - Protein summary of digests from cattle and chicken MBM gelatine analysed by LC-MALDI using standard scoring.

** indicates several accession numbers (see Appendix 3.3 for entire list)*

Sample	Protein	Organism Identity	Acc. No	No. peptides	Total Ion Score
Cattle MBM 145°C	Collagen α1(I)	Cattle	P02453	49	1790
	Collagen α2(I)	Cattle	P02465	26	1346
	Actin, muscle	29 species	*	8	235
	Collagen α1(II)	Mouse	P28481	10	225
	Myosin heavy chain	7 avian to mammal species	*	6	203
	Haemoglobin subunit Beta	4 bovid species	*	5	154
	Haemoglobin subunit alpha	3 bovine species	*	2	124
	Myoglobin	6 mammal species	*	1	106
	Alpha-actinin-2	3 avian to mammal species	*	2	93
	RNA polymerase Beta chain	Marine bacterium	Q7URW6	3	74
	Serum albumin	Cattle	P02769	3	74
Chicken MBM 145°C	Collagen α1(I)	Chicken	P02457	29	1695
	Collagen α2(I)	Chicken	P02467	28	1374
	Myosin heavy chain (adult)	Chicken	P13538	19	805
	Myosin heavy chain (embryonic)	Chicken	P02565	18	743
	Myosin-1	Mouse & Human	Q5SX40 & P12882	14	618
	Actin	26 fish to mammal species	*	6	401
	Collagen α1 (II)	Rat & Mouse	P05539 & P28481	7	199
	Collagen α1 (III)	Chicken	P12105	6	156
	Haemoglobin subunit alpha-A	Chicken & Turkey	P01994 & P81023	1	133
	Glyceraldehyde-3-phosphate dehydrogenase	8 species	*	1	87
	Haemoglobin subunit α-D	Chicken & Turkey	P02001 & P81024	1	74
	Tropomyosin 1 alpha chain	7 fish to mammal species	*	2	72
	Troponin C, skeletal muscle	Turkey	P10246	2	68
	Creatine kinase M-type	Chicken	P00565	2	76
	Alpha-actinin-2	Chicken	P20111	4	75

Cattle MBM Sample α1 (I)

```
   1 MFSFVDLRLL LLLAATALLT HGQEEGQEEG QEEDIPPVTC VQNGLRYHDR
  51 DVWKPVPCQI CVCDNGNVLC DDVICDELKD CPNAKVPTDE CCPVCPEGQE
 101 SPTDQETTGV EGPKGDTGPR GPRGPAGPPG RDGIPGQPGL PGPPGPPGPP
 151 GPPGLGGNFA PQLSYGYDEK STGISVPGPM GPSGPRGLPG PPGAPGPQGF
 201 QGPPGEPGEP GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG
 251 ARGLPGTAGL PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM
 301 GPRGLPGERG RPGAPGPAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA
 351 KGEGGPQGPR GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG
 401 APGIAGAPGF PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP
 451 GPTGIQGPPG PAGEEGKRGA RGEPGPAGLP GPPGERGGPG SRGFPGADGV
 501 AGPKGPAGER GAPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG
 551 KTGPPGPAGQ DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP
 601 GPPGAVGPAG KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP
 651 PGEAGKPGEQ GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG
 701 APGNDGAKGD AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK
 751 GADGAPGKDG VRGLTGPIGP PGPAGAPGDK GEAGPSGPAG PTGARGAPGD
 801 RGEPGPPGPA GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPAGPPG
 851 PIGNVGAPGP KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA
 901 GKEGSKGPRG ETGPAGRPGE VGPPGPPGPA GEKGAPGADG PAGAPGTPGP
 951 QGIAGQRGVV GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG
1001 PPGLAGPPGE SGREGAPGAE GSPGRDGSPG AKGDRGETGP AGPPGAPGAP
1051 GAPGPVGPAG KSGDRGETGP AGPAGPIGPV GARGPAGPQG PRGDKGETGE
1101 QGDRGIKGHR GFSGLQGPPG PPGSPGEQGP SGASGPAGPR GPPGSAGSPG
1151 KDGLNGLPGP IGPPGPRGRT GDAGPAGPPG PPGPPGPPGP PSGGYDLSFL
1201 PQPPQEKAHD GGRYYRADDA NVVRDRDLEV DTTLKSLSQQ IENIRSPEGS
1251 RKNPARTCRD LKMCHSDWKS GEYWIDPNQG CNLDAIKVFC NMETGETCVY
1301 PTQPSVAQKN WYISKNPKEK RHVWYGESMT GGFQFEYGGQ GSDPADVAIQ
1351 LTFLRLMSTE ASQNITYHCK NSVAYMDQQT GNLKKALLLQ GSNEIEIRAE
1401 GNSRFTYSVT YDGCTSHTGA WGKTVIEYKT TKTSRLPIID VAPLDVGAPD
1451 QEFGFDVGPA CFL
```

Cattle MBM Sample α2 (I)

```
   1 MLSFVDTRTL LLLAVTSCLA TCQSLQEATA RKGPSGDRGP RGERGPPGPP
  51 GRDGDDGIPG PPGPPGPPGP PGLGGNFAAQ FDAKGGGPGP MGLMGPRGPP
 101 GASGAPGPQG FQGPPGEPGE PGQTGPAGAR GPPGPPGKAG EDGHPGKPGR
 151 PGERGVVGPQ GARGFPGTPG LPGFKGIRGH NGLDGLKGQP GAPGVKGEPG
 201 APGENGTPGQ TGARGLPGER GRVGAPGPAG ARGSDGSVGP VGPAGPIGSA
 251 GPPGFPGAPG PKGELGPVGN PGPAGPAGPR GEVGLPGLSG PVGPPGNPGA
 301 NGLPGAKGAA GLPGVAGAPG LPGPRGIPGP VGAAGATGAR GLVGEPGPAG
 351 SKGESGNKGE PGAVGQPGPP GPSGEEGKRG STGEIGPAGP PGPPGLRGNP
 401 GSRGLPGADG RAGVMGPAGS RGATGPAGVR GPNGDSGRPG EPGLMGPRGF
 451 PGSPGNIGPA GKEGPVGLPG IDGRPGPIGP AGARGEPGNI GFPGPKGPSG
 501 DPGKAGEKGH AGLAGARGAP GPDGNNGAQG PPGLQGVQGG KGEQGPAGPP
 551 GFQGLPGPAG TAGEAGKPGE RGIPGEFGLP GPAGARGERG PPGESGAAGP
 601 TGPIGSRGPS GPPGPDGNKG EPGVVGAPGT AGPSGPSGLP GERGAAGIPG
 651 GKGEKGETGL RGDIGSPGRD GARGAPGAIG APGPAGANGD RGEAGPAGPA
 701 GPAGPRGSPG ERGEVGPAGP NGFAGPAGAA GQPGAKGERG TKGPKGENGP
 751 VGPTGPVGAA GPSGPNGPPG PAGSRGDGGP PGATGFPGAA GRTGPPGPSG
 801 ISGPPGPPGP AGKEGLRGPR GDQGPVGRSG ETGASGPPGF VGEKGPSGEP
 851 GTAGPPGTPG PQGLLGAPGF LGLPGSRGER GLPGVAGSVG EPGPLGIAGP
 901 PGARGPPGNV GNPGVNGAPG EAGRDGNPGN DGPPGRDGQP GHKGERGYPG
 951 NAGPVGAAGA PGPQGPVGPV GKHGNRGEPG PAGAVGPAGA VGPRGPSGPQ
1001 GIRGDKGEPG DKGPRGLPGL KGHNGLQGLP GLAGHHGDQG APGAVGPAGP
1051 RGPAGPSGPA GKDGRIGQPG AVGPAGIRGS QGSQGPAGPP GPPGPPGPPG
1101 PSGGGYEFGF DGDFYRADQP RSPTSLRPKD YEVDATLKSL NNQIETLLTP
1151 EGSRKNPART CRDLRLSHPE WSSGYYWIDP NQGCTMDAIK VYCDFSTGET
1201 CIRAQPEDIP VKNWYRNSKA KKHVWVGETI NGGTQFEYNV EGVTTKEMAT
1251 QLAFMRLLAN HASQNITYHC KNSIAYMDEE TGNLKKAVIL QGSNDVELVA
1301 EGNSRFTYTV LVDGCSKKTN EWQKTIIEYK TNKPSRLPIL DIAPLDIGGA
1351 DQEIRLNIGP VCFK
```

Chicken MBM Sample α1 (I)

```
   1 MFSFVDSRLL LLIAATVLLT RGEGEEDIQT GSCVQDGLTY NDKDVWKPEP
  51 CQICVCDSGN ILCDEVICED TSDCPNAEIP FGECCPICPD VDASPVYPES
 101 AGVEGPKGDT GPRGDRGLPG PPGRDGIPGQ PGLPGPPGPP GPPGLGGNFA
 151 PQMSYGYDEK SAGVAVPGPM GPAGPRGLPG PPGAPGPQGF QGPPGEPGEP
 201 GASGPMGPRG PAGPPGKNGD DGEAGKPGRP GQRGPPGPQG ARGLPGTAGL
 251 PGMKGHRGFS GLDGAKGQPG PAGPKGEPGS PGENGAPGQM GPRGLPGERG
 301 RPGPSGPAGA RGNDGAPGAA GPPGPTGPAG PPGFPGAAGA KGETGPQGAR
 351 GSEGPQGSRG EPGPPGPAGA AGPAGNPGAD GQPGAKGATG APGIAGAPGF
 401 PGARGPSGPQ GPSGAPGPKG NSGEPGAPGN KGDTGAKGEP GPAGVQGPPG
 451 PAGEEGKRGA RGEPGPAGLP GPAGERGAPG SRGFPGADGI AGPKGPPGER
 501 GSPGAVGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG KTGPPGPAGQ
 551 DGRPGPAGPP GARGQAGVMG FPGPKGAAGE PGKPGERGAP GPPGAVGAAG
 601 KDGEAGAQGP PGPTGPAGER GEQGPAGAPG FQGLPGPAGP PGEAGKPGEQ
 651 GVPGNAGAPG PAGARGERGF PGERGVQGPP GPQGPRGANG APGNDGAKGD
 701 AGAPGAPGNE GPPGLEGMPG ERGAAGLPGA KGDRGDPGPK GADGAPGKDG
 751 LRGLTGPIGP PGPAGAPGDK GEAGPPGPAG PTGARGAPGD RGEPGPPGPA
 801 GFAGPPGADG QPGAKGETGD AGAKGDAGPP GPAGPTGAPG PAGZVGAPGP
 851 KGARGSAGPP GATGFPGAAG RVGPPGPSGN IGLPGPPGPA GKZGSKGPRG
 901 ETGPAGRPGE PGPAGPPGPP GEKGSPGADG PIGAPGTPGP QGIAGQRGVV
 951 GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG PPGLAGPPGE
1001 AGREGAPGAE GAPGRDGAAG PKGDRGETGP AGPPGAPGAP GAPGPVGPAG
1051 KNGDRGETGP AGPAGPPGPA GARGPAGPQG PRGDKGETGE QGDRGMKGHR
1101 GFSGLQGPPG PPGAPGEQGP SGASGPAGPR GPPGSAGAAG KDGLNGLPGP
1151 IGPPGPRGRT GEVGPVGPPG PPGPPGPPGP PSGGFDFSFL PQPPQEKAHD
1201 GGRYYRADDA NVMRDRDLEV DTTLKSLSQQ IENIRSPEGT RKNPARTCRD
1251 LKMCHGDWKS GEYWIDPNQG CNLDAIKVYC NMETGETCVY PTQATIAQKN
1301 WYLSKNPKEK KHVWFGETMS DGFQFEYGGE GSNPADVAIQ LTFLRLMSTE
1351 ATQNVTYHCK NSVAYMDHDT GNLKKALLLQ GANEIEIRAE GNSRFTYGVT
1401 EDGCTSHTGA WGKTVIEYKT TKTSRLPIID LAPMDVGAPD QEFGIDIGPV
1451 CFL
```

Chicken MBM Sample α2 (I)

```
   1 MLSFVDTRTL LLLAVTSCLA TCQSLQEATA RKGPSGDRGP RGERGPPGPP
  51 GRDGDDGIPG PPGPPGPPGP PGLGGNFAAQ FDAKGGGPGP MGLMGPRGPP
 101 GASGAPGPQG FQGPPGEPGE PGQTGPAGAR GPPGPPGKAG EDGHPGKPGR
 151 PGERGVVGPQ GARGFPGTPG LPGFKGIRGH NGLDGLKGQP GAPGVKGEPG
 201 APGENGTPGQ TGARGLPGER GRVGAPGPAG ARGSDGSVGP VGPAGPIGSA
 251 GPPGFPGAPG PKGELGPVGN PGPAGPAGPR GEVGLPGLSG PVGPPGNPGA
 301 NGLPGAKGAA GLPGVAGAPG LPGPRGIPGP VGAAGATGAR GLVGEPGPAG
 351 SKGESGNKGE PGAVGQPGPP GPSGEEGKRG STGEIGPAGP PGPPGLRGNP
 401 GSRGLPGADG RAGVMGPAGS RGATGPAGVR GPNGDSGRPG EPGLMGPRGF
 451 PGSPGNIGPA GKEGPVGLPG IDGRPGPIGP AGARGEPGNI GFPGPKGPSG
 501 DPGKAGEKGH AGLAGARGAP GPDGNNGAQG PPGLQGVQGG KGEQGPAGPP
 551 GFQGLPGPAG TAGEAGKPGE RGIPGEFGLP GPAGARGERG PPGESGAAGP
 601 TGPIGSRGPS GPPGPDGNKG EPGVVGAPGT AGPSGPSGLP GERGAAGIPG
 651 GKGEKGETGL RGDIGSPGRD GARGAPGAIG APGPAGANGD RGEAGPAGPA
 701 GPAGPRGSPG ERGEVGPAGP NGFAGPAGAA GQPGAKGERG TKGPKGENGP
 751 VGPTGPVGAA GPSGPNGPPG PAGSRGDGGP PGATGFPGAA GRTGPPGPSG
 801 ISGPPGPPGP AGKEGLRGPR GDQGPVGRSG ETGASGPPGF VGEKGPSGEP
 851 GTAGPPGTPG PQGLLGAPGF LGLPGSRGER GLPGVAGSVG EPGPLGIAGP
 901 PGARGPPGNV GNPGVNGAPG EAGRDGNPGN DGPPGRDGQP GHKGERGYPG
 951 NAGPVGAAGA PGPQGPVGPV GKHGNRGEPG PAGAVGPAGA VGPRGPSGPQ
1001 GIRGDKGEPG DKGPRGLPGL KGHNGLQGLP GLAGHHGDQG APGAVGPAGP
1051 RGPAGPSGPA GKDGRIGQPG AVGPAGIRGS QGSQGPAGPP GPPGPPGPPG
1101 PSGGGYEFGF DGDFYRADQP RSPTSLRPKD YEVDATLKSL NNQIETLLTP
1151 EGSRKNPART CRDLRLSHPE WSSGYYWIDP NQGCTMDAIK VYCDFSTGET
1201 CIRAQPEDIP VKNWYRNSKA KKHVWVGETI NGGTQFEYNV EGVTTKEMAT
1251 QLAFMRLLAN HASQNITYHC KNSIAYMDEE TGNLKKAVIL QGSNDVELVA
1301 EGNSRFTYTV LVDGCSKKTN EWQKTIIEYK TNKPSRLPIL DIAPLDIGGA
1351 DQEIRLNIGP VCFK
```

Figure 4.3 - Sequence coverage of cattle MBM collagen a1 (I) coverage (51%), a2 (I) coverage (33%) and chicken MBM collagen a1 (I) coverage (36%), a2 (I) coverage (39%) following tryptic digests and LC-MALDI MS/MS analyses of acid-insoluble residues. Grey shading indicates procollagen sequence that is not part of the tropocollagen molecule, red text indicates peptide matches in LC-MALDI analyses.

*Table 4.4 - Protein summary of digests from cattle and chicken raw MBM sediment gelatine analysed by LC-MALDI using standard scoring. * indicates several accession numbers (see Appendix 3.3 for detailed list)*

Sample	Protein	Organism Identity	Acc. No	No. peptides	Total Ion Score
Cattle MBM Sediment 145°C	Collagen α1(I)	Cattle	P02453	14	487
	Collagen α2(I)	Cattle	P02465	6	345
	Actin	25 fish to mammal species	*	5	226
	Haemoglobin subunit beta	6 bovid species	*	4	180
	Serum albumin	Cattle	P02769	4	105
	Myosin-1 (heavy chain)	4 avian to mammal species	*	2	94
Chicken MBM Sediment 145°C	Collagen α1(I)	Chicken	P02457	31	1913
	Collagen α2(I)	Chicken	P02467	26	1576
	Myosin heavy chain (adult)	Chicken	P13538	14	779
	Myosin heavy chain (embryonic)	Chicken	P02565	12	751
	Actin	29 fish to mammal species	*	7	477
	Alpha-actinin-2	3 avian to mammal species	*	7	318
	Collagen α1 (II)	Rat	P05539	8	263
	Collagen α1 (III)	Chicken	P12105	4	178
	Haemoglobin subunit α-D	Chicken & Turkey	P02001 & P81024	2	166
	Haemoglobin subunit α-A	Chicken	P01994	1	134
	Tropomyosin 1 alpha chain	4 fish to mammal species	*	3	129
	Fructose-bisphosphate aldolase C	3 avian to mammal species	*	1	125
	Arginine kinase	Prawn	P51545	2	101
	Beta-enolase	Chicken	P07322	1	90
	Glyceraldehyde-3-phosphate dehydrogenase	8 avian species	*	1	69

Cattle MBM sediment α1 (I)

```
   1 MFSFVDLRLL LLLAATALLT HGQEEGQEEG QEEDIPPVTC VQNGLRYHDR
  51 DVWKPVPCQI CVCDNGNVLC DDVICDELKD CPNAKVPTDE CCPVCPEGQE
 101 SPTDQETTGV EGPKGDTGPR GPRGPAGPPG RDGIPGQPGL PGPPGPPGPP
 151 GPPGLGGNFA PQLSYGYDEK STGISVPGPM GPSGPRGLPG PPGAPGPQGF
 201 QGPPGEPGEP GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG
 251 ARGLPGTAGL PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM
 301 GPRGLPGERG RPGAPGPAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA
 351 KGEGGPQGPR GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG
 401 APGIAGAPGF PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP
 451 GPTGIQGPPG PAGEEGKRGA RGEPGPAGLP GPPGERGGPG SRGFPGADGV
 501 AGPKGPAGER GAPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG
 551 KTGPPGPAGQ DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP
 601 GPPGAVGPAG KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP
 651 PGEAGKPGEQ GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG
 701 APGNDGAKGD AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK
 751 GADGAPGKDG VRGLTGPIGP PGPAGAPGDK GEAGPSGPAG PTGARGAPGD
 801 RGEPGPPGPA GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPAGPPG
 851 PIGNVGAPGP KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA
 901 GKEGSKGPRG ETGPAGRPGE VGPPGPPGPA GEKGAPGADG PAGAPGTPGP
 951 QGIAGQRGVV GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG
1001 PPGLAGPPGE SGREGAPGAE GSPGRDGSPG AKGDRGETGP AGPPGAPGAP
1051 GAPGPVGPAG KSGDRGETGP AGPAGPIGPV GARGPAGPQG PRGDKGETGE
1101 QGDRGIKGHR GFSGLQGPPG PPGSPGEQGP SGASGPAGPR GPPGSAGSPG
1151 KDGLNGLPGP IGPPGPRGRT GDAGPAGPPG PPGPPGPPGP PSGGYDLSFL
1201 PQPPQEKAHD GGRYYRADDA NVVRDRDLEV DTTLKSLSQQ IENIRSPEGS
1251 RKNPARTCRD LKMCHSDWKS GEYWIDPNQG CNLDAIKVFC NMETGETCVY
1301 PTQPSVAQKN WYISKNPKEK RHVWYGESMT GGFQFEYGGQ GSDPADVAIQ
1351 LTFLRLMSTE ASQNITYHCK NSVAYMDQQT GNLKKALLLQ GSNEIEIRAE
1401 GNSRFTYSVT YDGCTSHTGA WGKTVIEYKT TKTSRLPIID VAPLDVGAPD
1451 QEFGFDVGPA CFL
```

Cattle MBM sediment α2 (I)

```
   1 MLSFVDTRTL LLLAVTSCLA TCQSLQEATA RKGPSGDRGP RGERGPPGPP
  51 GRDGDDGIPG PPGPPGPPGP PGLGGNFAAQ FDAKGGGPGP MGLMGPRGPP
 101 GASGAPGPQG FQGPPGEPGE PGQTGPAGAR GPPGPPGKAG EDGHPGKPGR
 151 PGERGVVGPQ GARGFPGTPG LPGFKGIRGH NGLDGLKGQP GAPGVKGEPG
 201 APGENGTPGQ TGARGLPGER GRVGAPGPAG ARGSDGSVGP VGPAGPIGSA
 251 GPPGFPGAPG PKGELGPVGN PGPAGPAGPR GEVGLPGLSG PVGPPGNPGA
 301 NGLPGAKGAA GLPGVAGAPG LPGPRGIPGP VGAAGATGAR GLVGEPGPAG
 351 SKGESGNKGE PGAVGQPGPP GPSGEEGKRG STGEIGPAGP PGPPGLRGNP
 401 GSRGLPGADG RAGVMGPAGS RGATGPAGVR GPNGDSGRPG EPGLMGPRGF
 451 PGSPGNIGPA GKEGPVGLPG IDGRPGPIGP AGARGEPGNI GFPGPKGPSG
 501 DPGKAGEKGH AGLAGARGAP GPDGNNGAQG PPGLQGVQGG KGEQGPAGPP
 551 GFQGLPGPAG TAGEAGKPGE RGIPGEFGLP GPAGARGERG PPGESGAAGP
 601 TGPIGSRGPS GPPGPDGNKG EPGVVGAPGT AGPSGPSGLP GERGAAGIPG
 651 GKGEKGETGL RGDIGSPGRD GARGAPGAIG APGPAGANGD RGEAGPAGPA
 701 GPAGPRGSPG ERGEVGPAGP NGFAGPAGAA GQPGAKGERG TKGPKGENGP
 751 VGPTGPVGAA GPSGPNGPPG PAGSRGDGGP PGATGFPGAA GRTGPPGPSG
 801 ISGPPGPPGP AGKEGLRGPR GDQGPVGRSG ETGASGPPGF VGEKGPSGEP
 851 GTAGPPGTPG PQGLLGAPGF LGLPGSRGER GLPGVAGSVG EPGPLGIAGP
 901 PGARGPPGNV GNPGVNGAPG EAGRDGNPGN DGPPGRDGQP GHKGERGYPG
 951 NAGPVGAAGA PGPQGPVGPV GKHGNRGEPG PAGAVGPAGA VGPRGPSGPQ
1001 GIRGDKGEPG DKGPRGLPGL KGHNGLQGLP GLAGHHGDQG APGAVGPAGP
1051 RGPAGPSGPA GKDGRIGQPG AVGPAGIRGS QGSQGPAGPP GPPGPPGPPG
1101 PSGGGYEFGF DGDFYRADQP RSPTSLRPKD YEVDATLKSL NNQIETLLTP
1151 EGSRKNPART CRDLRLSHPE WSSGYYWIDP NQGCTMDAIK VYCDFSTGET
1201 CIRAQPEDIP VKNWYRNSKA KKHVWVGETI NGGTQFEYNV EGVTTKEMAT
1251 QLAFMRLLAN HASQNITYHC KNSIAYMDEE TGNLKKAVIL QGSNDVELVA
1301 EGNSRFTYTV LVDGCSKKTN EWQKTIIEYK TNKPSRLPIL DIAPLDIGGA
1351 DQEIRLNIGP VCFK
```

Chicken MBM sediment α1 (I)

```
   1 MFSFVDSRLL LLIAATVLLT RGEGEEDIQT GSCVQDGLTY NDKDVWKPEP
  51 CQICVCDSGN ILCDEVICED TSDCPNAEIP FGECCPICPD VDASPVYPES
 101 AGVEGPKGDT GPRGDRGLPG PPGRDGIPGQ PGLPGPPGPP GPPGLGGNFA
 151 PQMSYGYDEK SAGVAVPGPM GPAGPRGLPG PPGAPGPQGF QGPPGEPGEP
 201 GASGPMGPRG PAGPPGKNGD DGEAGKPGRP GQRGPPGPQG ARGLPGTAGL
 251 PGMKGHRGFS GLDGAKGQPG PAGPKGEPGS PGENGAPGQM GPRGLPGERG
 301 RPGPSGPAGA RGNDGAPGAA GPPGPTGPAG PPGFPGAAGA KGETGPQGAR
 351 GSEGPQGSRG EPGPPGPAGA AGPAGNPGAD GQPGAKGATG APGIAGAPGF
 401 PGARGPSGPQ GPSGAPGPKG NSGEPGAPGN KGDTGAKGEP GPAGVQGPPG
 451 PAGEEGKRGA RGEPGPAGLP GPAGERGAPG SRGFPGADGI AGPKGPPGER
 501 GSPGAVGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG KTGPPGPAGQ
 551 DGRPGPAGPP GARGQAGVMG FPGPKGAAGE PGKPGERGAP GPPGAVGAAG
 601 KDGEAGAQGP PGPTGPAGER GEQGPAGAPG FQGLPGPAGP PGEAGKPGEQ
 651 GVPGNAGAPG PAGARGERGF PGERGVQGPP GPQGPRGANG APGNDGAKGD
 701 AGAPGAPGNE GPPGLEGMPG ERGAAGLPGA KGDRGDPGPK GADGAPGKDG
 751 LRGLTGPIGP PGPAGAPGDK GEAGPPGPAG PTGARGAPGD RGEPGPPGPA
 801 GFAGPPGADG QPGAKGETGD AGAKGDAGPP GPAGPTGAPG PAGZVGAPGP
 851 KGARGSAGPP GATGFPGAAG RVGPPGPSGN IGLPGPPGPA GKZGSKGPRG
 901 ETGPAGRPGE PGPAGPPGPP GEKGSPGADG PIGAPGTPGP QGIAGQRGVV
 951 GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG PPGLAGPPGE
1001 AGREGAPGAE GAPGRDGAAG PKGDRGETGP AGPPGAPGAP GAPGPVGPAG
1051 KNGDRGETGP AGPAGPPGPA GARGPAGPQG PRGDKGETGE QGDRGMKGHR
1101 GFSGLQGPPG PPGAPGEQGP SGASGPAGPR GPPGSAGAAG KDGLNGLPGP
1151 IGPPGPRGRT GEVGPVGPPG PPGPPGPPGP PSGGFDFSFL PQPPQEKAHD
1201 GGRYYRADDA NVMRDRDLEV DTTLKSLSQQ IENIRSPEGT RKNPARTCRD
1251 LKMCHGDWKS GEYWIDPNQG CNLDAIKVYC NMETGETCVY PTQATIAQKN
1301 WYLSKNPKEK KHVWFGETMS DGFQFEYGGE GSNPADVAIQ LTFLRLMSTE
1351 ATQNVTYHCK NSVAYMDHDT GNLKKALLLQ GANEIEIRAE GNSRFTYGVT
1401 EDGCTSHTGA WGKTVIEYKT TKTSRLPIID LAPMDVGAPD QEFGIDIGPV
1451 CFL
```

Chicken MBM sediment α2 (I)

```
   1 MLSFVDTRIL LLLAVTSYLA TSQHVSEASA GRKGPRGDKG PQGERGPPGP
  51 PGRDGEDGPP GPPGPPGPPG LGGNFAAQYD PSKAADFGPG PMGLMGPRGP
 101 PGASGPPGPP GFQGVPGEPG EPGQTGPQGP RGPPGPPGKA GEDGHPGKPG
 151 RPGERGVAGP QGARGFPGTP GPPGFKGIRG HNGLDGLTGQ PGAPGTKGEP
 201 GAPGENGTPG QPGARGLPGE RGRIGAPGPA GARGSDGSAG PTGPAXXXXX
 251 XXXXXXXXXX XXGEIGPAGN EGPTGPAGPR GEIGLPGSSG PVGPPGNPGA
 301 NGLPGAKGAA GLPGVAGAPG LPGPRGIPGP PGPAGPSGAR GLVGEPGPAG
 351 AKGESGNKGE PGAAGPPGPP GPSGEEGKRG SNGEPGSAGP PGPAGLRGEP
 401 GSRGLPGADG RAGVMGPAGN RGASGPVGAK GPNGDAGRPG EPGLMGPRXX
 451 XXXXXXXXXX XXXXXXGFPG ADGRVGPIGP AGNRGEPGNI GFPGPKGPTG
 501 EPGKPGEKGN VGLAGPRGAP GPEGNNGAQG PPGVTGNQGA KGETGPAGPP
 551 GFQGLPGPSG PAGEAGKPGE RGLHGEFGVP GPAGPRGERG LPGESGAVGP
 601 AGPIGSRGPS GPPGPDGNKG EPGNVGPAGA PGPAGPGGIP GERGVAGVPG
 651 GKGEKGAPGL RGDTGATGRD GARGLPGAIG APGPAGGAGD RGEGGPAGPA
 701 GPAGARGIPG ERGEPGPVGP SGFAGPPGAA GQPGAKGERG PKGPKGETGP
 751 TGAIGPIGAS GPPGPVGAAG PAGPRGDAGP PGMTGFPGAA GRVGPPGPAG
 801 ITGPPGPPGP AGKDGPRGLR GDVGPVGRTG EQGIAGPPGF AGEKGPSGEA
 851 GAAGPPGTPG PQGILGAPGI LGLPGSRGER GLPGIAGATG EPGPLGVSGP
 901 PGARGPSGPV GSPGPNGAPG EAGRDGNPGN DGPPGRDGAP GFKGERGAPG
 951 NPGPSGALGA PGPHGQVGPS GKPGNRGDPG PVGPVGPAGA FGPRGLAGPQ
1001 GPRGEKGEPG DKGHRGLPGL KGHNGLQGLP GLAGQHGDQG PPGNNGPAGP
1051 RGPPGPSGPP GKDGRNGLPG PIGPAGVRGS HGSQGPAGPP GPPGPPGPPG
1101 PNGGGYEVGF DAEYYRADQP SLRPKDYEVD ATLKTLNNQI ETLLTPEGSK
1151 KNPARTCRDL RLSHPEWSSG FYWIDPNQGC TADAIRAYCD FATGETCIHA
1201 SLEDIPTKTW YVSKNPKDKK HIWFGETING GTQFEYNGEG VTTKDMATQL
1251 AFMRLLANHA SQNITYHCKN SIAYMDEETG NLKKAVILQG SNDVELRAEG
1301 NSRFTFSVLV DGCSKKNNKW GKTIIEYRTN KPSRLPILDI APLDIGGADQ
1351 EFGLHIGPVC FK
```

Figure 4.4 - Sequence coverage of cattle MBM sediment collagen a1 (I) coverage (15%), a2 (I) coverage (8%) and chicken MBM sediment collagen a1 (I) coverage (36%), a2 (I) coverage (35%) following gelatine digests and LC-MS analyses. Grey shading indicates procollagen sequence not part of the tropocollagen molecule, red text indicates peptide matches in LC-MALDI analyses.

The dominance of collagen (I) was clear in the LC-MALDI MS analyses of MBM samples from both species, with additional peptide matches to collagen (II), and collagen (III) varying throughout the two species. These could be misidentifications of collagen (I) sequences due to their similarities (for example the sequence Gly-Ala-Hyp has a similar mass value to Gly-Pro-Ser) and poor understanding of PTMs in bone collagen. The only NCP observed in MBM that was also identified in the analyses of bone was serum albumin in the cattle sample. Haemoglobin was observed in both samples along with the muscle proteins

myosin, actin and alpha-actinin. Several other muscle proteins (myoglobin, tropomyosin and troponin) and other proteins that are not tissue-specific such as cullin, glyceralehyde-3-phosphate dehydrogenase and creatine kinase were also identified in one of the samples (cullin in the cattle MBM sample and the latter in the chicken MBM sample).

As with the MBM samples, the dominance of collagen (I) was also clear in the LC-MALDI MS analyses of MBM sediment samples from cattle and chicken, with additional matches to collagen types II, and III in the chicken samples. As with the unsedimented MBM samples, the only NCP observed in MBM sediment that was also identified in the analyses of bone was serum albumin (in the cattle sample only). Haemoglobin, actin and myosin were also observed in both sedimented samples with some of the other muscle proteins (namely tropomyosin, beta-enolase and alpha-actinin) only being observed in the chicken MBM sediment sample (as was glyceralehyde-3-phosphate dehydrogenase). The presence of these proteins is consistent with their associations with skeletal muscle as well as skeletal bone tissue, but indicates that sedimentation was not enough to separate bone-associated proteins from the muscle proteins also present in the original mixture.

4.4.3 Amino Acid Composition and Racemisation Analyses

The range of increasing rendering temperature cattle, pig, sheep and chicken samples (133°C, 137°C, 141°C & 145°C) were analysed for amino acid concentrations and extent of racemisation in order to investigate the thermal stability of the dominant proteins. Amino acid racemisation values were measured for each of the 'Set B-EFPRA' samples, including unsedimented MBM and sedimented MBM (Fig. 4.5). Principal component analyses of the total amino acid compositions of the MBM samples were carried out to investigate differences in overall composition of MBM with increasing rendering temperature (Fig. 4.5).

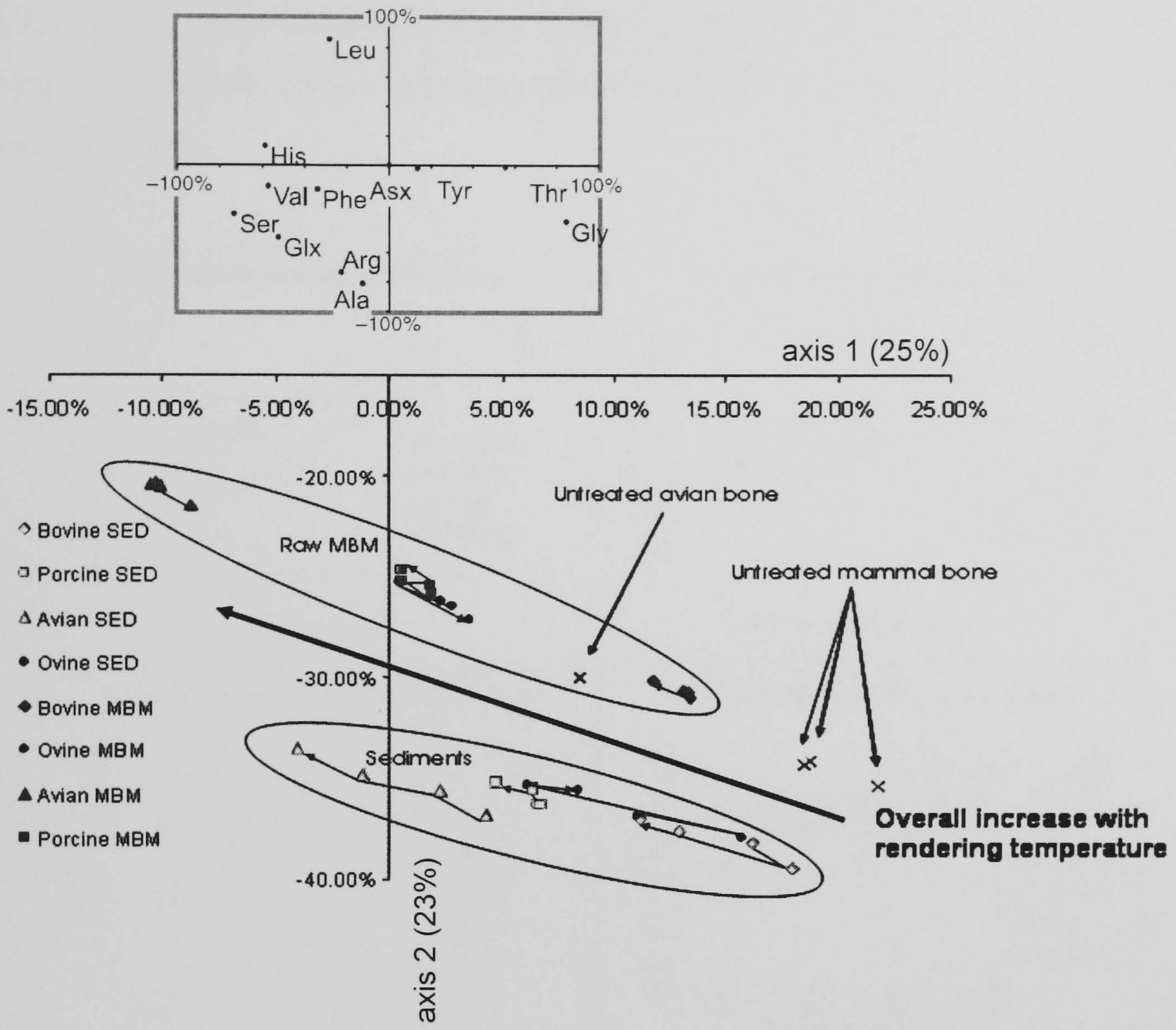


Figure 4.5 - Principal component analysis of total amino acid concentrations in untreated bone, MBM and MBM sediment (SED) samples using XLstat.

The amino acid composition of the MBM samples appears to be less influenced by glycine concentrations with increases in rendering temperature, and more influenced by the amounts of all other amino acids reported (see Fig. 4.5 inset), similar to previous work on heated bone (M. Collins unpublished data). However, the sedimented MBM samples form a distinct group from the unsedimented MBM samples that is more related with alanine and arginine concentrations than previously expected (Fig. 4.5), presumably due to additional proteins present in the unsedimented MBM. Figure 4.5 also highlights the substantial difference in composition between avian and mammalian samples which are off-set from each other in bone and MBM samples.

Asx DL values were plotted for each MBM and MBM sediment sample in order to investigate the effects of the increasing temperature rending processes on each of these samples (Fig. 4.6).

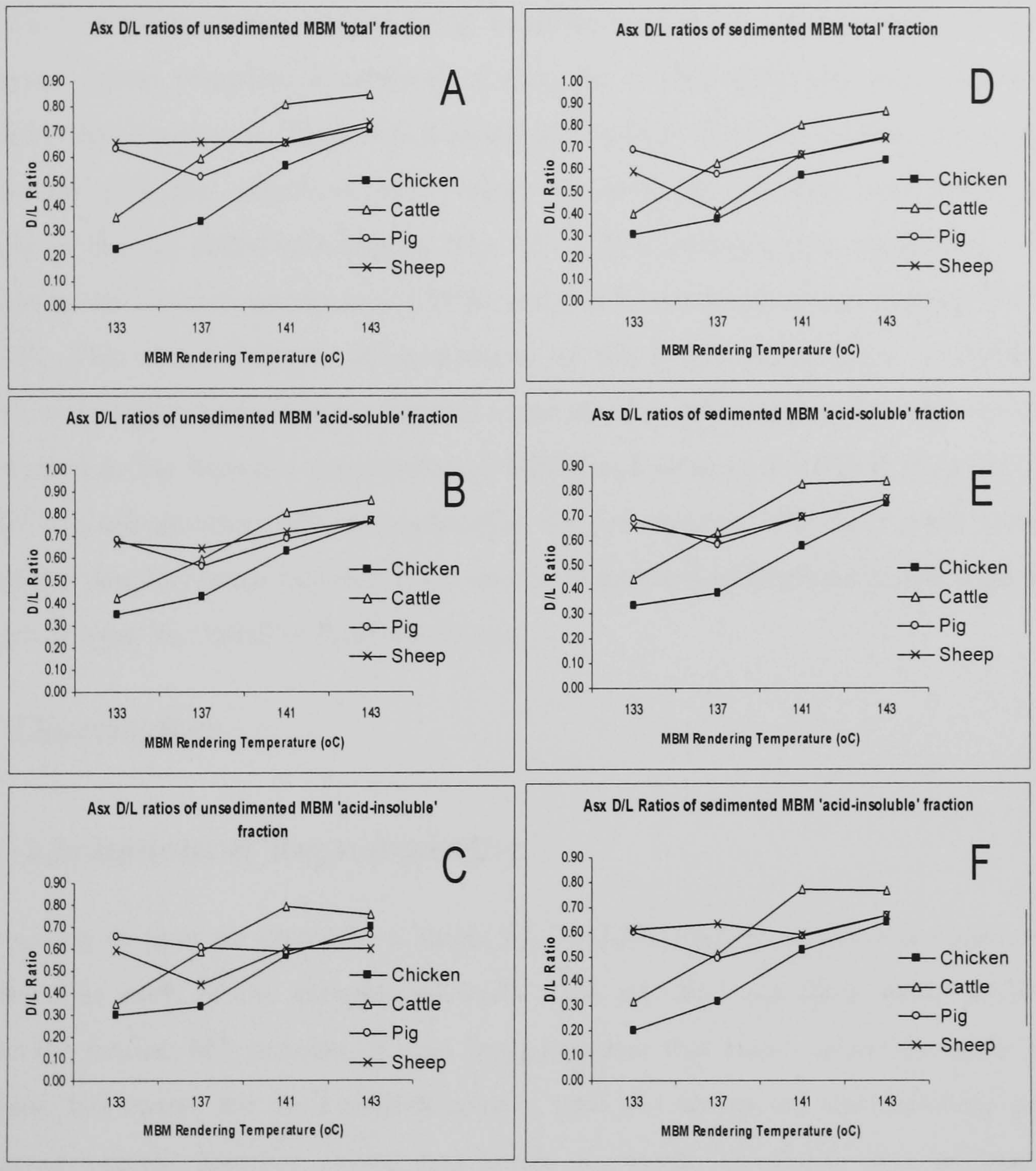


Figure 4.6 - Asx D/L values of both unsedimented (A,B,C) and sedimented (D,E,F) MBM for three different fractions; 'total', 'acid-soluble', and 'acid-insoluble'. Four species rendered to four different temperatures are shown in each. Data shown in Appendix 3.5.

The relationships between MBM rendering temperature and Asx DL ratios were observed to be similar in both the unsedimented and sedimented MBM (compare Figure 4.6 A-C, to D-F). The extent of degradation is also very similar throughout the different fractions, for

example, the sheep unsedimented MBM DL ratio of samples rendered at all four temperatures was approximately 0.7 (Fig. 4.6 A-C). The DL ratios of the acid-insoluble residues (Fig. 4.6 C & F) are far greater than expected for intact triple helical collagen (Hare 1980). This implies that, assuming the model of a constrained Asx racemisation by van Duin and Collins (1998) is correct, despite being insoluble in acid, the triple-helical molecules have undergone almost complete denaturation during the rendering process, and renatured into an acid-insoluble network. There was a linear relationship observed between the increase in temperature and the observed D/L value for both the chicken and cattle samples throughout the six different fractions (Fig. 4.6 A-F). However, this relationship was not apparent in the lower temperatures (133°C- and 137°C-rendered) sheep and pig MBM (Fig. 4.6 A-F). This could indicate contamination of the lower temperature rendered MBM samples with higher temperature rendered material. The only samples that did not follow a similar relationship between unsedimented MBM and sedimented MBM were the 133°C- and 137°C-rendered sheep MBM samples (Fig. 4.6 A-F crosses). The three other species had very similar Asx DL ratios between both unsedimented and sedimented MBM, regardless of the fraction type (ie., 'total' or 'acid-insoluble').

4.5 Discussion

4.5.1 Limitations & Reproducibility

The amount of data produced in a single LC-MALDI-MS/MS analysis is enormous and complex. For each of the samples analysed, there are data regarding peaks analysed by product ion tandem MS, number of peak fragmentations that match sequences in the protein databases, ion scores for each peptide match, total ion scores for the matching proteins, ranking of protein matches above and below a specific threshold, lists of unmatched peptides, etc. To simplify the vast amounts of data available, the information taken from the search result files was a list of the identified proteins (ie., Table 4.1), and in the case of collagen (I), their respective coverage sequences (ie., Fig. 4.1), although it is often desirable to use unmatched data such as the list of the unidentified peak m/z values that were analysed by product ion tandem MS (Appendix 3.3). One inherent problem with analysing these data presented as a Mascot search result file was that of the top 10 best matches, most of them

were to collagen sequences from the α1 (I) and α2 (I) chains of the several different species' sequences present in the protein databases used (UniProt). Ideally, these unique peptide matches that are identified as belonging to a collagen (I) sequence that is not the highest ranking match still should be considered, even with analyses on known samples with published collagen (I) sequences. The large amount of unidentified peaks is also an important source of information for species identification as many of the species of economic interest, and particularly in studies of ancient bone that do not have published collagen sequences. The large amount of unidentified peaks (only 50-80% sequence coverage of the two chain sequences are achieved) in the samples of known sequences could be associated with incorrect published collagen sequences, but is more likely due to either unknown modifications (such as additional glycations) or the presence of many NCPs that are likely associated with collagen but not matched (particularly as disulphide bridges (cysteine residues) were not reduced and alkylated; many NCPs contain disulphide bridges). The reason for why these peaks were not matched to any particular sequence should be further investigated as to whether it was due to the poor quality of the product ion tandem mass spectra, the databases lacking sequence information, or unknown and complicated PTMs (including the presence of disulphide bridges within NCPs or cross-links within collagen molecules). Data handling remains a key issue in the analysis of proteomic data and has been discussed thoroughly by many authors (Carr *et al.* 2004; Nesvizhskii and Aebersold 2005; Rappsilber and Mann 2002; Resing *et al.* 2004).

The dissimilarity of same-sample replicates, as emphasised by the clear differences in sequence coverage observed in two LC-MALDI-MS analyses of cattle bone collagen (I) of the same bone element (with 'collagen' gelatinisation, digestion, peptide separation and MS analysis on the same days); despite having similar % coverages, Sample 1 and Sample 2 exhibit clear differences in peptides that are present/absent. However, the aims of using LC-MALDI-MS herein were to investigate the most thermostable proteins that could survive in MBM, and hence long-term survival in ancient bone. This dissimilarity most likely reflects the heterogeneity of the samples and, for the reproducible study of bone collagen, further sample preparation would be required to remove the NCP fraction of bone.

4.5.2 Non-Collagenous Proteins

The association of NCPs, bone proteins as well as serum proteins, with the 'acid-insoluble' fraction of bone considered loosely to be pure collagen, is also of interest. For example, the survival of some of these NCPs in degraded bone would be useful due to the greater amounts of inter-species sequence variation than collagen, as well as much fewer complications with compiling long primary sequences from LC-MS data. Interestingly, the presence of NCPs such as OC in both cattle and pig 'collagen', and serum proteins like A2HS in cattle and sheep 'collagen' and albumin in sheep and pig 'collagen', is in contrast to the chicken 'collagen' sample. In the chicken 'collagen' only a single identification of a serum protein (apolipoprotein) and the more confident identification of the blood protein haemoglobin is indicated. Although the methods used do not show adequate reproducibility, this complete absence of bone NCPs in the avian sample is more likely due to sample pre-treatments rather than absence *in situ*, because of their identical functions in vertebrate tissue.

4.5.3 Thermostable Proteins; Raw MBM vs Sedimented MBM

In terms of modern unheated bone, only a small weight fraction of mature mammalian collagenous tissue can be put into solution in the form of acid soluble tropocollagen; the major portion of the collagen is insoluble in most aqueous systems. This insoluble fraction can be converted into gelatin by heating at 60-65°C (the melting temperature of unmineralised cross-linked collagen) under controlled laboratory conditions that minimises the amount of hydrolysis. This is what also occurs at the much higher temperatures in the MBM rendering processes. Although the amino acid compositions of both raw MBM and sedimented MBM seem to be very similar and both 'collagen-like' (Appendix 3.6; see Elster *et al.* 1991), the principal component analyses of individual amino acid concentrations shown in Figure 4.5 clearly indicates that there are some clear differences. As the LC-MS data of the two types of sample supports the similarity of the proteins that survive within these MBM samples, where the dominant protein in both raw and sedimented MBM is collagen (I), the subtle differences shown in Figure 4.5 could be due to differences in the actual NCPs present and/or differences in the decomposition products remaining. Both the raw and sedimented MBM also indicated the presence of muscle proteins (actin, myosin, myoglobin, troponin, tropomyosin), and blood proteins (albumin and haemoglobin). The chicken

samples, both unsedimented MBM and sedimented MBM, contained much larger amounts of myosin than any other non-collagenous protein, which could explain the off-set in the chicken samples from the mammalian samples plotted on the principal component analysis (Fig. 4.5), where the non-collagenous component of untreated chicken 'collagen' is even more distinct (Table 4.1).

4.5.4 Amino Acid Racemisation Analyses

The expected dominance of collagen in bone tissue is supported by the amino acid ratios of all three fractions ('total', 'acid-soluble', and 'acid-insoluble') in both the unsedimented and sedimented MBM samples (see amino acid concentration tables in Appendix 3.6). Because Asx is thought not to racemise in intact triple helical collagen, the extent of Asx racemisation within a sample of bone can be used to study the integrity of the collagen (Collins *et al.* 1999), where upon heating beyond the melting temperature of collagen, the intra- and intermolecular cross-links and hydrogen bonds that stabilise the molecules of collagen are broken and shrinkage occurs. As the aspartic acid residues are no longer strained they undergo a succinimide reaction, resulting in aspartic acid racemisation (van Duin and Collins 1998). The extent of this racemisation is believed associated to the extent of denaturation of the collagen triple helix and thus the degradation of the sample.

The racemisation results indicate that the applications of aspartic acid racemisation to the measure of heat-induced degradation can be potentially useful to establish food processing history in MBM, as has already been reported by other authors for processed foods (Liardon and Hurrel 1983). The results presented here complementarily show the problems with using the sheep and pig samples of this particular set of standards. However, the D/L values being sufficiently high to imply that, at the higher temperatures at least, the ability to extract authentic DNA would seem unlikely as they are greater than 0.12-0.15 (Poinar *et al.* 1996; Poinar and Stankiewicz 1999) unless this proposal for screening bone samples at such D/L values is incorrect. In the AAR data shown in Figure 4.6, the cattle MBM proteins appeared to be much more degraded than the chicken MBM proteins. This is unexpected as chicken collagen (I) is much more soluble than cattle collagen (I) due to a lower concentration in cross-links (Gerstenfeld *et al.* 1994). However, the degradation could be influenced by the

relative ages of the livestock being slaughtered, i.e., young poultry vs. mature cattle (or the precise conditions of heating). Miller *et al.* (1983) reported increased percentages of soluble (muscle) collagen with the feeding of high-energy diets to mature cattle. This is because during the resulting periods of rapid growth, the rate of protein synthesis is elevated, which results in an increased proportion of newly synthesised collagen resulting in less stable collagen fibers with higher solubility (Boleman *et al.* 1996).

4.6 Concluding Remarks

The dominance of collagen (I) and the similarities of both the unsedimented and the sedimented samples of MBM material are indicative of bone being the dominant tissue present (the presence of skin in not supported by any other identifications, ie., no elastin was identified). However, the presence of highly-racemised collagen in the 'acid-insoluble' fractions indicate extensive denaturation and renaturation. The need for an investigative study of any set of standards that are to be used for method development is made clear by the flaws observed within the 'B-EFPRA' set of MBM standards, in particular the likely contamination of lower temperature (133°C and 137°C) rendered MBM with higher temperature material. The under-estimated levels of protein degradation observed in two of the sixteen samples could have influenced the successes or failures of new methodologies created within the STRATFEED project using this sample set, particularly because most experiments were carried out at the lowest rendering temperature (133°C).

Collagen (I) was shown to be the dominant protein in the highest temperature rendered MBM sample, and hence the most thermostable protein. This supports the concept that it should survive in ancient bone for longer periods of time than either aDNA or OC. Although the potential information gained is vast, the obvious limitations of the LC-MS technique are the extent of protein sequences in the databases, for example there are only five complete protein sequences for both chains of collagen (I). As well as problems inherent in analysing species without a known sequence where only two of the four MBM sample species have published protein sequences, there are also problems with reproducibility of the technique in application to bone samples. Clear differences were also shown to exist between mammalian and avian bone specimens, which could affect how avian peptides mixed into mammalian feeds are analysed and observed. The poor reproducibility from

different samples of the same untreated bone, let alone different individuals of the same species (of different biological ages), indicates that with the additional complications of the MBM rendering processes modifying peptides, either improved genomic sequences of all possible species needs to be achieved, or a robust method for isolating a select few peptides needs to be developed for applications of species identification.

5 A method of isolating the collagen α2 (I) chain carboxytelopeptide for species identification in bone fragments

Reprinted from *Analytical Biochemistry, Volume 374, Issue 2, 15 March 2008, Pages 325-334,*

I proposed the development of the method outlined in this chapter, designed the analyses, carried out the analyses and wrote up the research in this chapter.

Available online at www.sciencedirect.com
ScienceDirect
Analytical Biochemistry 374 (2008) 325–334
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

A method of isolating the collagen (I) α2 chain carboxytelopeptide for species identification in bone fragments

Michael Buckley [a,*], Matthew Collins [b], Jane Thomas-Oates [c]

[a] *Department of Biology, University of York, York YO10 5YW, UK*
[b] *Department of Archaeology, University of York, York YO1 7EP, UK*
[c] *Department of Chemistry, University of York, York YO10 5DD, UK*

Received 20 September 2007
Available online 27 December 2007

Abstract

We present a novel method for the isolation and analysis of the bone collagen (I) α2 chain carboxytelopeptide as a species biomarker. Conventional methods for the analysis and sequencing of mixtures of proteins and peptides commonly involve using the protease trypsin to cleave proteins present in the sample. However, in the study of collagen, these methods result in very complex mixtures of peptides that are difficult to analyze and the acquired results are not reproducible. Here we use bacterial collagenase (from *Clostridium histolyticum*) for its ability to cleave the highly unusual Gly-Xaa-Yaa repeating sequence of collagen, where Xaa usually is Pro and Yaa often is Hyp. Followed by a simple isolation step using a reverse phase solid phase extraction cartridge, the α2 (I) chain carboxytelopeptide can be readily analyzed by matrix-assisted laser desorption/ionization–mass spectrometry (MALDI–MS) and the results can be used to distinguish between different species of origin.

Keywords: Bacterial collagenase; Collagen; Telopeptide; Species variability; Meat and bone meal; Archaeology; Gelatin; Mass spectrometry

Using biomolecules for species identification and phylogenetic investigations has been widely investigated in a number of tissues since the introduction of molecular methods of DNA manipulation to archaeology [1–6] and food and agricultural science [7–12]. Archaeologists use species information for interpreting site-husbanding practices [13–15], as paleontologists do for environmental conditions [16,17]. The meat and livestock industry has been concerned with identifying contaminant proteins in animal feed since the outbreak of bovine spongiform encephalopathy (BSE),[1] which was believed to have been caused by the feeding of ruminant tissues to cattle [18]. The subsequent ban on using processed animal proteins, including meat and bone meal (MBM), as feed ingredients for farmed animals was a pre-preventative measure against the spread of transmissible spongiform encephalopathies (TSEs). To police the subsequent ban on the feeding of ruminant tissues to ruminants, a robust method of species identification must be developed and has been widely investigated [19].

DNA-based methods of species identification are transforming both archaeology and food safety. Archaeological tissues (principally bone) suffer from difficulties regarding contamination with exogenous DNA, not only from excavators [20] but also from pre-excavation contamination such as that deriving from husbanding practices above archaeological sites (e.g., from permeating sheep urine) [21] and from contamination of laboratory reagents [22]. Agricultural tissues such as MBM are purposely sterilized at high temperatures, which degrade the DNA [23]. As a consequence, alternatives to DNA analysis for species identification are being investigated, including methods

* Corresponding author.
E-mail address: mbuckley82@gmail.com (M. Buckley).

[1] *Abbreviations used:* BSE, bovine spongiform encephalopathy; MBM, meat and bone meal; TSE, transmissible spongiform encephalopathy; LC–MS, liquid chromatography–mass spectrometry; SPE, solid phase extraction; MALDI, matrix-assisted laser desorption/ionization; OG, octyl-β-D-glucopyranoside; HCl, hydrochloric acid; ACN, acetonitrile; MeOH, methanol; TFA, trifluoroacetic acid; ddH_2O, distilled and deionized water; CID, collision-induced dissociation; MS/MS, tandem mass spectrometry.

0003-2697/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2007.12.002

based on near infrared microscopy [24,25]. One promising alternative for species identification of archaeological [26–31] and processed [32–34] bone are the bone proteins themselves, the most abundant of which is collagen (I).

The collagen (I) molecule is a heterotrimer composed of two α1 (I) and one α2 (I) chains and contains an uninterrupted triple helical domain (~1000 residues from each chain), approximately 300 nm in length and 1.5 nm in diameter, flanked by short nonhelical telopeptides. The helical region contains a repeating Gly-Xaa-Yaa motif (where Xaa is commonly Pro and Yaa is commonly Hyp) that is highly conserved among species [35]. We have found that conventional liquid chromatography–mass spectrometry (LC–MS) methods for analyzing bone collagen are expensive and time-consuming, and the data analysis is complicated by the within-sample heterogeneity of extractable collagen molecules [36]. To use collagen sequence to infer species, a reproducible method was developed by using the matrix metalloprotease bacterial collagenase to release telopeptides. Bacterial collagenase from *Clostridium histolyticum* is thought to degrade the native protein at the Yaa-Gly bond in the repeating sequence Pro-Yaa-Gly-Pro, where Yaa usually is a neutral amino acid [37,38]. Following digestion, the product is mainly species-uninformative tripeptides [39] and the telopeptides.

The telopeptides, which do not have a repeating amino acid sequence and do not adopt a triple helical conformation, account for 2% of the molecule and are essential for fibril formation [40]. The telopeptides are the most immunogenic regions of the type I molecule, and one specific property of the carboxyterminal telopeptide is that its α1 chain adopts a folded conformation with a sharp hairpin turn around residues 13 and 14 of the 25-residue telopeptide [41]. Most of the telopeptide regions take part in cross-link formation except for the carboxyterminal telopeptide of the α2 chain, which possesses no lysine residues for cross-link formation to occur.

We have developed a solid phase extraction (SPE) method to remove the cleaved tripeptides of digested collagen and to isolate the few surviving peptides, which were predominantly from the α2 (I) carboxytelopeptide. The method was developed using samples of cattle, human, dog, rabbit, and chicken bone collagen to establish reproducibility. The study was then extended to cover bone samples of 17 different species (5 with published telopeptide sequences and 12 of unknown sequences). These sequences were then compared with an additional 19 published and genome-derived sequences in a phylogenetic analysis to establish the usefulness of this approach for species identification and to rationalize the results.

Materials and methods

Materials

Type III bacterial collagenase from *C. histolyticum* with activity of ≥400 collagen digestion units/mg (C0255), matrix-assisted laser desorption/ionization (MALDI) calibration peptides, α-cyano-4-hydroxycinnamic acid, and octyl-β-D-glucopyranoside (OG) were purchased from Sigma–Aldrich. Hydrochloric acid (HCl) and chloroform were purchased from BDH, acetonitrile (ACN) and methanol (MeOH) were purchased from Fisher Scientific, and trifluoroacetic acid (TFA) was purchased from Fluka. C18 SPE cartridges (100 mg) were purchased from Amersham Biosciences (RPN1900). All water was distilled and deionized (ddH_2O).

Bone samples of *Meles meles* (badger), *Oryctolagus cuniculus* (rabbit), *Lepus europaeus* (hare), and *Phasianus colchicus* (pheasant) were provided by A. Chamberlain (University of Sheffield, UK); *Canis familiaris* (dog) and *Vulpes vulpes* (fox) were provided by T. O'Connor (University of York, UK); *Capra hircus* (goat), *Bubalus bubalis* (water buffalo), and *Cervus elaphus* (red deer) bone were obtained from Langthorne Farm (Northallerton, UK); and *Bos taurus* (cattle), *Ovis aries* (sheep), *Sus scrofa* (pig), *Gallus gallus* (chicken), *Meleagris gallopavo* (turkey), and *Anas platyrhynos* (duck) were obtained from Prosper De Mulder (Doncaster, UK). Archaeological *Bison priscus* (bison) was donated by P. Ostrom (Michigan State University, USA), and archaeological *Homo sapiens* (human) was provided by H. Koon (University of York, UK).

Collagen preparation and digestion

Bone shards were defatted twice in chloroform/methanol (87:17, v/v) under constant mixing for 1 h. The bone shards were then powdered under liquid nitrogen (Spexmill 6750) and demineralized for 4 h with 0.6 M HCl at 4 °C. Following centrifugation at 13,000 rpm, the acid-insoluble pellet was rinsed five times with 2 mL of ddH_2O until the pH reached neutral. The rinsed pellet was then lyophilized overnight, and 2 mg of this resulting "collagen" was then weighed out and digested with type III bacterial collagenase for 5 h at 37 °C following the manufacturer's protocol.

Telopeptide isolation and MALDI analysis

The SPE cartridges were prepared with one bed volume each of 100% MeOH followed by 5% ACN, 0.1% TFA, and then 90% ACN/0.1% TFA. Following enzymatic digestion, the samples were centrifuged at 13,000 rpm for 10 min and the supernatant was applied to the SPE cartridges. A stepped gradient of increasing ACN concentrations following several wash steps (2 × 5% ACN and 1 × 22% ACN in 0.1% TFA) was applied to the column, and the eluting fraction (26% ACN in 0.1% TFA) was collected and dried down in a centrifugal evaporator. The samples were resuspended with detergent (1% [w/v] OG in 50 mM Tris–HCl [pH 7.4]) and diluted in 5% ACN/0.1% TFA for mass spectrometric analysis. Then 0.5 μL of sample solution from each SPE fraction was spotted with 0.5 μL of α-cyano-4-hydroxycinnamic acid matrix

solution (1% in ACN/H_2O, 1:1, v/v) and allowed to dry. Each spot was analyzed by MALDI–MS using a calibrated Applied Biosystems 4700 Proteomics Analyzer. Peptide masses of a mix of standard peptides were used for a wide range calibration from 0.8 to 4 kDa. (The standards used were des-Arg^1–bradykinin, angiotensin I, Glu^1–fibrinopeptide, ACTH 1–17 clip, ACTH 18–39 clip, and ACTH 7–39 clip.) *De novo* sequencing of the samples was carried out by manually interpreting the spectra obtained on collision-induced dissociation (CID) product ion MS (tandem mass spectrometry [MS/MS]) of selected peptides.

Results

Isolation and identification of species-diagnostic biomarkers

Following digestion of acid-insoluble cattle, human, dog, rabbit, and chicken collagen with bacterial collagenase and separation using a C18 SPE cartridge, the peptides isolated in the 26% ACN fractions were analyzed by MALDI–MS. The spectra were reproducible and contain peaks that make it possible to distinguish among these five species (Fig. 1). The five most intense peptide ions from the telopeptide-containing fraction (26% ACN) of human, cow, dog, rabbit, and chicken gelatinized collagen digests were selected for CID MS/MS (e.g., see Figs. 2 and 3) and the peptide sequences derived by manual interpretation. Interestingly, all precursor ions analyzed for these five species of known sequence match fragments of the carboxyterminal 18-amino-acid-residue peptide of the collagen (I) α2 chain (hereafter referred to as the α2 (I) chain carboxytelopeptide). We propose the use of the *m*/*z* values of the five most common fragments from this 18-residue α2 (I) carboxytelopeptide as a peptide mass fingerprint diagnostic of particular species. The five peaks, labeled A to E, include the intact telopeptide (E) but also a fragment with only the

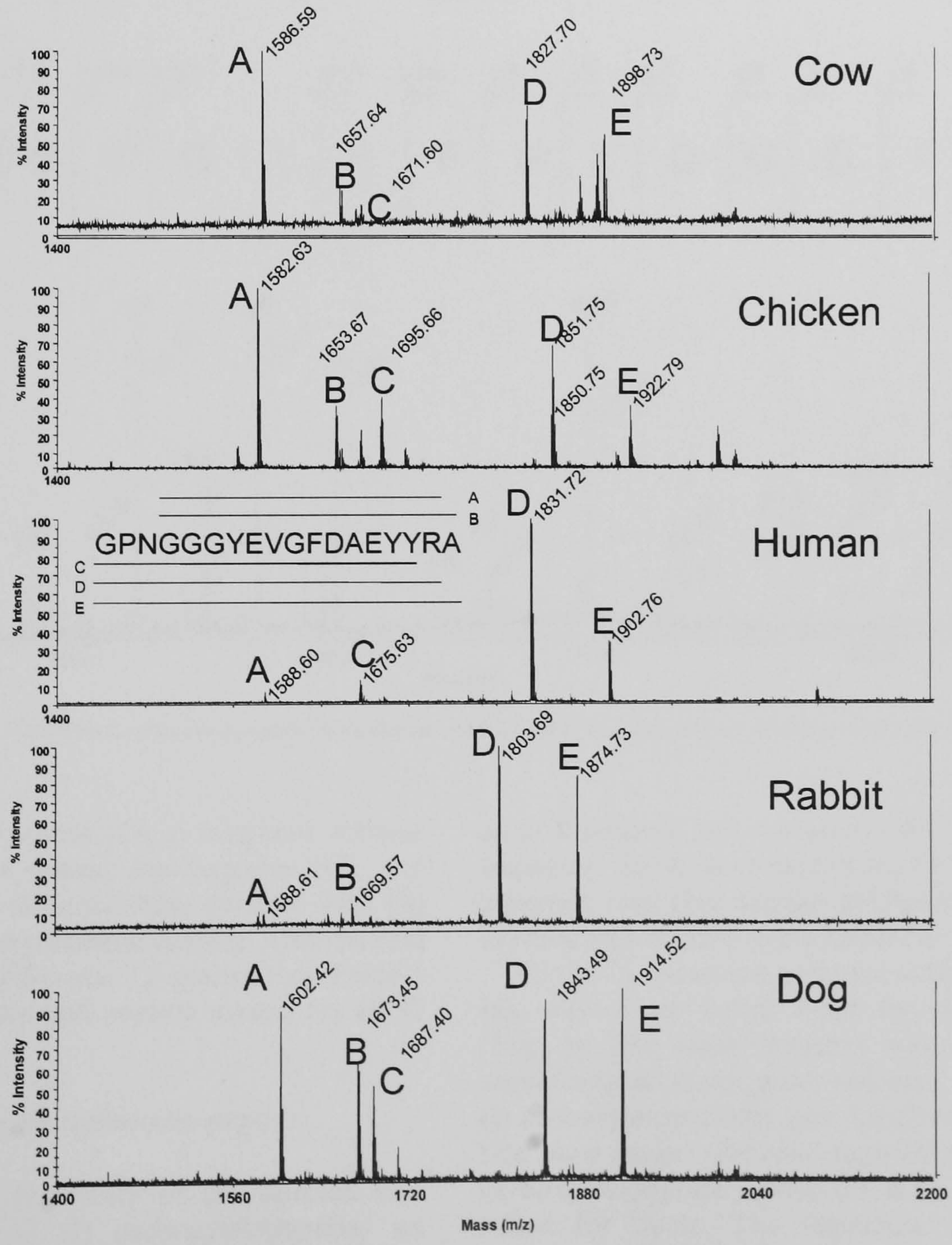
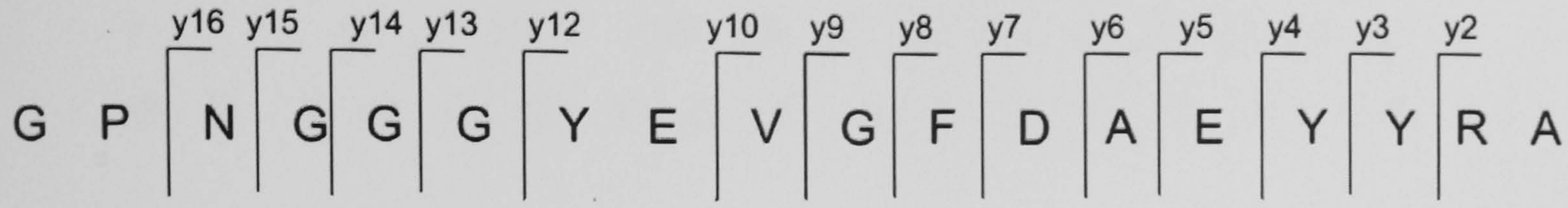
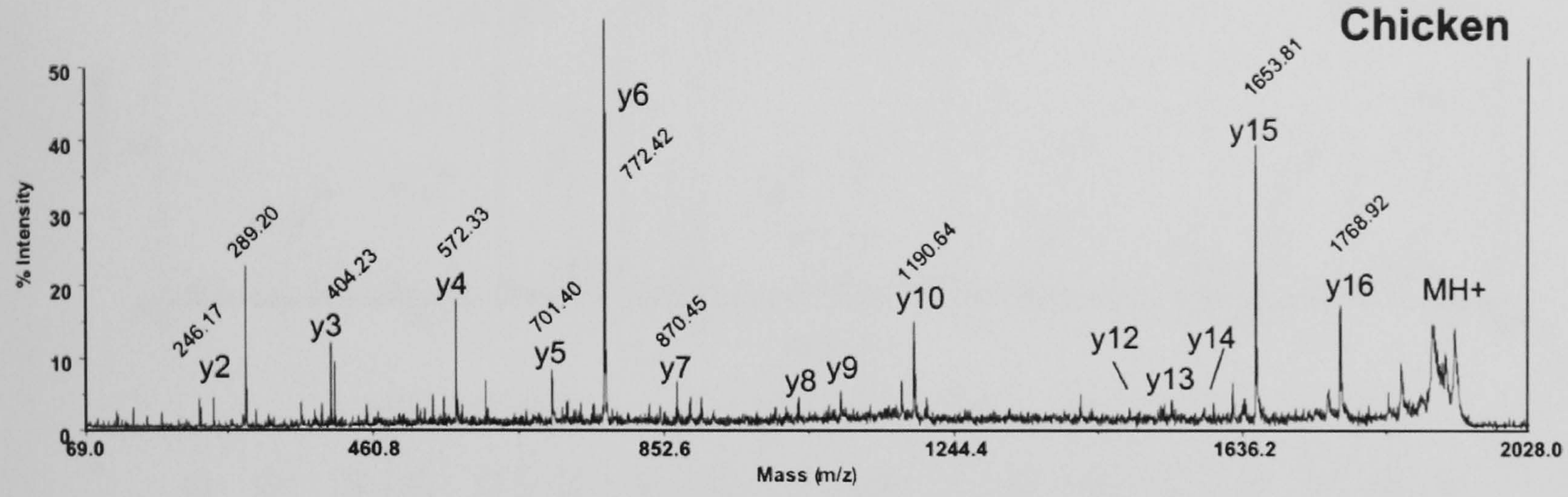
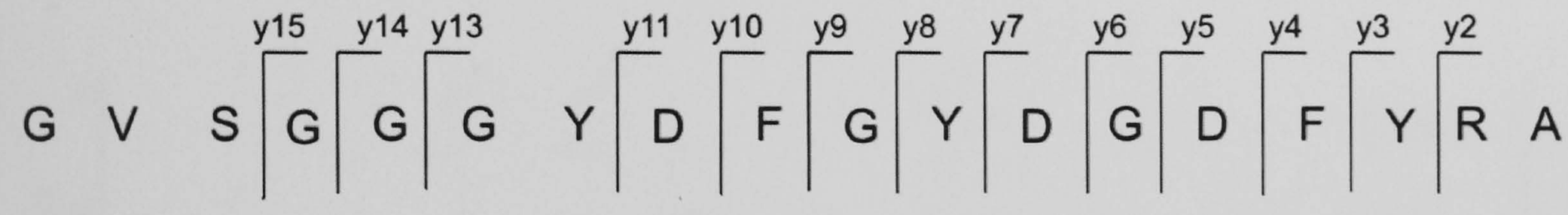


Fig. 1. MALDI mass spectrum of collagen (I) α2 chain carboxytelopeptide isolated from cow, chicken, human, rabbit, and dog bone. Each fraction contains up to five fragments of the 18-residue telopeptide: residues 4 to 17 (A), 4 to 18 (B), 1 to 16 (C), 1 to 17 (D), and 1 to 18 (E), where residues 4 to 6 always are GGG. The mass spectra from the remaining 15 species are shown in the supplementary material.







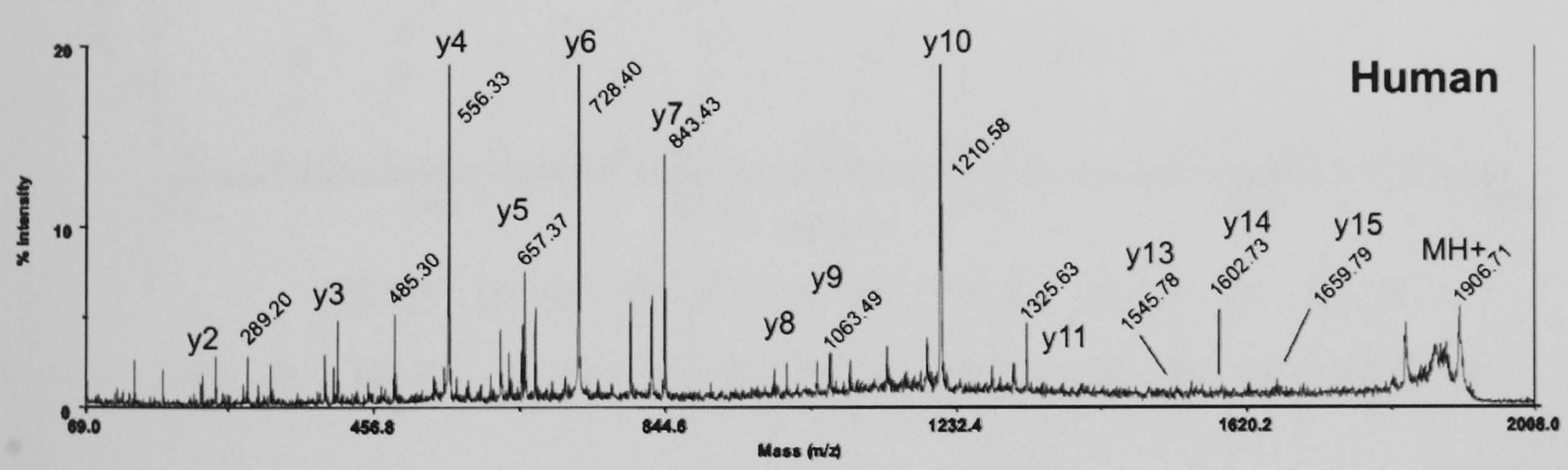


Fig. 2. Tandem mass spectra of 18-residue carboxytelopeptide from chicken and human bone collagen (I) α2 chain with amino acid sequences determined by *de novo* sequencing.

carboxyterminal alanine missing (D), a fragment without both the carboxyterminal alanine and arginine (C), and fragments without the N-terminal three residues with (B) and without (A) the carboxyterminal alanine. This method was then applied to an additional 12 species (see Table 1 for a list of the A–E protonated peptide masses for all 17 species).

Species variability of α2 chain carboxytelopeptide

To assess the species variability of the selected biomarker, the 18-residue α2 (I) carboxytelopeptide, we obtained 9 sequences from the UniProt database (5 of which we confirmed by manual interpretation of our CID MS/MS data), 15 sequences from whole-genome alignment as in Karolchik and coworkers [42], and an additional 12 sequences via *de novo* sequencing (Table 2). The sequences obtained were then aligned in Clustal and a neighbor-joining tree was created using Geneious Basic 3.0.4 (Fig. 4).

Of the 12 sequences obtained solely by *de novo* sequencing, only 3 are novel: those for sheep, hare, and duck (Fig. 3). The same sequence was obtained on *de novo* sequencing of sheep, goat, red deer, and water buffalo α2 (I) carboxytelopeptides (see Supplementary Fig. 5 for peptide mass maps). *De novo* sequencing of the bison α2 (I) carboxytelopeptide yielded the same sequence as that published for cattle. The sequences obtained by *de novo* sequencing of both turkey and pheasant α2 (I) carboxytelopeptides (see Supplementary Fig. 6 for peptide mass maps) matched the sequence of chicken. Likewise, the

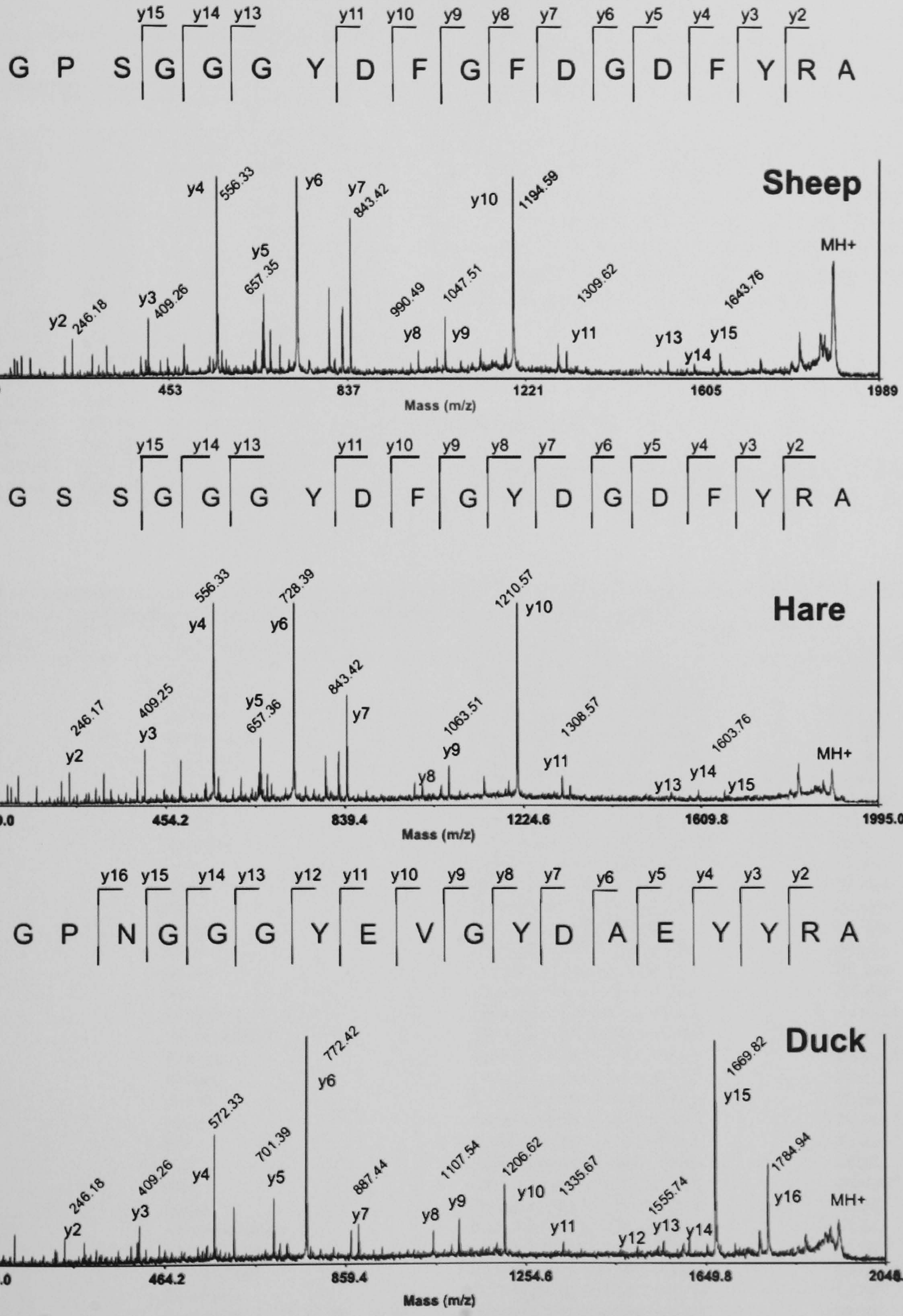


Fig. 3. Tandem mass spectra of 18-residue carboxytelopeptide from sheep, hare, and duck bone collagen (I) α2 chain with amino acid sequence determined by *de novo* sequencing.

sequences obtained by *de novo* sequencing of fox, badger, and pig α2 (I) carboxytelopeptides matched the sequence of dog (see Supplementary Fig. 7 for peptide mass maps).

The analysis of the 18-amino-acid α2 (I) carboxytelopeptide sequences from the 36 species (Fig. 4) indicates some degree of phylogenetic patterning given that the fish,

Table 1
Calculated α2 (I) chain carboxytelopeptide fragment mass/charge values from 17 species analyzed

Common name	Telopeptide fragment *m/z*				
	A	B	C	D	E
Cow	1586.65	1657.68	1671.65	1827.75	1898.79
Chicken	1582.67	1653.71	1694.69	1850.79	1921.83
Human	1588.63	1659.66	1675.65	1831.75	1902.79
Rabbit	1588.63	1659.66	1647.62	1803.72	1874.75
Hare	1588.63	1659.66	1663.61	1819.71	1890.75
Dog	1602.64	1673.68	1687.65	1843.75	1914.79
Sheep	1572.63	1643.67	1657.64	1813.74	1884.78
Goat	1572.63	1643.67	1657.64	1813.74	1884.78
Water buffalo	1572.63	1643.67	1657.64	1813.74	1884.78
Red deer	1572.63	1643.67	1657.64	1813.74	1884.78
Turkey	1582.67	1653.71	1694.69	1850.79	1921.83
Duck	1598.67	1669.71	1710.68	1867.79	1938.82
Pheasant	1582.67	1653.71	1694.69	1850.79	1921.83
Bison	1586.65	1657.68	1671.65	1827.75	1898.79
Pig	1602.64	1673.68	1687.65	1843.75	1914.79
Fox	1602.64	1673.68	1687.65	1843.75	1914.79
Badger	1602.64	1673.68	1687.65	1843.75	1914.79

amphibians, avians, and mammals all form distinct lineages. However, within mammals, although all artiodactyls fall under one clade (with the armadillo forming a sister clade) and both rodents form their own clade, other relationships are not consistent with the known phylogeny of the mammals. This suggests a degree of convergent evolution within a restricted sequence (e.g. the two lagomorphs [hare and rabbit] are separate, with bushbaby having the same sequence as rabbit). This indicates that prior information regarding the potential species (as is the case with most archaeological specimens as well as MBM) is required to confidently identify the unknown sample.

Discussion

Telopeptide isolation

The cleavage site of bacterial collagenase is believed to be the Yaa-Gly of triple helical collagen [37], supported

Table 2
Collagen (I) α2 chain carboxytelopeptide sequences derived from either MALDI–TOF/TOF tandem MS (labeled *de novo*), the UniProt protein database (given an accession number), or whole-genome alignment (labeled UCSC) as in Karolchik and coworkers [42]

Genus and species name	Common name	Telopeptide sequence	Accession number
Danio rerio	Zebrafish	GPSGGGYDTSGGYDEYRA	UCSC
Oncorhyncus mykiss	Rainbow trout	GPAGGGYDQSGGYDEYRA	O93484
Rana catesbiana	Bullfrog	GPSGGGYDGGDGGEYYRA	O42350
Anolis carolinensis	Green anole	GVNGGGYEIGYDMEYYRA	UCSC
Xenopus tropicalis	Pipid frog	GASGGGYDGGFGGEFFRA	UCSC
Gallus gallus	Chicken	GPNGGGYEVGFDAEYYRA	P02467*
Meleagris gallopavo	Turkey	GPNGGGYEVGFDAEYYRA	*De novo**
Anas platyrhyncos	Duck	GPNGGGYEVGYDAEYYRA	*De novo**
Phasianus colchicus	Pheasant	GPNGGGYEVGFDAEYYRA	*De novo**
Bos taurus	Cow	GPSGGGYEFGFDGDFYRA	P02465*
Bison priscus	Bison	GPSGGGYEFGFDGDFYRA	*De novo**
Sus scrofa	Pig	GPSGGGYDFGYEGDFYRA	*De novo**
Ovis aries	Sheep	GPSGGGYDFGFDGDFYRA	*De novo**
Capra hircus	Goat	GPSGGGYDFGFDGDFYRA	*De novo**
Mus musculus	Mouse	GVSGGGYDFGFEGDFYRA	Q01149
Rattus norvegicus	Rat	GVSGGGYDFGFEGGFYRA	P02466
Meles meles	Badger	GPSGGGYDFGYEGDFYRA	*De novo**
Bubalus bubalus	Water buffalo	GPSGGGYDFGFDGDFYRA	*De novo**
Cervus elaphus	Red deer	GPSGGGYDFGFDGDFYRA	*De novo**
Homo sapiens	Human	GVSGGGYDFGYDGDFYRA	P08123*
Oryctolagus cuniculus	Rabbit	GASGGGYDFGYDGDFYRA	Q28668*
Lepus europaeus	Hare	GSSGGGYDFGYDGDFYRA	*De novo**
Vulpes vulpes	Fox	GPSGGGYDFGYEGDFYRA	*De novo**
Felis catus	Cat	GPSGGGYDFGYEGDFYRA	UCSC
Canis familiaris	Dog	GPSGGGYDFGYEGDFYRA	O46392*
Equus caballus	Horse	GPSGGGYDFGYDGDFYRA	UCSC
Loxodonta africana	African elephant	GPSGGGYDFGYDGDFYRA	UCSC
Macaca mulatta	Rhesus macaque	GVSGGGYDFGYDGDFYRA	UCSC
Pan troglodytes	Chimpanzee	GVSGGGYDFGYDGDFYRA	UCSC
Otolemur garnettii	Bushbaby	GASGGGYDFGYDGDFYRA	UCSC
Tupaia belangeri	Tree shrew	GVSGGGYDFGYDGDFYRA	UCSC
Cavia porcellus	Guinea pig	GASGGGYDFGFDGDFYRA	UCSC
Echinops telfairi	Tenrec/Lesser hedgehog	GVSGGGYDFGYDGDFYRA	UCSC
Erinaceus europaeus	Hedgehog	GPSGGGYDFGYEGDFYRA	UCSC
Dasypus novmecinctus	Armadillo	GASGGGYDFGYEGDFYRA	UCSC
Ornithorhynchus anatinus	Platypus	GPSGGGYDFGYEGDFYRA	UCSC

Note. Accession numbers marked with an asterisk (*) indicate sequences that were supported by *de novo* sequencing

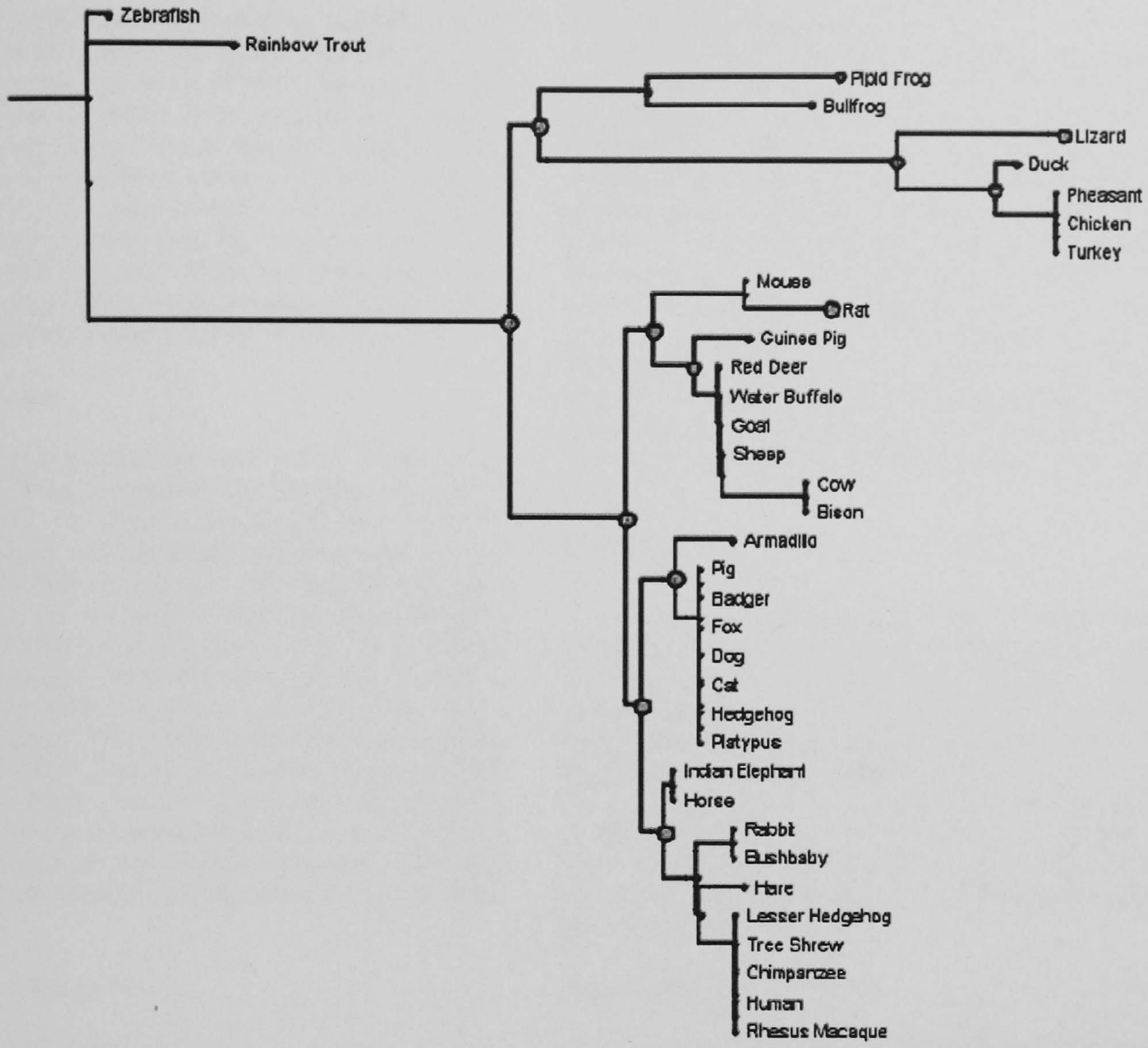


Fig. 4. Phylogram of collagen (I) α2 chain carboxytelopeptide sequences from 36 species showing amounts of amino acid sequence variation. In addition to the telopeptide sequences of the 17 species we investigated, sequences from 2 other species (mouse and rat) were obtained from UniProt and sequences from 11 other species were obtained from the UCSC human (hg18) referenced 28-way whole-genome alignment [42].

by its ability to convert the bulk of the fibrous protein into tripeptides [39]. The majority of these small peptides are not retained or are relatively easily eluted from the C18 SPE cartridge (<22% ACN). The last peptides retained are the fragments of the α2 (I) carboxytelopeptide and can be eluted by a more hydrophobic (26% ACN) wash; no additional peptides were eluted at higher organic concentrations. Although the α2 (I) carboxytelopeptide (and fragments thereof) eluted last of all, it has a lower p*I* (~4) than that of most collagen peptides (p*I* 6–9). It is interesting to note that the peptides isolated in this way belong to the only telopeptide that does not contain a lysine residue and, as a result, does not get posttranslationally modified to hydroxylysine. This modification is used elsewhere in collagen as an integral part of cross-link formation [43].

Despite the number of peptides identified in this fraction, all were products of the α2 (I) carboxytelopeptide. Partial enzymatic cleavage of the GPS–GGG is not surprising considering the Pro-Yaa-Gly cleavage site (between Yaa and Gly), although the apparent abundance of the peptide lacking the carboxyterminal Ala was unexpected. A similar effect was observed in previous studies of osteocalcin, whereby the loss of the nonpolar carboxyterminal residue (often Val) occurred with increasing temperatures [27].

The five protonated telopeptide fragments observed in the MALDI mass spectra have different relative abundances (Fig. 1 and Supplementary Figs. 5–7). However, these differences in relative abundance do not appear to be species specific or species dependent; only the *m*/*z* values for the 18-residue protonated telopeptide are used to identify species. The *m*/*z* values for the smaller fragments can be used in conjunction with the MS/MS data to identify the sequence of the telopeptide. The variability in relative intensities of the ions for these telopeptide fragments could be caused by a number of reasons such as variations in substrate/enzyme ratios (due to difficulties in accurately

weighing the lyophilized enzyme and/or variability in the amount of protein in the lyophilized collagen residue) and potential impurities, the levels of which presumably vary slightly from batch to batch, in the commercially available enzyme; the manufacturers note that the collagenase may contain traces of clostripain, which is an Arg-C endoproteinase. Alternatively, these intensity variations may reflect sample-to-sample variations in the complex matrix from which the samples were fractionated and, thus, simply reflect differences in ion yield due to the presence of slightly different competing/suppressing components in the sample.

Sequence variation

From the known sequences (Table 1), 7 of the 18 residues are conserved throughout all organisms: the amino-terminal glycine, the triglycine motif (GGG) followed by a tyrosine residue, and the three carboxyterminal residues YRA (with the only exception to this being FRA in pipid frog). Among the mammals, there are also conserved phenylalanine residues at positions 9 and 15. The highly conserved triglycine motif has been shown to have a greater binding affinity for other proteins than that of the diglycine motif [44] present in the other telopeptides. This speculatively makes the α2 (I) chain carboxytelopeptide in bone collagen (I) an ideal candidate for the site of osteocalcin binding, particularly because it is the only telopeptide not able to cross-link and because both bone collagen (I) and osteocalcin molecules exist in equimolar amounts [45].

Applications as biomarkers

Paleontology and archaeology

Sequence identification has proven to be very useful in analyzing taxonomic identifications in both modern and ancient species, most commonly using DNA [1–6]. Both animal and plant species often are used as indicators of climate as well as other environmental conditions and have been particularly widely investigated for the British record [46–48]. The applicability of biomolecular methods of species identification is apparent in the many paleontological site species lists for which a significant proportion of samples remain unidentified [49–51]. However, the older the material, the more likely that depurination effects of bone diagenesis will occur throughout the sample. More recently, the longer term survival of bone proteins over DNA is being demonstrated [52,53]. As a result, there is potential for methods using bone collagen to be readily applied to answer paleontological and archaeozoological questions in fragmentary bone specimens.

In addition, the ability to follow the fate of the telopeptide would be useful for understanding the diagenetic history of archaeological bone collagen, for example, to interpret racemization kinetics, where Collins and coworkers [54] argued that the accumulation of D-Asx is concentrated in the telopeptides.

Agricultural and food science

Methods using biomolecules to determine the origin of species in agricultural and food science have also been on the increase [7–12]. This is because regulation (EC) 999/2001 prohibits the feeding of mammalian processed animal proteins to ruminants; however, due to the lack of appropriate detection methods, the ban was extended to all farmed animals. Because other authors have shown the presence of gelatin, a degraded form of collagen, in rendered MBM [33], the method presented here using species-specific peptides of collagen could be applied to differentiate between species. There is also sufficient sequence variation in the 18-residue biomarker to distinguish among the three major mammal species previously used in the production of MBM: cattle, sheep, and pig (Table 2).

Conclusions

A simple method that is both cheap and reproducible of isolating the collagen (I) α2 chain carboxytelopeptide has been described. As a biomarker for species, it is capable of discerning between farmed domesticates (i.e., differentiating cattle from sheep and pig) and, thus, is of potential use for identifying contaminated MBM. It also has potential for applications to paleontology and archaeology due to the well-known survival of collagen in ancient bone. There is additional potential for its application as a biomarker of bone turnover as a supplementary method to those currently in use.

Acknowledgments

The authors thank Andrew Chamberlain, Terry O'Connor, Hannah Koon, and Paul and Kate Langthorne for samples. The authors also thank Brian Raney and Simon Ho for bioinformatics support. The authors gratefully acknowledge the National Environment Research Council for providing a studentship grant (NER/S/J/2004/13017) to M.C. for M.B.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2007.12.002.

References

[1] R.H. Thomas, W. Schaffner, A.C. Wilson, S. Pääbo, DNA phylogeny of the extinct marsupial wolf, Nature 340 (1989) 465–467.
[2] R. Higuchi, B. Bowman, M. Freiberger, O.A. Ryder, A.C. Wilson, DNA sequences from the quagga, an extinct member of the horse family, Nature 312 (1984) 282–284.
[3] R.C. Fleischer, S.L. Olson, H.F. James, A.C. Cooper, Identification of the extinct Hawaiian eagle (*Haliaeetus*) by mtDNA sequence analysis, Auk 117 (2000) 1051–1056.

[4] A.D. Greenwood, C. Capelli, G. Possnert, S. Paabo, Nuclear DNA sequences from late Pleistocene megafauna, Mol. Biol. Evol. 16 (1999) 1466–1473.
[5] C.M. Nielsen-Marsh, P.H. Ostrom, H. Gandhi, B. Shapiro, A. Cooper, P.V. Hauschka, M.J. Collins, Exceptional preservation of bison bones >55 ka as demonstrated by protein and DNA sequences, Geology 30 (2002) 1099–1102.
[6] D.Y. Yang, J.R. Woiderski, J.C. Driver, DNA analysis of archaeological rabbit remains from the American Southwest, J. Archaeol. Sci. 32 (2005) 567–578.
[7] B. Chiappini, G. Brambilla, U. Agrimi, G. Vaccari, H.J.M. Aarts, G. Berben, D. Frezza, V. Giambra, Real-time polymerase chain reaction approach for quantitation of ruminant-specific DNA to indicate a correlation between DNA amount and meat and bone meal heat treatments, J. AOAC Intl. 88 (2005) 1399–1403.
[8] A. Toyoda, M. Nakajo, H. Kawachi, T. Matsui, H. Yano, PCR detection of bovine mitochondrial DNA derived from meat and bone meal in feed, J. Food Protect. 67 (2004) 2829–2832.
[9] Y.H. Cheng, C.M. Wen, S.T. Ding, C.C. Kao, T.Y. Kuo, Detecting meat-and-bone meal in ruminant's feeds by species-specific PCR, J. Anim. Feed Sci. 12 (2003) 849–858.
[10] K. Tajima, O. Enishi, M. Amari, M. Mitsumori, H. Kajikawa, M. Kurihara, S. Yanai, H. Matsui, H. Yasue, T. Mitsuhashi, T. Kawashima, M. Matsumoto, PCR detection of DNAs of animal origin in feed by primers based on sequences of short and long interspersed repetitive elements, Biosci. Biotechnol. Biochem. 66 (2002) 2247–2250.
[11] S. Lahiff, M. Glennon, L. O'Brien, J. Lyng, T. Smith, M. Maher, N. Shilton, Species-specific PCR for the identification of ovine, porcine, and chicken species in meat and bone meal (MBM), Mol. Cell. Probes 15 (2001) 27–35.
[12] F. Bellagamba, V.M. Moretti, S. Comincini, F. Valfre, Identification of species in animal feedstuffs by polymerase chain reaction–restriction fragment length polymorphism analysis of mitochondrial DNA, J. Agric. Food Chem. 49 (2001) 3775–3781.
[13] P. Halstead, Pastoralism or household herding? Problems of scale and specialization in early Greek animal husbandry, World Archaeol. 28 (1996) 20–42.
[14] U. Albarella, The mystery of husbandry: Medieval animals and the problem of integrating historical and archaeological evidence, Antiquity 73 (1999) 867–875.
[15] I. Mainland, P. Halstead, The economics of sheep and goat husbandry in Norse Greenland, Arctic Anthropol. 42 (2005) 103–120.
[16] G.R. Coope, P.L. Gibbard, A.R. Hall, R.C. Preece, J.E. Robinson, A.J. Sutcliffe, Climatic and environmental reconstructions based on fossil assemblages from Middle Devensian (Weichselian) deposits of the River Thames at South Kensington, Central London, UK, Quat. Sci. Rev. 16 (1997) 1163–1195.
[17] T. van Kolfschoten, The Eemian mammal fauna of Central Europe, Geologie En Mijnbouw [Netherlands Journal of Geosciences] 79 (2000) 269–281.
[18] M.G. Doherr, Etiology, transmission, and epidemiological situation of BSE in Europe, Praktische Tierarzt 83 (2002) 156–161.
[19] P. Vermeulen, V. Baeten, P. Dardenne, The Stratfeed Project: A European initiative to tackle the problematic detection of MBM in compound feed, in: P. Dardenne (Ed.), Stratfeed: Strategies and methods to detect and quantify mammalian tissues in feedingstuffs, European Communities, Luxembourg, 2005.
[20] T.A. Brown, K. Brown, Ancient DNA and the archaeologist, Antiquity 66 (1992) 10–23.
[21] J. Haile, R. Holdaway, K. Oliver, M. Bunce, M.T.P. Gilbert, R. Nielsen, K. Munch, S.Y.W. Ho, B. Shapiro, E. Willerslev, Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor? Mol. Biol. Evol. 24 (2007) 982–989.
[22] J.A. Leonard, O. Shanks, M. Hofreiter, E. Kreuz, L. Hodges, W. Ream, R.K. Wayne, R.C. Fleischer, Animal DNA in PCR reagents plagues ancient DNA research, J. Archaeol. Sci. 34 (2007) 1361–1366.
[23] T. Lindahl, B. Nyberg, Rate of depurination of native deoxyribonucleic acid, Biochemistry 11 (1972) 3610–3618.
[24] D.C. Perez-Marin, A. Garrido-Varo, J.E. Guerrero-Ginel, A. Gomez-Cabrera, Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: Chemical composition and open-declaration, Anim. Feed Sci. Technol. 116 (2004) 333–349.
[25] V. Baeten, A. M. Reiner, G. Sinnaeve, P. Dardenne, Analysis of feedingstuffs by near-infrared microscopy (NIRM): Detection and quantification of meat and bone meal (MBM), in: Proceedings of the 6th International Symposium on Food Authenticity and Safety, FASIS, Nantes, France, 2001.
[26] C. Cattaneo, K. Gelsthorpe, P. Phillips, R.J. Sokol, Detection of blood proteins in ancient human bone using ELISA: A comparative study of the survival of IgG and albumin, Intl. J. Osteoarchaeol. 2 (1992) 3103–3107.
[27] P.H. Ostrom, M. Schall, H. Gandhi, T.L. Shen, P.V. Hauschka, J.R. Strahler, D.A. Gage, New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry, Geochim. Cosmochim. Acta 64 (2000) 1043–1050.
[28] M.J. Collins, A.M. Child, A.T.C. van Duin, C. Vermeer, Ancient osteocalcin: The most stable bone protein? Ancient Biomol. 2 (1998) 223–238.
[29] N. Tuross, L. Stathoplos, Ancient proteins in fossil bones, in: E.A. Zimmer, T.J. White, R.L. Cann, A.C. Wilson (Eds.), Molecular Evolution: Producing the Biochemical Data, vol. 224, Academic Press, San Diego, 1993, pp. 121–129.
[30] H.N. Poinar, B.A. Stankiewicz, Protein preservation and DNA retrieval from ancient tissues, Proc. Natl. Acad. Sci USA 96 (1999) 8426–8431.
[31] P.H. Ostrom, H. Gandhi, J.R. Strahler, A.K. Walker, P.C. Andrews, J. Leykam, T.W. Stafford, R.L. Kelly, D.N. Walker, M. Buckley, J. Humpula, Unraveling the sequence and structure of the protein osteocalcin from a 42 ka fossil horse, Geochim. Cosmochim. Acta 70 (2006) 2034–2044.
[32] S.H. Kim, T.S. Huang, T.A. Seymour, C.I. Wei, S.C. Kempf, C.R. Bridgman, D. Momcilovic, R.A. Clemens, H.J. An, Development of immunoassay for detection of meat and bone meal in animal feed, J. Food Protect. 68 (2005) 1860–1865.
[33] M.F. Ocana, H. Neubert, A. Przyborowska, R. Parker, P. Bramley, J. Halket, R. Patel, BSE control: Detection of gelatine-derived peptides in animal feed by mass spectrometry, Analyst 129 (2004) 111–115.
[34] M. Fernandez Ocaña, J. Jarvis, R. Parker, P.M. Bramley, J.M. Halket, R.K.P. Patel, H. Neubert, C-terminal sequencing by mass spectrometry: Application to gelatine-derived proline-rich peptides, Proteomics 5 (2005) 1209–1216.
[35] M.L. Chu, W. de Wet, M. Bernard, J.F. Ding, M. Morabito, J. Myers, C. Williams, F. Ramirez, Human pro α1 (I) collagen gene structure reveals evolutionary conservation of a pattern of introns and exons, Nature 310 (1984) 337–340.
[36] D.S. Jackson, J.P. Bentley, On the significance of the extractable collagens, J. Cell Biol. 7 (1960) 37–42.
[37] M. Soberano, G. Schoellmann, Specificity of bacterial collagenase: Studies with peptides newly synthesized using the solid-phase method, Biochim. Biophys. Acta 271 (1972) 133–144.
[38] E. Harris, M. Farrell, Resistance to collagenase: A characteristic of collagen fibrils cross-linked by formaldehyde, Biochim. Biophys. Acta 278 (1972) 133–141.
[39] M.J. DeNiro, S. Weiner, Use of collagenase to purify collagen from prehistoric bones for stable isotopic analysis, Geochim. Cosmochim. Acta 52 (1988) 2425–2431.
[40] K.E. Kadler, D.F. Holmes, J.A. Trotter, J. Chapman, Collagen fibril formation, Biochem. J. 316 (1996) 1–11.
[41] J.P. Orgel, T.J. Wess, A. Miller, The in situ conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen, Struct. Folding Des. 8 (2000) 137–142.
[42] D. Karolchik, R. Baertsch, M. Diekhans, T.S. Furey, A. Hinrichs, Y.T. Lu, K.M. Roskin, M. Schwartz, C.W. Sugnet, D.J. Thomas, The

UCSC Genome Browser Database, Nucleic Acids Res. 31 (2003) 51–54.
[43] L. Knott, A.J. Bailey, Collagen cross-links in mineralizing tissues: A review of their chemistry, function, and clinical relevance, Bone 22 (1998) 181–187.
[44] J.K. Klarlund, W. Tsiaras, J.J. Holik, A. Chawla, M.P. Czech, Distinct polyphosphoinositide binding selectivities for pleckstrin homology domains of GRP1-like proteins based on diglycine versus triglycine motifs, J. Biol. Chem. 275 (2000) 32816–32821.
[45] R.V. Prigodich, M.R. Vesely, Characterization of the complex between bovine osteocalcin and type I collagen, Arch. Biochem. Biophys. 345 (1997) 339–341.
[46] A. Currant, R. Jacobi, A formal mammalian biostratigraphy for the Late Pleistocene of Britain, Quat. Sci. Rev. 20 (2001) 1707–1716.
[47] D.C. Schreve, Differentiation of the British late Middle Pleistocene interglacials: The evidence from mammalian biostratigraphy, Quat. Sci. Rev. 20 (2001) 1693–1705.
[48] A.J. Stuart, A.M. Lister, The mammalian faunas of Pakefield/Kessingland and Corton, Suffolk, UK: Evidence for a new temperate episode in the British early Middle Pleistocene, Quat. Sci. Rev. 20 (2001) 1677–1692.
[49] A.J. Stuart, A comparison of the Middle Pleistocene mammal faunas of Voigstedt (Thuringia German Democratic Republic) and West Runton, Norfolk, England, Quartarpalaontologie Berlin 4 (1981) 155–163.
[50] A.J. Stuart, The vertebrate fauna of the type Cromerian, Boreas 4 (1975) 63–76.
[51] A.J. Stuart, Pleistocene Vertebrates in the British Isles, Longman, London, 1982.
[52] M. Schweitzer, C.L. Hill, J.M. Asara, W.S. Lane, S.H. Pincus, Identification of immunoreactive material in mammoth fossils, J. Mol. Evol. 55 (2002) 696–705.
[53] J.M. Asara, M.H. Schweitzer, L.M. Freimark, M. Phillips, L.C. Cantley, Protein sequences from Mastodon and Tyrannosaurus Rex revealed by mass spectrometry, Science 316 (2007) 280–285.
[54] M.J. Collins, E.R. Waite, A.C.T. van Duin, Predicting protein decomposition: The case of aspartic-acid racemization kinetics, Phil. Trans. Roy. Soc. Lond. Ser. B 354 (1999) 51–64.

6 The survival of the collagen α2 (I) chain carboxytelopeptide beyond osteocalcin and DNA in archaeological bone

6.1 Abstract

The novel method for the isolation and analysis of the collagen α2 (I) chain carboxytelopeptide as a species marker using bacterial collagenase (from Clostridium histolyticum) for its ability to cleave the highly unusual Gly-Xaa-Yaa repeating sequence of collagen followed by a simple isolation step using a reverse phase solid phase extraction cartridge, was investigated in 19 archaeological samples that failed osteocalcin isolation, 12 of which failed aDNA analyses. The results show 100% success rates even in samples where the amino acid concentrations are tenfold lower than for modern bone samples. Additional sequences of extinct giant tortoise and dodo, obtained from specimens that also failed previous attempts of aDNA extraction, are also presented along with proposed sequences of five other species (common shrew, chum salmon, microbat, thirteen-lined ground squirrel and Japanese medaka). These additional seven sequences are compared to 36 previously reported sequences in a phylogenetic analysis to determine the taxonomic resolution of the telopeptide as a marker for species identification.

6.2 Introduction

One of the most significant revelations of biomolecular archaeology was the use of ancient biomolecules for species identification and phylogenetic investigations which has been widely investigated in a number of tissues following the introduction of molecular methods of DNA manipulation to archaeology (Fleischer *et al.* 2000; Greenwood *et al.* 1999; Higuchi *et al.* 1984; Nielsen-Marsh *et al.* 2002; Thomas *et al.* 1989; Yang *et al.* 2005). Archaeologists utilise species information for interpreting site-husbanding practices (Albarella 1999; Halstead 1996; Mainland and Halstead 2005) as palaeontologists do for environmental conditions (Coope *et al.* 1997; van Kolfschoten 2000). However, archaeological tissues (principally bone) suffer from difficulties regarding several forms of contamination (from permeating urine resulting

from husbanding practices above an archaeological site and/or from laboratory reagents (Brown and Brown 1992; Haile *et al.* 2007; Leonard *et al.* 2007) as well as limited survival rates (Nielsen-Marsh 2002). Increasingly, proteins are becoming the subject of archaeological studies (Evershed and Tuross 1996; Nielsen-Marsh *et al.* 2005; Smith *et al.* 2005; Stutz and Tuross 1997), where the dominant protein in bone, collagen (I), has been demonstrated to survive (as indicated by amino acid analyses) in samples that failed aDNA analyses (see Chapter 2). The collagen (I) molecule is a heterotrimer (in higher vertebrates) composed of two α1 (I) and one α2 (I) chains and contains an uninterrupted triple helical domain (~1000 residues from each chain), flanked by short non-helical telopeptides. Because the helical region contains a repeating Gly-Xaa-Yaa motif (where Xaa is commonly Pro and Yaa commonly Hyp) and was considered highly conserved among species (Chu *et al.* 1984), whereas the telopeptides are not restricted to such limited conformations (and thus particular amino acid substitutions), in order to use a partial collagen sequence to infer species, a simple and reproducible method was developed using bacterial collagenase to digest all but the α2 (I) chain carboxytelopeptides (Chapter 5).

However, following a similar principle to the comparison of OC and aDNA survival that was presented in Chapter 2, the usefulness of this carboxytelopeptide species-specific marker method would only be justified if it can be shown to be successful in a number of archaeological samples that have failed for aDNA analyses. Thus, the 19 archaeological bone samples from four European sites that were part of an ancient DNA study (Anderung *et al.* 2005), 10 of which failed to yield aDNA amplicons, were analysed for collagen α2 (I) carboxytelopeptide survival and the results presented here. In addition to these 19 bovine (*Bos primigenius*) specimens, samples of the extinct dodo (*Raphus cucullatus*) and giant tortoise (*Cylindraspis sp.*) from Mauritius were analysed for their telopeptide sequences having also previously failed aDNA analyses following the method of Shapiro *et al.* (2004; Shapiro *et al.* unpublished data). As well as these two species, a further five more α2 (I) carboxytelopeptide sequences were obtained from genomic data alignment (in collaboration with B. Raney, UCSC)

6.3 Experimental

6.3.1 Collagen Preparation and Digestion

The α2 (I) carboxytelopeptide fragments were isolated from archaeological bone specimens following the methods described in section 5.2. For this analysis, the S/N value of each telopeptide fragment peak was also recorded as described for OC in section 2.2.2, the S/N value for the protonated peptide ion peak provides a crude estimate of the quality of data from recoverable collagen (I) telopeptide. As well as sample origin and approximate age, the telopeptide fragment S/N values were tabulated along with with Asx D/L; Asx D/L is a common measure of the extent of deterioration of bone protein (see Poinar *et al.* 1996).

6.3.2 Amino Acid Composition and Racemisation Analyses

Amino acid racemization analyses were also carried out on the 21 archaeological bone samples. Approximately 1 mg bone powder was treated with 100 μL 7M HCl under N_2 at 110°C for 18 h to demineralise the hydroxyapatite and release any peptide-bound amino acids by hydrolysis. Samples were then dried with a centrifugal evaporator and rehydrated with 500 μL 0.01 mM L-homo-arginine as an internal standard for RP-HPLC analysis. The amino acid compositions were analysed by RP-HPLC using fluorescence detection, following a modification of the method of Kaufman and Manley (1998) as described in section 2.2.3.

6.4 Results

6.4.1 Comparative Survival of Collagen Telopeptide with Osteocalcin and DNA

In order to assess the survival of this carboxytelopeptide marker in comparison to other ancient biomolecules, 19 archaeological samples from four European sites previously investigated for both aDNA and OC survival (Chapter 2) were analysed. These 19 samples were selected because all failed to yield OC although DNA was successfully amplified from nine (section 2.3.1). Two further samples that also failed to amplify aDNA (dodo and giant

tortoise) were also analysed; thermal age estimates of these two samples (>50 kyrs $_{@10°C}$) suggested these are at the limits of amplification success. Table 6.1 shows the S/N for the five most dominant telopeptide fragments (Fig. 6.1) recorded from the spectra of the carboxytelopeptide-containing fractions for each archaeological sample as well as a modern cattle (*Bos taurus*) specimen. The identity of the carboxytelopeptide was also confirmed by MS/MS acquisition and subsequent *de novo* sequencing of the dominant peak.

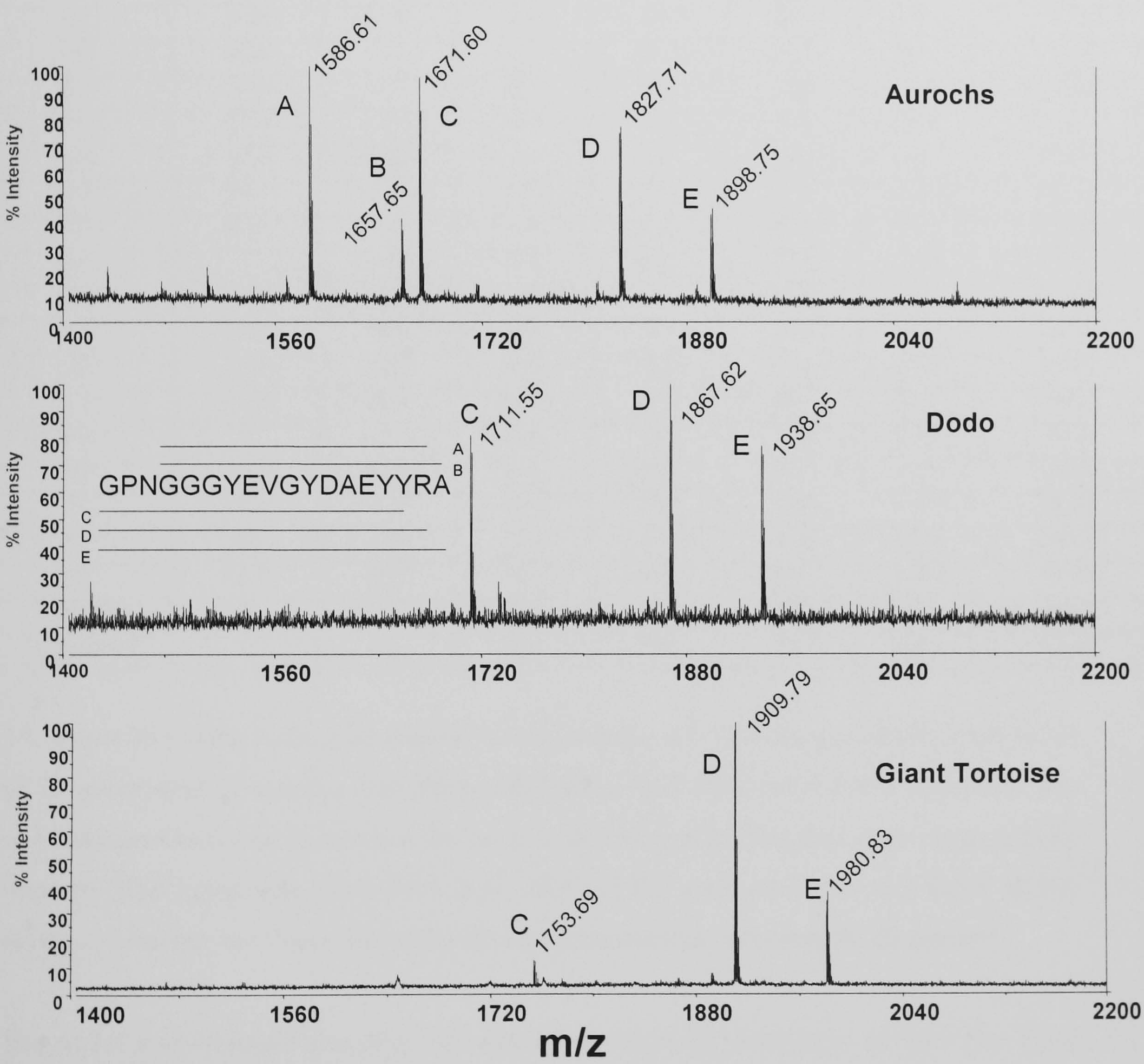


Figure 6.1 - MALDI-MS spectra of collagen a2 (I) chain carboxy-telopeptide isolated from aurochs (sample S13), dodo, and giant tortoise. The fragments are labeled A-E and the fragmentation shown.

Table 6.1 - S/N values of dominant telopeptide peaks from 19 archaeological cattle samples and two Holocene sub-fossils (dodo and giant tortoise) compared with modern cattle standard (indicates failure to amplify ancient DNA). For more sample information on the 19 archaeological cattle samples, see Table 2.2.*

Sample ID	Fragment A S/N	Fragment B S/N	Fragment C S/N	Fragment D S/N	Fragment E S/N	Asx D/L	Total Amino Acid Conc. (nmol/mg)
Modern	76.03	18.27	7.55	47.84	39.59	0.070	141.744
AS1	57.86	25.64	24.75	28.55	13.86	0.147	24.445
AS11	76.26	21.13	28.75	28.80	12.96	0.112	96.081
AS12	62.82	61.80	32.72	26.83	42.51	0.121	70.702
AS3	24.28	1	1	24.63	26.71	0.116	21.168
AS4*	199.32	70.37	52.59	51.72	38.45	0.123	97.882
AS5*	116.15	17.70	30.25	25.69	7.20	0.134	86.493
AS6	84.14	31.89	31.44	43.73	26.09	0.117	72.909
AS8*	195.47	113.62	49.70	41.71	32.26	0.124	75.243
AS9*	82.61	73.08	65.48	34.67	47.06	0.124	93.952
DD11*	1	6.42	39.47	42.17	88.75	0.147	20.528
DD3*	278.17	46.84	40.43	56.43	64.04	0.167	42.235
LE1*	59.61	48.72	30.88	25.22	31.40	0.112	8.051
LE2*	87.06	13.10	30.84	22.67	17.57	0.140	12.189
LE3	107.25	14.34	52.26	36.75	12.29	0.116	55.672
LE5*	1	1	26.08	15.55	33.55	0.132	16.810
M54	109.56	87.91	43.45	44.83	62.80	0.093	155.732
S12	144.24	127.11	64.00	94.69	138.19	0.104	50.550
S13*	71.37	27.88	71.19	60.27	31.44	0.305	8.105
S2	126.29	104.11	37.26	51.64	64.25	0.107	52.443
Dodo*	1	1	16.32	20.63	20.22	0.202	99.895
Giant Tortoise*	1	1	22.88	232.21	78.42	0.181	113.521

The intact 18 residue carboxytelopeptide of the collagen α2 (I) chain was clearly observed in all 21 archaeological samples and often with greater S/N than the modern standards. The only samples where one or more of the peaks with lower m/z (15 residue carboxytelopeptide fragment) had apparently lower S/N (e.g., DD11, LE5) were ones that also failed aDNA analyses, however, the diagnostic carboxytelopeptide markers were still clearly present.

The amino acid compositions (Fig. 6.2) were observed to be identical in samples that failed aDNA analysis to those that succeeded and indicative of the dominance of collagen in the samples (following Elster *et al.* 1991). As shown in Figure 6.3, the concentrations of amino acids in samples that succeeded aDNA analysis were not significantly greater (Mann Whitney U $p=0.086$, $N=22$; Table 6.1) than those that failed aDNA analysis, although the amino acid

racemisation values (Asx D/L) did appear to be significantly different (Mann Whitney U p=0.02, N=22).

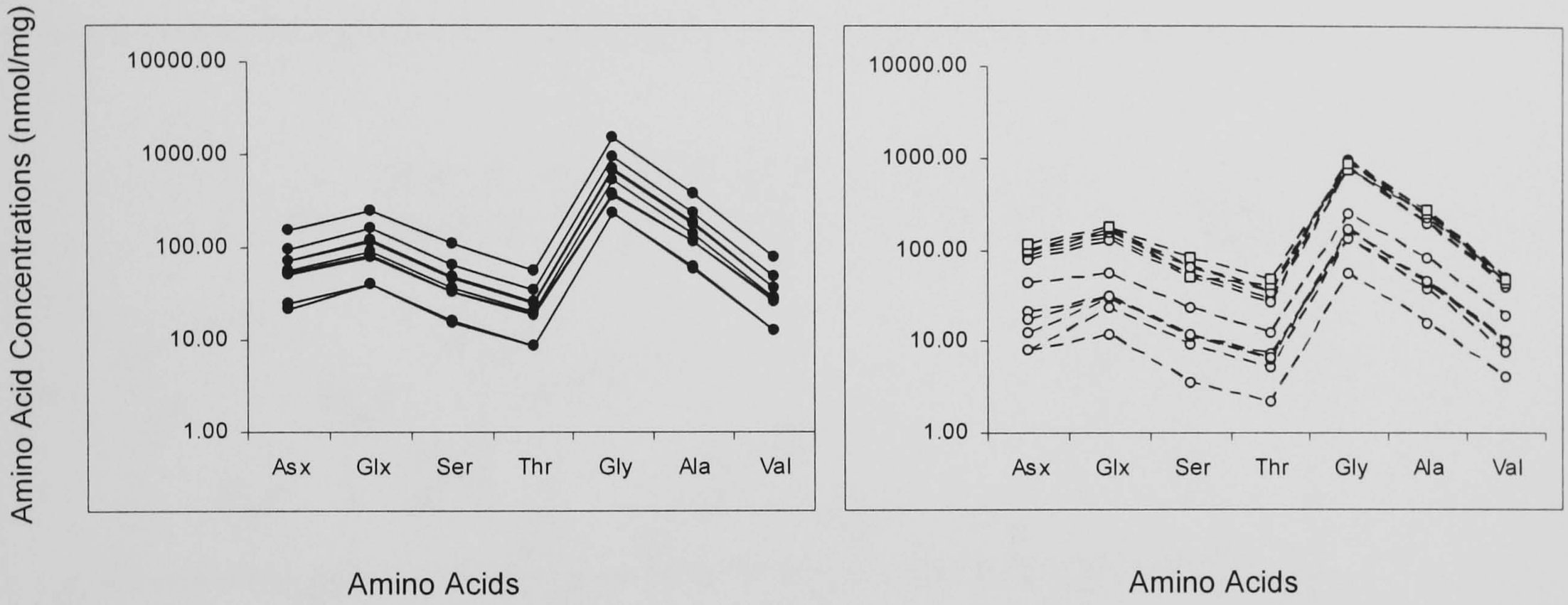


Figure 6.2 – Amino acid composition plots of the 21 archaeological samples (Table 6.1) showing those from which aDNA was amplified (left) and those that failed aDNA amplification (right).

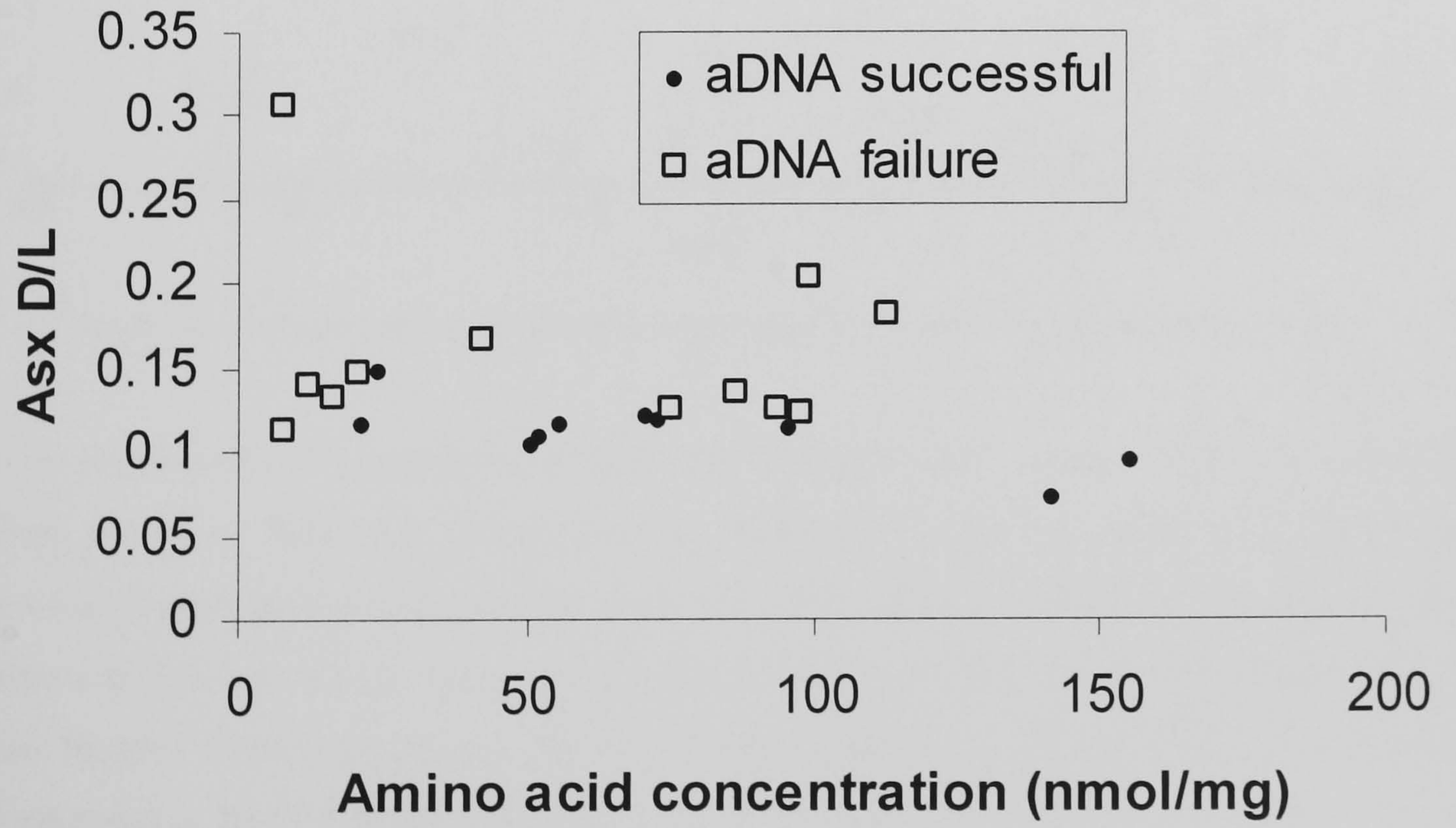


Figure 6.3 – Plot of amino acid concentrations against Asx D/L values showing archaeological samples (Table 6.1) that succeeded aDNA amplification and samples that failed aDNA amplification.

6.4.2 Sequencing Unknown Species and Sequence Variability

Manual interpretation by *de novo* sequencing of product ion tandem mass spectra from species of previously unknown sequence information (Fig. 6.4) was simplified by comparisons to product ion tandem mass spectra of other species, aided by the presence of particular conserved regions such as the triglycine motif at positions 4-6 of the peptide.

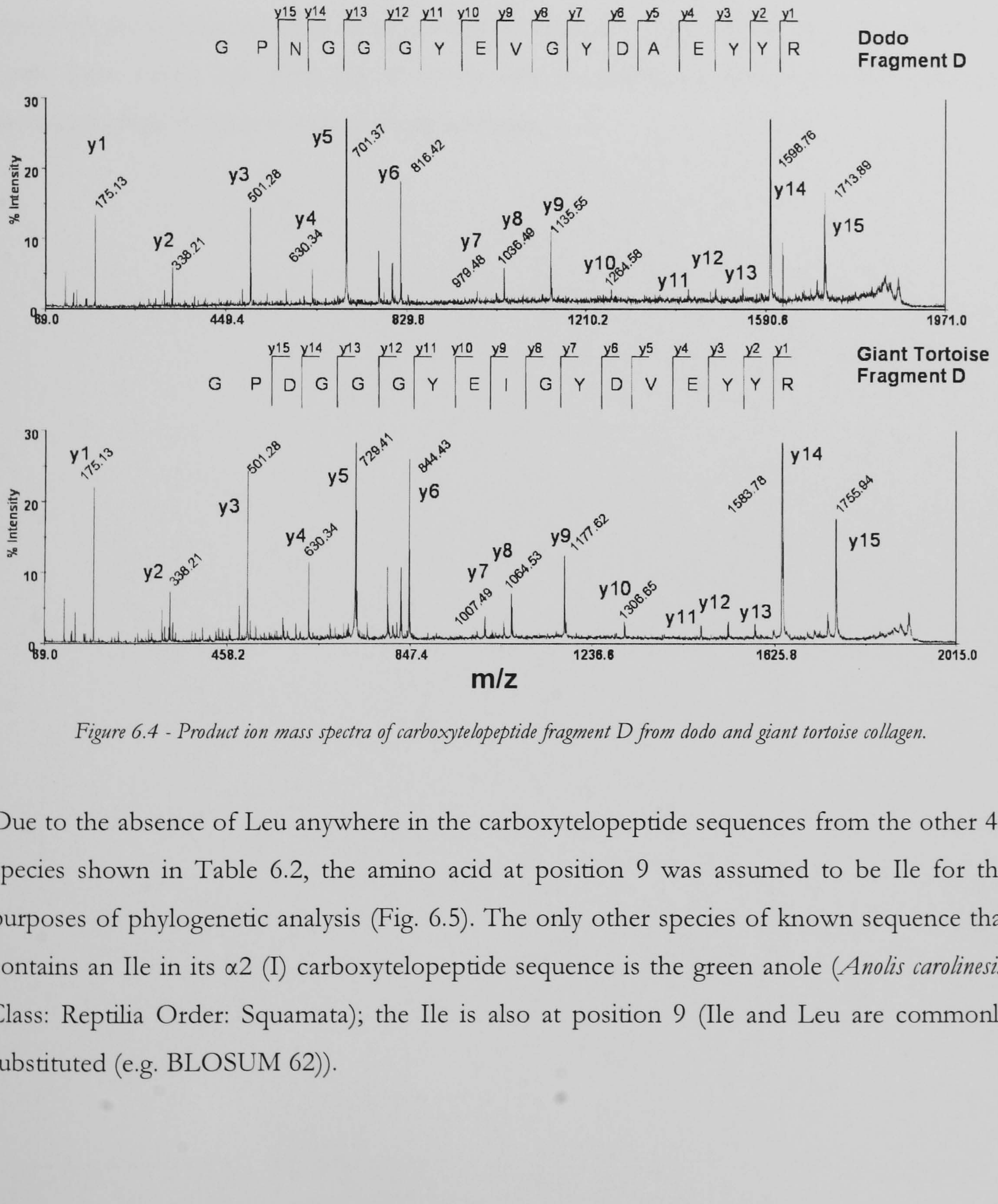


Figure 6.4 - Product ion mass spectra of carboxytelopeptide fragment D from dodo and giant tortoise collagen.

Due to the absence of Leu anywhere in the carboxytelopeptide sequences from the other 42 species shown in Table 6.2, the amino acid at position 9 was assumed to be Ile for the purposes of phylogenetic analysis (Fig. 6.5). The only other species of known sequence that contains an Ile in its α2 (I) carboxytelopeptide sequence is the green anole (*Anolis carolinesis;* Class: Reptilia Order: Squamata); the Ile is also at position 9 (Ile and Leu are commonly substituted (e.g. BLOSUM 62)).

In addition to sequences of dodo and giant tortoise determined by *de novo* sequencing, the sequences of five more species, the common shrew (*Sorex araneus*), chum salmon (*Oncorhynchus keta*), microbat (*Myotis lucifugus*), ground squirrel (*Spermophilus tridecemlineatus*) and the Japanese medaka (*Oryzias latipes*) were also obtained by alignment of available genomic data with the human genome as described in section 5.3.2 (Table 6.2; B. Raney pers. comm). A phylogenetic analysis of the carboxytelopeptide sequences from 43 species (Fig. 6.5) clearly shows the limitations of using this peptide as a marker for species determination. Although particularly useful at distinguishing between some agricultural domesticates such as cattle from sheep and from pig, it is not able to distinguish between other species of archaeozoological interest such as sheep and goat.

Table 6.2 - Table of a2 (I) carboxytelopeptide sequences obtained from either the UniProt protein database, alignment with the human genome by the University of California Santa Cruz (UCSC) or by manual interpretation of analysed specimens (de novo). Conserved residues to the sequence for the zebrafish are represented with a dot. UCSC sequences additional from the previous chapter are marked with an ''.*

Species Name	Common Name	Telopeptide Sequence	Accession No.
Danio rerio	Zebrafish	GPSGGGYDTSGGYDEYRA	UCSC
Oncorhyncus mykiss	Rainbow Trout	..A.....Q....E....	O93484
Rana catesbiana	Bullfrog	.......GGD.GEY...	O42350
Anolis carolinensis	Green Anole	.VN....EIGYDMEY...	UCSC
Xenopus tropicalis	Pipid frog	.A.....GGF.GEFF..	UCSC
Gallus gallus	Chicken	..N....EVGFDAEY...	P02467
Meleagris gallopavo	Turkey	..N....EVGFDAEY...	*Chapter 5*
Anas platyrhyncos	Duck	..N....EVGYDAEY...	*Chapter 5*
Phasianus colchicus	Pheasant	..N....EVGFDAEY...	*Chapter 5*
Bos Taurus	Cow	.......EFGFDG.F...	P02465
Bison priscus	Bison	.......EFGFDG.F...	*Chapter 5*
Sus scrofa	Pig	.......FGYEG.F...	*Chapter 5*
Ovis aries	Sheep	.......FGFDG.F...	*Chapter 5*
Capra hircus	Goat	.......FGFDG.F...	*Chapter 5*
Mus musculus	Mouse	.V.....FGFEG.F...	Q01149
Rattus norvegicus	Rat	.V.....FGFEGGF...	P02466
Meles meles	Badger	.......FGYEG.F...	*Chapter 5*
Bubalus bubalus	Water Buffalo	.......FGFDG.F...	*Chapter 5*
Cervus elaphus	Red Deer	.......FGFDG.F...	*Chapter 5*
Homo sapiens	Human	.V.....FGYDG.F...	P08123
Oryctolagus cuniculus	Rabbit	.A.....FGYDG.F...	Q28668
Lepus europaeus	Hare	.S.....FGYDG.F...	*Chapter 5*
Vulpes vulpes	Fox	.......FGYEG.F...	*Chapter 5*
Felis catus	Cat	.......FGYEG.F...	UCSC
Canis familiaris	Dog	.......FGYEG.F...	O46392
Equus caballus	Horse	.......FGYDG.F...	UCSC
Loxodonta africana	African Elephant	.......FGYDG.F...	UCSC
Macaca mulatta	Rhesus Macaque	.V.....FGYDG.F...	UCSC
Pan troglodytes	Chimpanzee	.V.....FGYDG.F...	UCSC
Otolemur garnettii	Bushbaby	.A.....FGYDG.F...	UCSC
Tupaia belangeri	Tree Shrew	.V.....FGYDG.F...	UCSC
Cavia porcellus	Guinea Pig	.A.....FGFDG.F...	UCSC
Echinops telfairi	Tenrec/Lesser Hedgehog	.V.....FGYDG.F...	UCSC
Erinaceus europaeus	Hedgehog	.......FGYEG.F...	UCSC
Dasypus novmecinctus	Armadillo	.A.....FGYEG.F...	UCSC
Ornithorhynchus anatinus	Platypus	.......FGYEG.F...	UCSC
Sorex araneus	Common Shrew	.......FGYDG.F...	UCSC*
Oncorhynchus keta	Chum salmon	..A.....Q....E....	UCSC*
Myotis lucifugus	Microbat	.I.....FGFDG.F...	UCSC*
Spermophillus tridecemlineatus	Thirteen-lined ground squirrel	.V.....FGYEG.F...	UCSC*
Oryzias latipes	Japanese medaka	..A.....V....E....	UCSC*
Cylindrsapis sp.	Giant Tortoise	..D....EIGYDVEY...	*de novo**
Raphus cucullatus	Dodo	..N....EVGYDAEY...	*de novo**

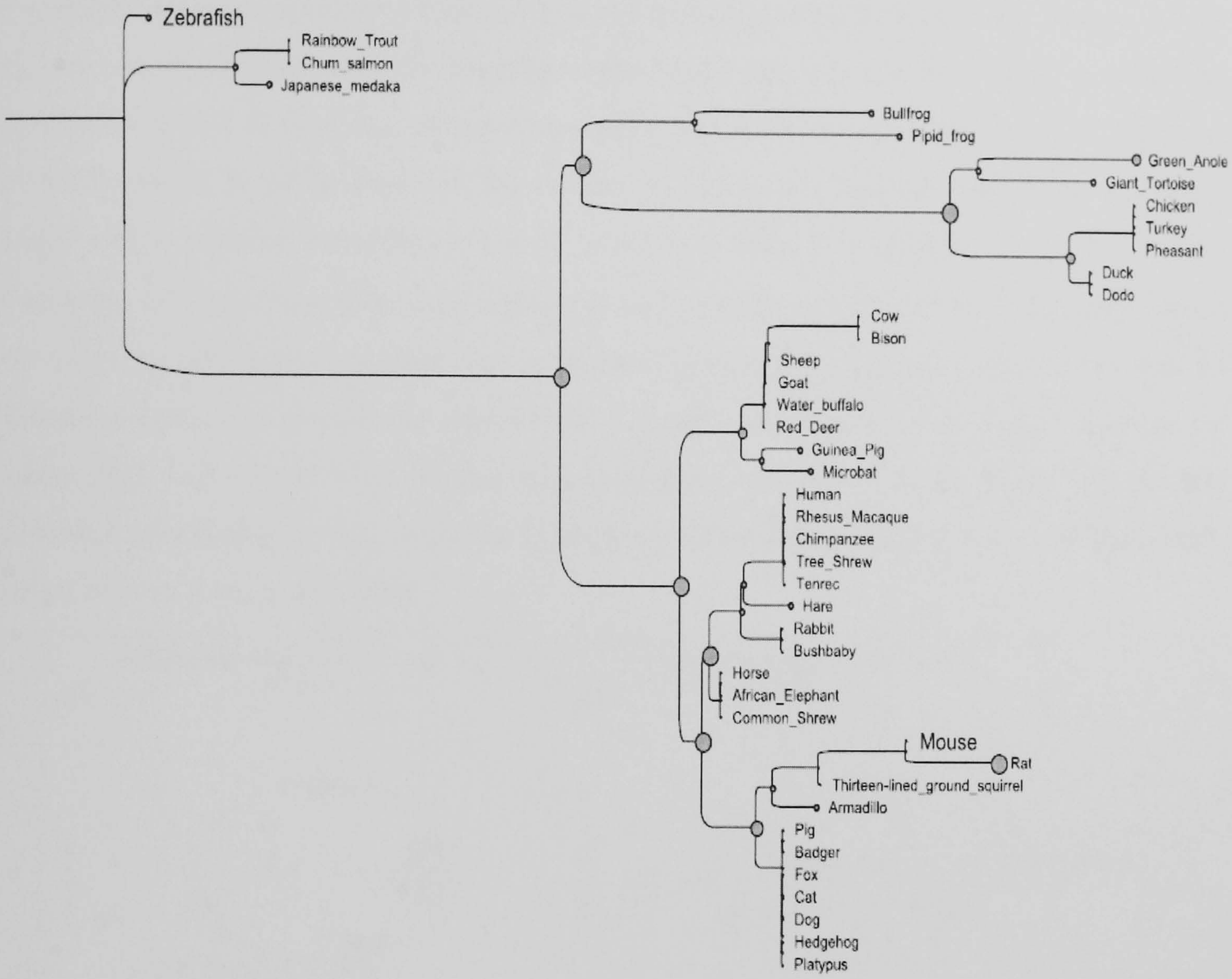


Figure 6.5 - Phylogram of the 43 a2 (I) carboxytelopeptide sequences where the giant tortoise sequence has an assumed I at position 9 despite being indistinguishable from L. Created using Geneious 3.5.6. Branch lengths are indicative of sequence difference, vertical bar indicates 100% sequence identity.

6.5 Discussion

The five protonated telopeptide fragments observed in the MALDI mass spectra have different relative abundances (Fig. 6.1). The m/z values for the smaller fragment peaks can be used in conjunction with the product ion tandem MS data to identify the sequence of the telopeptide. As hypothesised in section 5.4.1, the variability in relative intensities of the ions for the five carboxytelopeptide fragments could be due to a number of reasons, such as variations in substrate:enzyme ratios (due to difficulties in accurately weighing the lyophilized enzyme, and/or potential impurities, the levels of which presumably vary slightly from batch to batch, in the commercially available enzyme; the manufacturers (Sigma-Aldrich) note that the collagenase may contain trace amounts of clostripain, which is an Arg-

C endoproteinase. Sample-specific factors such as variability in the amount of protein in the lyophilised collagen residue or variability in the extent of degradation in the samples could also be a major influence on the fragment ratios observed. It is also interesting to note that fragments A and B were not observed in either the dodo or the giant tortoise sequence, possibly caused by difficulties for the enzyme to cleave the carboxytelopeptide Asn_3-Gly_4 bond which replaces Ser_3-Gly_4 found in mammals. This is probably as a result of the formation of a modified β-aspartyl residue linked through the side-chain β-carboxyl group. In order to investigate possible factors further, principal component analysis of the 19 bovine samples was carried out that included glycine concentrations and Asx racemisation values (Fig. 6.6). The S/N of all five fragments were observed to be co-related on the first principal component but this is primarily caused by increased concentrations, and apparently not related to Asx racemisation.

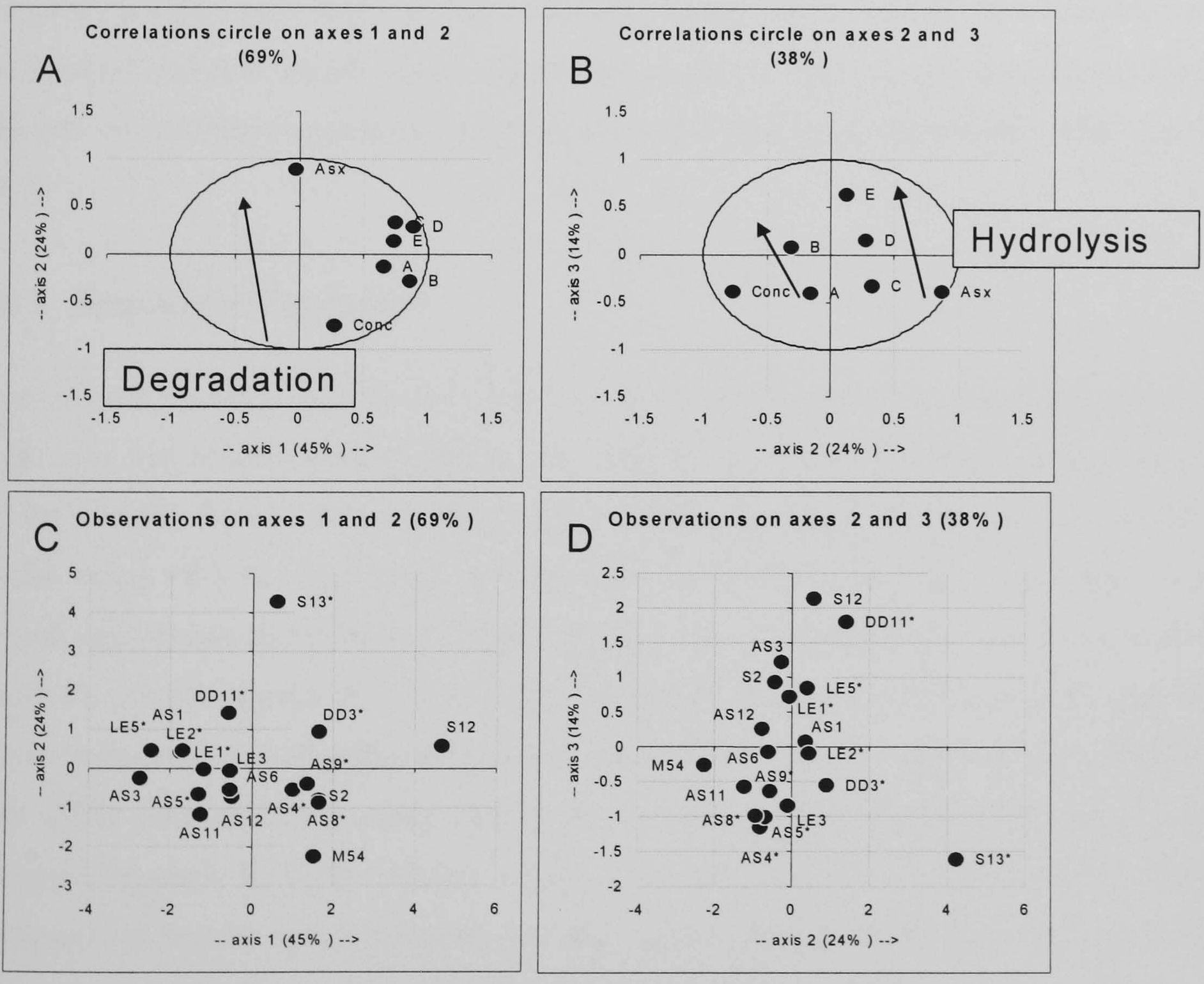


Figure 6.6 – PCA plots of the S/N values of all five a2 (I) carboxytelopeptide fragments, glycine concentrations and Asx racemisation values from 19 bovine archaeological samples; A) loadings of 1st and 2nd principal components, B) loadings of 2nd and 3rd principal components C) distribution of samples in 1st and 2nd principal components and D) distribution of samples in of 2nd and 3rd princpal components. Arrows indicate direction of increasing carboxy-terminal hydrolysis.

The principal component analyses shown in Figure 6.6 suggests that enzymatic cleavage increases with concentration (compare fragments A and B to C, D and E), whereas samples with higher Asx D/L are less likely to undergo enzymatic cleavage at the Ser_3-Gly_4 bond of the carboxytelopeptide. The enzyme may not be interacting with the more-degraded collagen due to changes in structure, and thus cleaving the peptide less. This cleavage does not appear to occur in the Asn_3-Gly_4 bond of the giant tortoise or dodo telopeptides. Interestingly, the increasing hydrolysis of carboxyterminal bonds (fragment D = E-1 carboxyterminal amino acid, fragment C = E-2 carboxyterminal amino acids, fragment A = B-1 carboxyterminal amino acid), does not appear to be related to either amino acid concentration or Asx D/L, and thus must be affected by some other factor not accounted for, such as environmental pH or preparation-induced effects. Alternatively, these intensity variations may only reflect sample-to-sample variations in the complex matrix from which the samples were fractionated and thus simply reflect differences in ion yield due to the presence of slightly different competing/suppressing components in the sample as mentioned earlier (section 5.4.1).

6.5.1 Sequence Variation

From the 43 sequences (Table 6.1), seven of the 18 residues are conserved throughout all organisms: the amino-terminal glycine, the triglycine motif (GGG) followed by a tyrosine residue and the three carboxyterminal residues YRA at positions 16-18 (the only exception to this being FRA in pipid frog). Amongst the mammals (including the common shrew, ground squirrel and microbat sequences), there are also conserved Phe residues at positions 9 and 15. As mentioned in section 5.4.2, the highly conserved triglycine motif has been shown to have a greater binding affinity for other proteins than the diglycine motif (Klarlund *et al.* 2000) present in the other telopeptides. This speculatively makes the α2 (I) chain carboxytelopeptide in bone collagen (I) an ideal candidate for the site of OC binding, particularly as it is the only telopeptide not able to cross-link, and both bone collagen (I) and OC molecules exist in equimolar amounts (Prigodich and Vesely 1997).

Sequence variations were observed between the four species of interest in the study of MBM: cattle, sheep, pig and chicken. However, no differences were observed in the

carboxytelopeptide sequences of sheep and goat, two species of interest to archaeology because of their morphological similarities. The taxonomic resolution of using this carboxytelopeptide alone varies from the sub-family level, such as Bovinae species (cattle and bison) from Caprinae species (sheep and goat) but with several inconsistencies with known phylogenetic relationships. For example, the *Bubalus* sequence is identical to the Caprinae sequence and not the Bovinae sequence. The cervid also appears to have an identical sequence as the caprines. All of the primates except bushbaby are observed on one clade, which forms a sister clade with the hare, and this combined clade forms a sister clade with the other lagomorph (rabbit) and other primate (bushbaby). This supports the closer relationship observed between lagomorphs and primates over other mammal orders in the genetic study by Murphy *et al.* (2001). The clade of carnivore species (cat, dog, fox and badger) also unexpectedly includes pig, hedgehog and platypus, each of different taxonomic order. The galliformes chicken and pheasant, of the same family (Phasianidae), and turkey of a different family (Meleagrididae) form a sister clade to the duck and dodo, each of completely different orders (Anseriformes and Colombiformes respectively). However, any speculation that avian collagen sequences are more conserved than mammalian collagen sequences would not be justified using such a small fragment (~2%) of the total sequence.

6.5.2 Applications as Species-Specific Markers Beyond aDNA

The applicability of biomolecular methods of species identification is apparent in the many palaeontological site species lists for which a substantial proportion of samples remain unidentified (Stuart 1975; Stuart 1982). Both animal and plant species are often used as indicators of climate as well as other environmental conditions and have been particularly widely investigated for the British record (Currant and Jacobi 2001; Schreve 2001; Stuart and Lister 2001). However, the older the material, the more likely that depurination effects of bone diagenesis will occur throughout the sample. This research clearly demonstrates the longer-term survival of even the non-helical telopeptide, which is not a cross-linked part of bone collagen (I), over aDNA and its potential as a species-diagnostic marker to be readily applied to answer particular archaeozoological questions in fragmentary bone specimens.

6.6 Conclusions

The simple method of isolating the collagen α2 (I) chain carboxytelopeptide described in Chapter 5 is demonstrated to be ideal for discriminating between morphologically-undiagnostic bone fragments that fail aDNA analyses. As a marker for species, it is capable of discerning between farmed domesticates (ie., differentiating cattle from sheep and pig) and between humans and most other non-primate animals, but is limited amongst other taxa, such as not being able to distinguish between sheep and goat. It has great potential for applications to palaeontology and archaeology due to the well-known survival of collagen (I) in ancient bone but a suite of collagen peptides should be targeted if more accurate species determination is required.

7 Isolating Collagen 'Helix' Peptides as Species-Specific Markers in Bone

7.1 Abstract

Although the collagen a2 (I) chain carboxyterminal telopeptide of bone collagen exhibits sufficient amino acid differences to distinguish between cattle, sheep, pig and chicken, the four species that form the major constituents of MBM feed, it is of limited application for speciation in archaeological and palaeontological bone samples. This is because the carboxytelopeptide is only capable of distinguishing between species that are simple to distinguish between using morphological criteria, even with all but the most severely fragmented material. The limitations of using collagen (I) peptide sequences obtained following digestion with trypsin and isolation by SPE for speciation of archaeological bone was investigated in greater detail. This alternative method of digestion and isolation of several potentially species-informative collagen a2 (I) chain peptides was developed and investigated for a range of 17 species. The results indicate that the four chosen peptides are capable of distinguishing between species that could not be distinguished using the carboxytelopeptide marker, thus their species-specificity are complementary to the respective carboxytelopeptide markers.

7.2 Introduction

Studying peptides derived from collagen (I) is a promising alternative for species identification of archaeological (Nielsen-Marsh *et al.* 2002; Ostrom *et al.* 2006; Ostrom *et al.* 2000; Tuross and Stathoplos 1993) and processed food (Kim *et al.* 2005; Ocana *et al.* 2004) samples from the more conventional methods using NIRM, NIRS, or PCR. However, the data analysis of results from conventional LC-MS methods of analysing collagen (I) is complicated by the within-sample heterogeneity of extractable collagen molecules where a single collagen extract contains collagen molecules of varying ages (and thus varying extent of cross-linking) (Jackson and Bentley 1960) that can potentially reduce the comparability between samples. In order to use collagen sequence to identify species, a method was developed by using the matrix metalloprotease bacterial collagenase to release

carboxytelopeptides (Chapter 5). This was found to be successful at recovering the α2 (I) chain carboxytelopeptide of collagen (I) in the 21 archaeological samples analysed, 14 of which previously failed attempts to analyse aDNA (see Chapter 6). Although the collagen α2 (I) chain carboxytelopeptide sequences indicated sufficient variations between species of interest to the industrial applications of MBM speciation (e.g., cattle from pig), they were too limited for particular applications to archaeology and palaeontology, such as distinguishing between morphologically similar (and phylogenetically more closely related) species like sheep and goat (Boessneck 1969). The results from section 5.3.2 showed that, of the mammalian species, the approach of isolating the α2 (I) chain carboxytelopeptide and analysing it by MALDI-MS could only be used to distinguish between species that are moderately simple to identify using morphological criteria, even with all but the most severely fragmented archaeological material.

The aims of the work described in this chapter were to investigate the several collagen (I) peptides for speciation of archaeological bone, using the enzyme trypsin and SPE for the isolation of potentially species-informative collagen peptides. A suitable method that is capable of distinguishing between cattle, pig, sheep and goat was determined, it allowed identification of the set of elution gradient concentrations that resulted in the best quality (determined by S/N) and most reproducible peaks (repeatedly observed in all of the triplicate analyses) observed in the analysis of cattle, pig, sheep and goat bone 'collagen' samples. Four α2 (I) chain peptides in one fraction and one α2 (I) chain peptide in a second fraction were repeatedly observed using the chosen method. The sequences of these four peptides were interpreted from product ion mass spectra for a total of 17 species and investigated for their species-specificity.

7.3 Experimental

7.3.1 Collagen Peptide Isolation and MALDI-MS Analysis

Samples of bone were defatted and powdered, and an 'acid-insoluble collagen' residue obtained as described (see section 3.3.3). Approximately 2 mg 'acid-insoluble' collagen was then gelatinised by heating at 65°C in 380 μL 50 mM ammonium bicarbonate for 3 h (for

method development see Appendix 3.1). Following gelatinisation, the samples were centrifuged at 13,000 *x g* for 15 min to precipitate ungelatinised protein and the supernatant digested with trypsin. 5 μL of 1 μg/μL trypsin solution was added to 380 μL sample and incubated at 37°C for 18 h. The SPE cartridges were prepared with one bed volume of 100% MeOH, followed by 5% ACN/95% H_2O/0.1% TFA and then 90% ACN/10% H_2O/0.1% TFA. Following enzymatic digestion, the samples were centrifuged at 13,000 *x g* for 10 min and the supernatant applied to the SPE cartridge. A stepped gradient of increasing ACN concentration following several wash steps (2 x 5% ACN and 1 x 22% ACN in 0.1% TFA in H_2O) was applied to the column and the eluting fractions (26% and 32% ACN in 0.1% TFA in H_2O) collected and dried down in a centrifugal evaporator (Fig. 7.1).

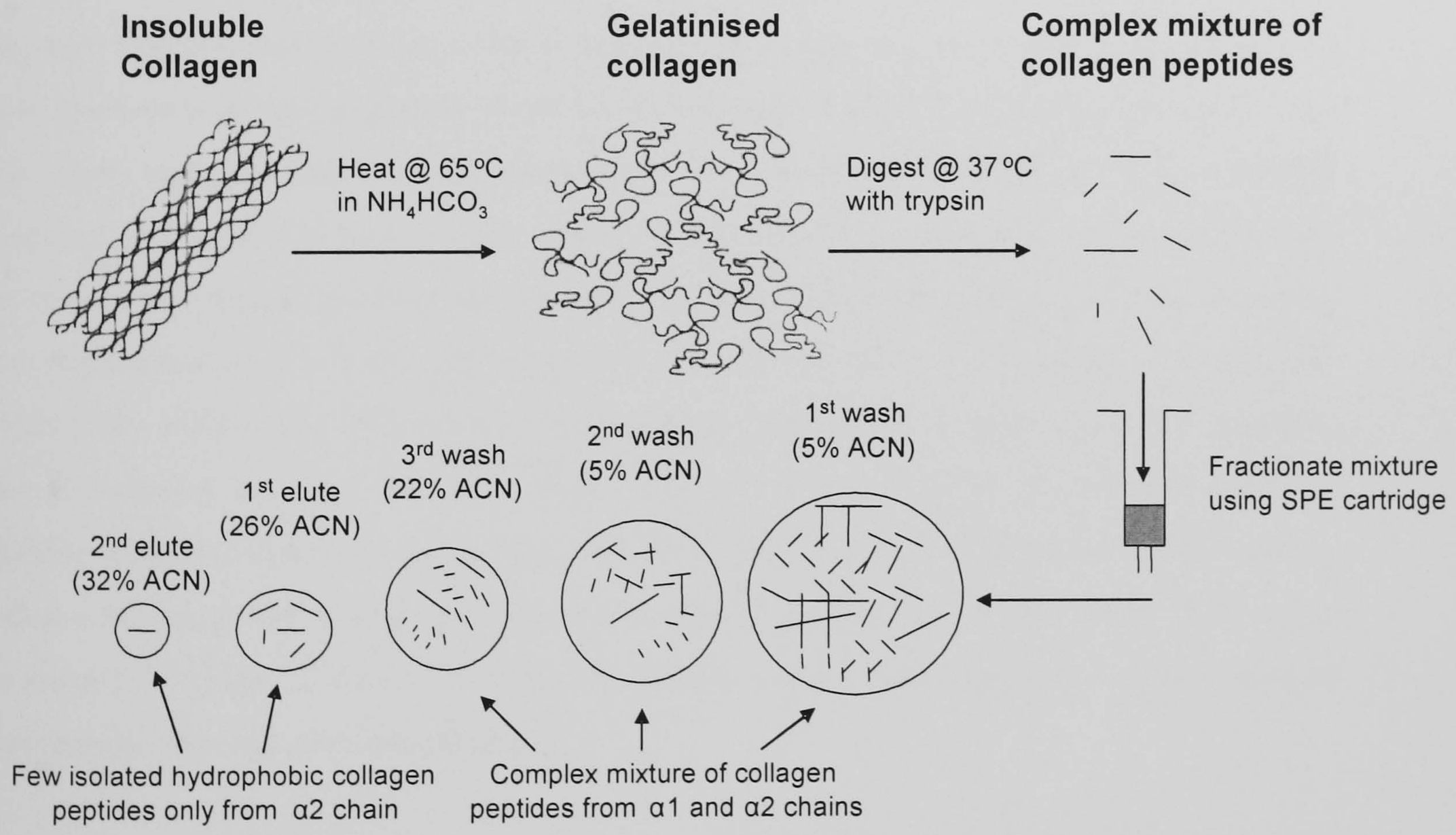


Figure 7.1 – Flow diagram of experimental procedure for isolating tryptic peptides from gelatinised bone collagen (I)

The samples were resuspended in buffer (1% OG in 50 mM ammonium bicarbonate) and diluted in 5% ACN/95% H_2O/0.1% TFA (v:v:v). 0.5 μL of sample solution from each SPE fraction was spotted onto an Applied Biosystems 192 well target plate with 0.5 μL of α-cyano-4-hydroxycinnamic acid matrix solution (1% in ACN/H_2O 1:1 v:v) and allowed to dry. Each spot was analysed by MALDI-MS in reflectron mode using a calibrated Applied

Biosystems 4700 Proteomics Analyzer. Peptide masses of a mix of standard peptides were used for a wide range calibration from 0.8 kDa to 4 kDa as described (section 5.2.3). *De novo* sequencing of the samples was carried out by manually interpreting the spectra obtained on CID product ion MS (MS/MS) of selected peptides. The product ion mass spectra were also searched against the UniProt database (January 31st 2007) via Mascot version 2.2 using Applied Biosystems GPS Explorer version 3.6. S/N values were obtained from Applied Biosystems Data Explorer version 4.3.

7.4 Results

7.4.1 The identification of the species-specific collagen (I) peptides

Initially, the most appropriate elution gradient for reproducibly obtaining a set of tryptic peptides that differ in m/z values for samples of cattle, pig, sheep and goat bone collagen (I) was developed by trial and error. Method development focussed on distinguishing between the sheep and goat bone samples because of the well-known difficulties in distinguishing between post-cranial bones of these species, particularly when fragmented (see Chapter 8). In the chosen elution gradient, the majority of peptides pass through the column following the application of 2 x 1 mL 5% ACN/95% H_2O/0.1% TFA (v:v:v) and 1 x 1 mL 22% ACN/78% H_2O/0.1% TFA (v:v:v). Two fractions of peptides are then eluted and collected, the first eluting in 1 x 1 mL 26% ACN/74% H_2O/0.1% TFA (v:v:v), which yields four distinct peptides within the range m/z 1400-3700 (Figs 7.2 & 7.3), and the second fraction often containing only a single peptide, also present in the previous (26% ACN) fraction but at lower S/N (Figs 7.2 & 7.3). In order to identify each of these peptides, MALDI product ion spectra were recorded and analysed.

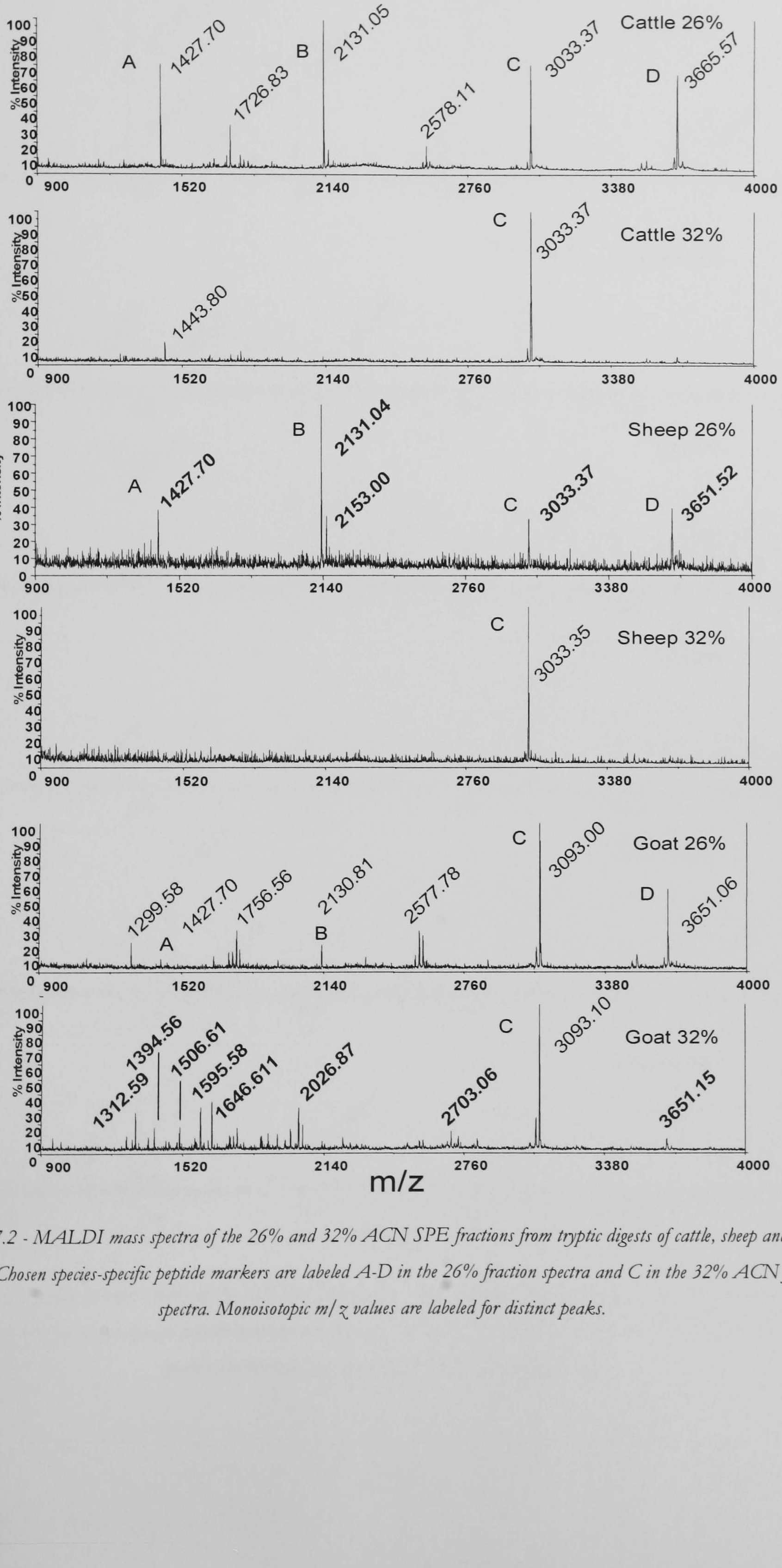


Figure 7.2 - MALDI mass spectra of the 26% and 32% ACN SPE fractions from tryptic digests of cattle, sheep and goat collagen.. Chosen species-specific peptide markers are labeled A-D in the 26% fraction spectra and C in the 32% ACN fraction spectra. Monoisotopic m/z values are labeled for distinct peaks.

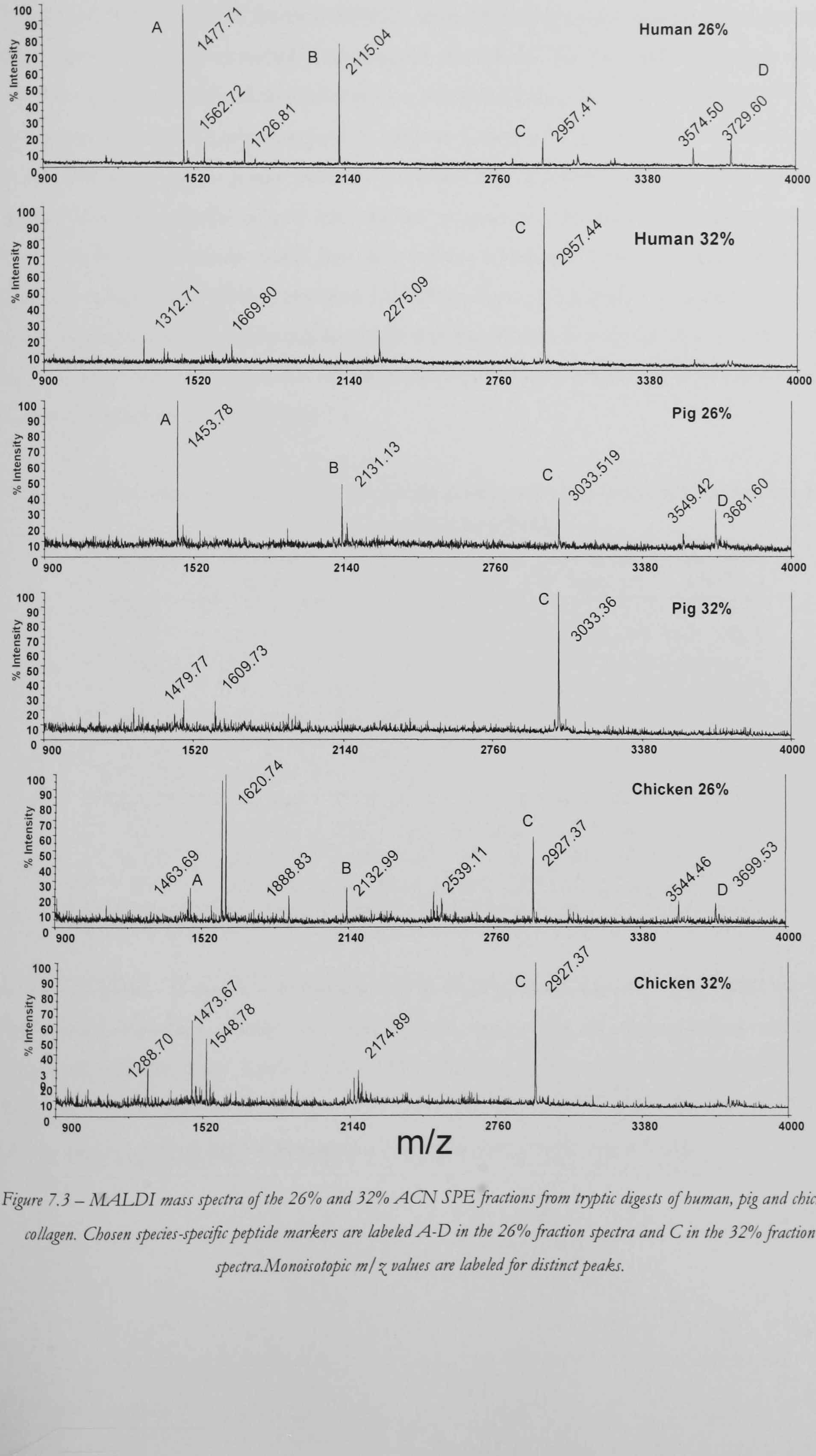


Figure 7.3 – MALDI mass spectra of the 26% and 32% ACN SPE fractions from tryptic digests of human, pig and chicken collagen. Chosen species-specific peptide markers are labeled A-D in the 26% fraction spectra and C in the 32% fraction spectra.Monoisotopic m/z values are labeled for distinct peaks.

In each of the 26% ACN fraction MALDI mass spectra for each species, there are often more than four peaks observed. For example, the spectra for the ruminant species (cattle, sheep and goat) samples often exhibited two additional peaks at m/z 1726.8 and 2578.1, but these were not consistently observed in replicate analyses. The 32% ACN fraction spectra usually contained only a single peak (m/z ~2900-3100 depending upon species) but in the spectra from the analysis of goat and chicken samples, several other peaks were observed. Many of these additional peaks had m/z values matching those of tryptic peptides of digested collagen observed in previous LC-MS analyses (see Chapter 4; Appendix 3.3) but these potential matches could not be confirmed by product ion spectra because the quality was too low. The m/z and S/N of the peaks of the four selected peptides (A-D) for 12 different species are listed in Table 7.1.

Table 7.1 – The m/z values (monoisotopic) and S/N values (in parentheses) of the species-specific peptide markers obtained from the 12 species analysed by MALDI-MS.

Sample	*m/z* Peptide 26%A	*m/z* Peptide 26%B	*m/z Peptide* 26%C	*m/z* Peptide 26%D	*m/z* Peptide 32%C
Cattle	1427.7 (39.15)	2131.1 (47.43)	3033.4 (25.76)	3665.6 (13.05)	3033.4 (38.46)
Pig	1453.7 (54.10)	2131.1 (24.33)	3033.4 (12.83)	3682.5 (14.67)	3033.4 (13.42)
Sheep	1427.7 (34.10)	2131.0 (70.83)	3033.4 (35.46)	3651.5 (45.38)	3033.4 (34.63)
Goat	1427.7 (21.48)	2131.0 (93.64)	3093.1 (64.35)	3651.4 (26.70)	3093.0 (53.28)
Chicken	1473.6 (14.48)	2133.0 (12.33)	2927.4 (28.06)	3699.5 (7.43)	2927.4 (18.79)
Human	1477.7 (179.98)	2115.0 (135.00)	2957.4 (20.93)	3729.6 (12.98)	2957.4 (70.16)
Water Buffalo	1455.6 (198.47)	2130.9 (135.54)	3075.2 (12.82)	3653.2 (47.70)	3075.2 (39.27)
Red Deer	1427.6 (12.18)	2130.9 (28.96)	3033.1 (28.81)	3651.1 (9.43)	3033.0 (44.29)
Hare	1453.6 (28.18)	2128.9 (45.46)	2957.1 (10.46)	3658.0 (8.78)	2957.1 (32.00)
Rabbit	1453.6 (22.76)	2128.9 (23.57)	2957.2 (5.00)	3642.0 (13.33)	2957.1 (9.42)
Fox	1437.3 (8.97)	2130.9 (53.30)	2999.2 (46.84)	3681.2 (13.08)	3000.2 (7.57)
Roe Deer	1427.5 (62.23)	2130.8 (87.58)	3059.1 (9.35)	3651.1 (1.52)	3060.2 (70.22)

Almost all of the 12 species analysed could be distinguished using the m/z values of these four peaks, including sheep and goat, which could not be distinguished using the carboxytelopeptide alone (section 5.3.2). The only two species analysed that could not be differentiated are sheep and red deer. This is unexpected because other more closely related species such as goat could be distinguished using at least one of these markers.

The relative S/N values of the peaks (between each sample) for Peptides A, B and C from the 26% ACN fraction varied between samples of different species, whereas the peak for Peptide D was often of relatively low S/N. Peptide C in the 32% ACN fraction was usually observed to have peaks of higher S/N than the Peptide C peak in the 26% ACN fraction. Thus in samples known to be either sheep or goat (see Chapter 8), the 32% ACN fraction may be more suitable for retrieving a better signal for Peptide C than the 26% ACN fraction.

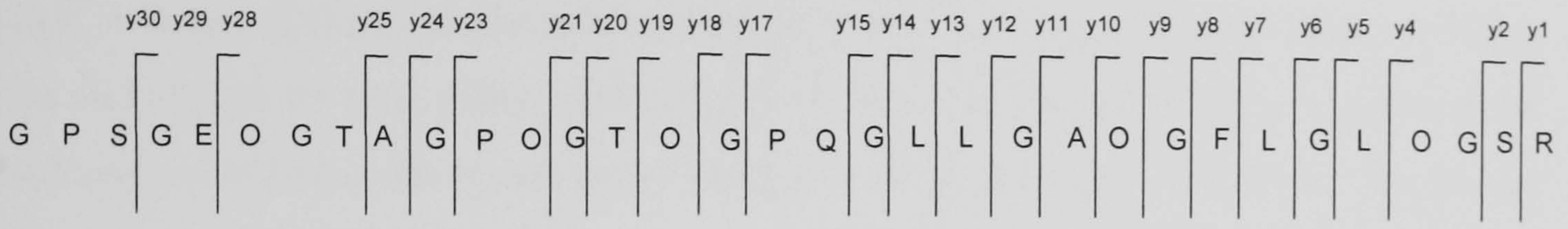
The peak lists generated for the product ion spectra of the four peptides from cattle, human and chicken collagen (I) were searched by Mascot against the UniProt public database because these were the only species that had published collagen sequences (Table 7.2).

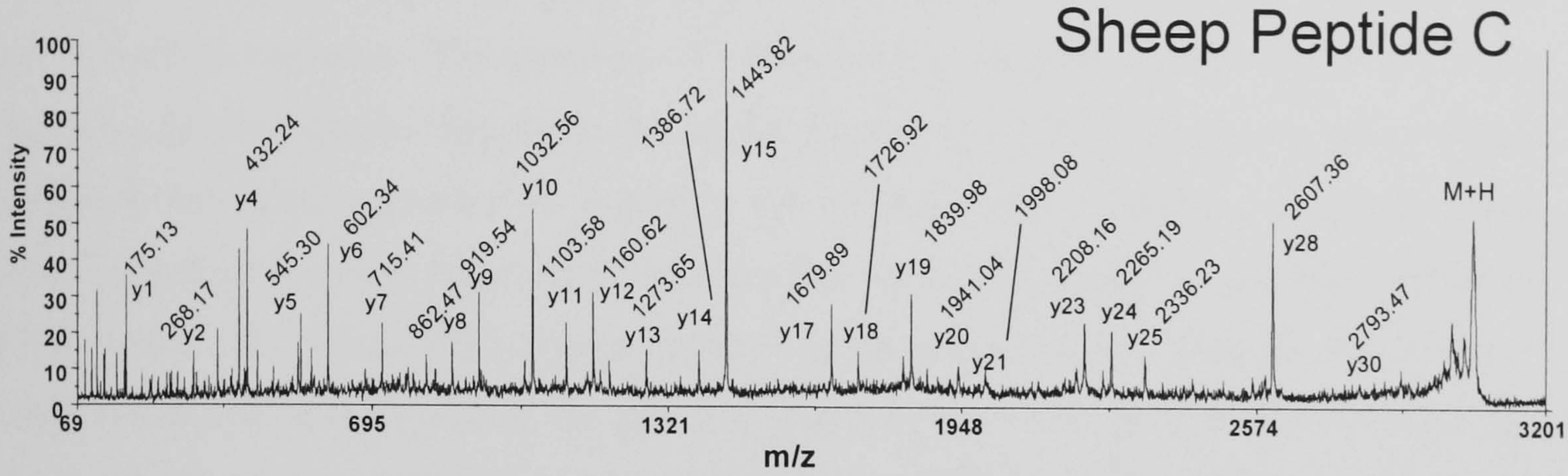
Table 7.2 - Sequences of the four matched cattle, human and chicken collagen tryptic peptide markers including observed (monoisotopic) m/z and relative position in the procollagen sequence.Underlining indicates hydroxylation of proline residues. Dots represent conserved sequence to the cattle sequences.

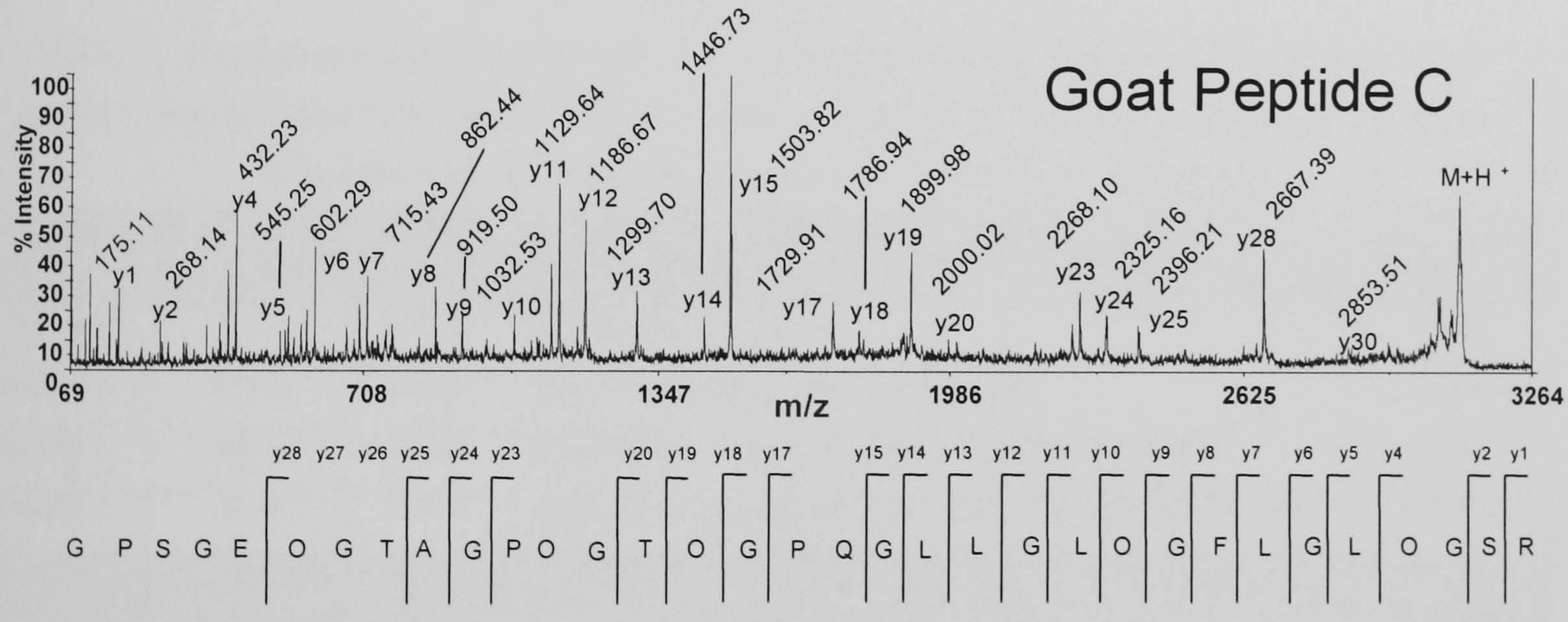
Cattle	**Observed** *m/z*	**Calculated** *m/z*	**Collagen α2 (I) chain sequence**	**Procollagen Position**
Peptide A	1427.7	1427.58	GIPGEFGLPGPAGAR	572-586
Peptide B	2131.1	2131.37	GLPGVAGSVGEPGPLGIAGPPGAR	881-904
Peptide C	3033.4	3033.29	GPSGEPGTAGPPGTPGPQGLLGAPGFLGLPGSR	845-877
Peptide D	3665.6	3665.79	GSQGSQGPAGPPGPPGPPGPPGPSGGGYEFGFDGDFYR	1079-1116
Human	**Observed** *m/z*	**Calculated** *m/z*	**Collagen α2 (I) chain sequence**	**Procollagen Position**
Peptide A	1477.7	1477.64	.LH.........P.	574-588
Peptide B	2115.0	2115.38	.......A...............	883-906
Peptide C	2957.4	2957.24	.....A..................I.....	847-879
Peptide D	3729.6	3729.88	.P..H................V....D...Y.....	1077-1114
Chicken	**Observed** *m/z*	**Calculated** *m/z*	**Collagen α2 (I) chain sequence**	**Procollagen Position**
Peptide A	1463.6	1463.62	.LH....V.....P.	572-586
Peptide B	2132.0	2032.35	....I..AT.......VS.....	881-904
Peptide C	2927.5	2928.21	.....A.A..........I.....I.....	845-877
Peptide D	3699.5	3698.84	..H..................N.....V...AE...	1079-1116

By identifying the four most prominent peptides of cattle, human and chicken 'collagen' in the isolated fractions it was clear that these peptides were from the same tryptic peptides (i.e., identical relative positions in the molecule) between samples from different species

(Table 7.2). Although suitable product ion spectra could be obtained for Peptides A, B and C of cattle, human and chicken samples, the sequence for Peptide D (the α2 (I) carboxytelopeptide-containing fragment) could only be identified by the interpretation of product ion spectrum from the cattle sample. However, the m/z value of the Peptide D ion appeared to match the expected value for this peptide from the known sequences of human and chicken collagen (m/z 3729.6 and m/z 3699.5 respectively). It is interesting to note that all four peptides are from the α2 (I) chain of collagen (I) rather than the twice as abundant α1 (I) chain (section 1.6.2). For those species that do not have collagen (I) sequences available in public protein databases, the product ion spectrum obtained for each peak listed in Table 7.1 was manually interpreted for *de novo* sequencing (e.g., Fig. 7.4).







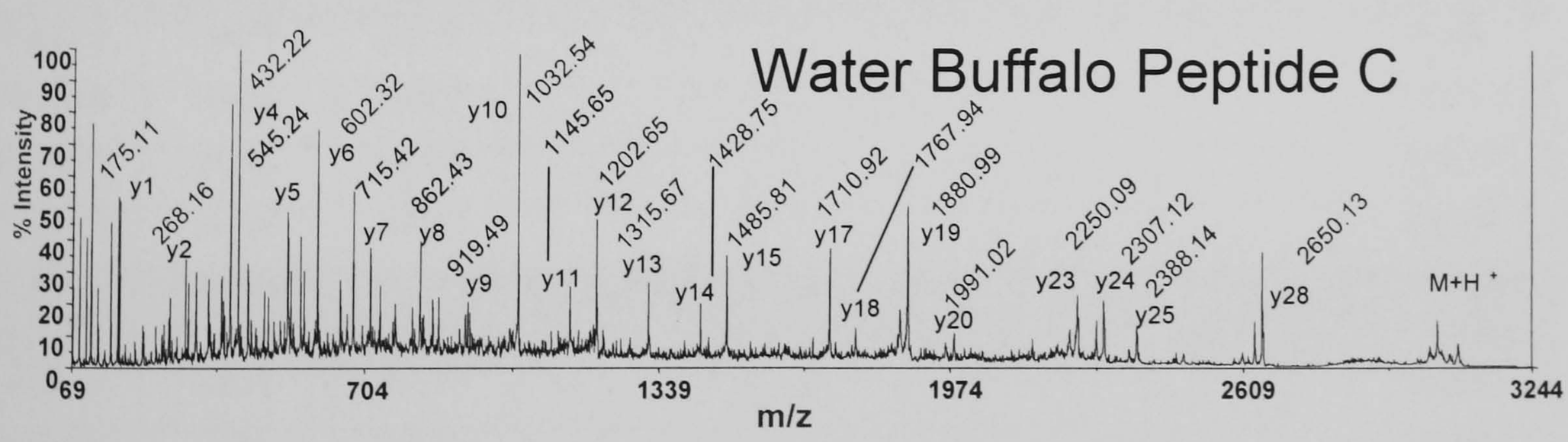


Figure 7.4 - Three examples of manual interpretation for de novo sequencing of product ion spectra showing the identification of Peptide C sequences for sheep, goat and water buffalo Peptide C.

Peptide C is clearly different in sheep (observed m/z 3033.4) and goat (observed m/z 3093.1) samples and differs by two amino acids (Fig. 7.4). The amino acid residue at position 20 within Peptide C is Leu in sheep and water buffalo, but Phe in the goat sequence. The amino acid at position 24 is Ala in sheep, Pro in goat and Leu (or Ile, an isobaric residue) in the water buffalo sequence. The patterns of fragmentation observed in the spectra of these typical collagen peptides appear to be similar throughout. The well-known effect where fragment ions amino-terminal to a proline (or hydroxyproline) residue are usually much more abundant than the fragment ions carboxy-terminal to the proline residue (Breci *et al.* 2003; Vaisar and Urban 1996) can be observed in these spectra. For example, the y9 ion is often much less abundant than the y10 ion (Fig. 7.4 sheep and water buffalo Peptide C), except in the goat sequence where the y10 amino-terminal amino acid is also a proline, and thus this effect is observed between the y10 ion and the y11 ion for the goat peptide (see Fig. 7.4).

Table 7.3 - Sequence information from interpretation of product ion spectra of the four sheep, goat and pig collagenous tryptic peptide markers including observed (monoisotopic) m/z and relative position in the procollagen sequence. Underlining indicates hydroxylation of proline residues. Dots represent conserved sequence to the sheep sequences.

Sheep	Observed *m/z*	Calculated *m/z*	Collagen α2 (I) chain sequence	Procollagen Position
Peptide A	1427.7	1427.58	GIPGEFGLPGPAGAR	572-586
Peptide B	2131.0	2131.37	GLPGVAGSVGEPGPLGIAGPPGAR	881-904
Peptide C	3033.4	3033.29	GPSGEPGTAGPPGTPGPQGLLGAPGFLGLPGSR	845-877
Peptide D	3651.5	3651.79	GSQGSQGPAGPPGPPGPPGPPGPSGGGYDFGFDGDFYR	1079-1116
Goat	**Observed *m/z***	**Calculated *m/z***	**Collagen α2 (I) chain sequence**	**Procollagen Position**
Peptide A	1427.7	1427.58	. . . . .	572-586
Peptide B	2131.0	2131.37	. . . . .	881-904
Peptide C	3093.3	3093.29	. . . . . F . . P	845-877
Peptide D	3651.4	3651.79	. . . . .	1079-1116
Pig	**Observed *m/z***	**Calculated *m/z***	**Collagen α2 (I) chain sequence**	**Procollagen Position**
Peptide A	1453.7	1453.58	. . . . . L . . P .	572-586
Peptide B	2131.1	2131.37	. . . . .	881-904
Peptide C	3033.4	3033.29	. . . . .	845-877
Peptide D	3683.5	3682.79	. . . . . YE	1079-1116

The similarities of the goat collagen α2 (I) chain peptides with sheep collagen α2 (I) chain peptides confirms that the chosen species-specific peptides are from the same position in the protein. Despite the lack of published sequence information for the three species listed in Table 7.3 (sheep, goat and pig), all four peptides were identified by interpretation of product ion spectra. However, the amino-terminal nine amino acids of Peptide D could not be confirmed by interpretation of product ion mass spectra due to the poor quality of the spectra; the presence of the highly conserved GSQGSQGPA sequence is supported by the observed m/z values for each precursor ion (where part of the Peptide D sequence was previously identified in the carboxytelopeptide sequences (position 22-38 of Peptide D) presented in section 5.3.2 for pig, sheep and goat specimens). Peptide C was the only peptide to show amino acid differences between sheep and goat (Leu to Phe and Ala to Pro transitions (see Table 7.3; Fig. 7.4)). However, Peptide C and Peptide B sequences of pig match those of cattle where Peptides A and D have only one and two amino acid differences respectively. Because these 'helix' peptide sequences (from the helical region of the collagen molecule), at least in conjunction with the telopeptide isolation method (by digestion with bacterial collagenase), showed variation between species that had identical carboxytelopeptide sequences (sheep and goat), these 'helix' peptides were investigated for a further six species (Table 7.1). The calculated $M+H^+$ m/z values of the peptides of an additional five species of sequences available through UniProt are presented in Table 7.4.

Table 7.4 - Sequence information of the chosen species-specific tryptic collagen-peptide markers from five species with published (in the UniProt protein database) sequence information (not analysed by MALDI-MS) including dog, mouse, rat, bullfrog and rainbow trout, showing calculated (monoisotopic) m/z values and relative position in the procollagen sequence. Underlining indicates hydroxylation of proline residues. Dots represent conserved sequence to the dog sequences.

Dog	*Calculated m/z*	*Position*	Peptide sequence (Acc. O46392)
A	1437.75 or 1453.75	574-588	GLPGEFGLPGPAGPR
B	2131.13	883-906	GLPGVAGSVGEPGPLGIAGPPGAR
C	2999.53	847-879	GPSGEPGTAGPPGTPGPQGLLGAPGILGLPGSR
D	3681.60	1081-1118	GSQGSQGPAGPPGPPGPPGPPGPPGPSGGGYDFGYEGDFYR
Mouse	*Calculated m/z*	*Position*	Peptide sequence (Acc. Q01149)
A	1437.75 or 1453.75	550-594	.....
B	2159.16	889-912	....I..AL.....S.....
C	2947.50	853-885	.....A...A.....
D	3661.62	1087-1124	.....V.....F.....
Rat	*Calculated m/z*	*Position*	Peptide sequence (Acc. P02466)
A	1437.75 or 1453.75	550-594	.....
B	2158. 14	889-912	.Q..I..AL.....
C	3003. 40	853-885	.....T.....A.....
D	3609.61	1087-1124	.....V.....F..G...
Bullfrog	*Calculated m/z*	*Position*	Peptide sequence (Acc. O42350)
A	1359.66 or 1375.63	565-579	.A..D..P..S..T.
B	2109.99	874-897	....GP..N.....S.L..SS.P.
C	2931.52	838-870	.....A.P.....AA..S.V..A.....T.
D	3501.51	1072-1009	.PA.FT.....HA.....G.DG.EY..
Rainbow Trout	*Calculated m/z*	*Position*	Peptide sequence (Acc. O93484)
A	1766.85 or 1782.85	564-582	AGNQGMPGDQGLPGPAGVK
B	2132.10	877-900	....GP.A.....RL..A.AS.P.
C	2981.45	841-873	.....S.PP.A...A....V..PS.FV.....
D	3570.60	1075-1112	..P.HL.....S..L...A..A.....QSGGY.E..

The 17 species investigated in this chapter (12 analysed by MALDI (Table 7.1) and the sequences of five additional species from UniProt (Table 7.4)) cover a wide phylogenetic range of animals (from fish to amphibians, avians and mammals), but focus predominantly on common mammal species often found on British archaeological sites, in particular the

major domesticates (cattle, sheep, goats, pigs, etc). Several deer species were included as well as commensal animals (mouse, rat, fox) and burrowing animals (such as rabbits and hares) often found as part of an archaeological assemblage. Phylogenetic analyses of the proposed sequences for each of the four species-specific peptide markers of 17 species (11 from Tables 7.2, 7.3 & 7.4 and 6 from UniProt protein database) were carried out to show their respective sequence variabilities (Fig. 7.5).

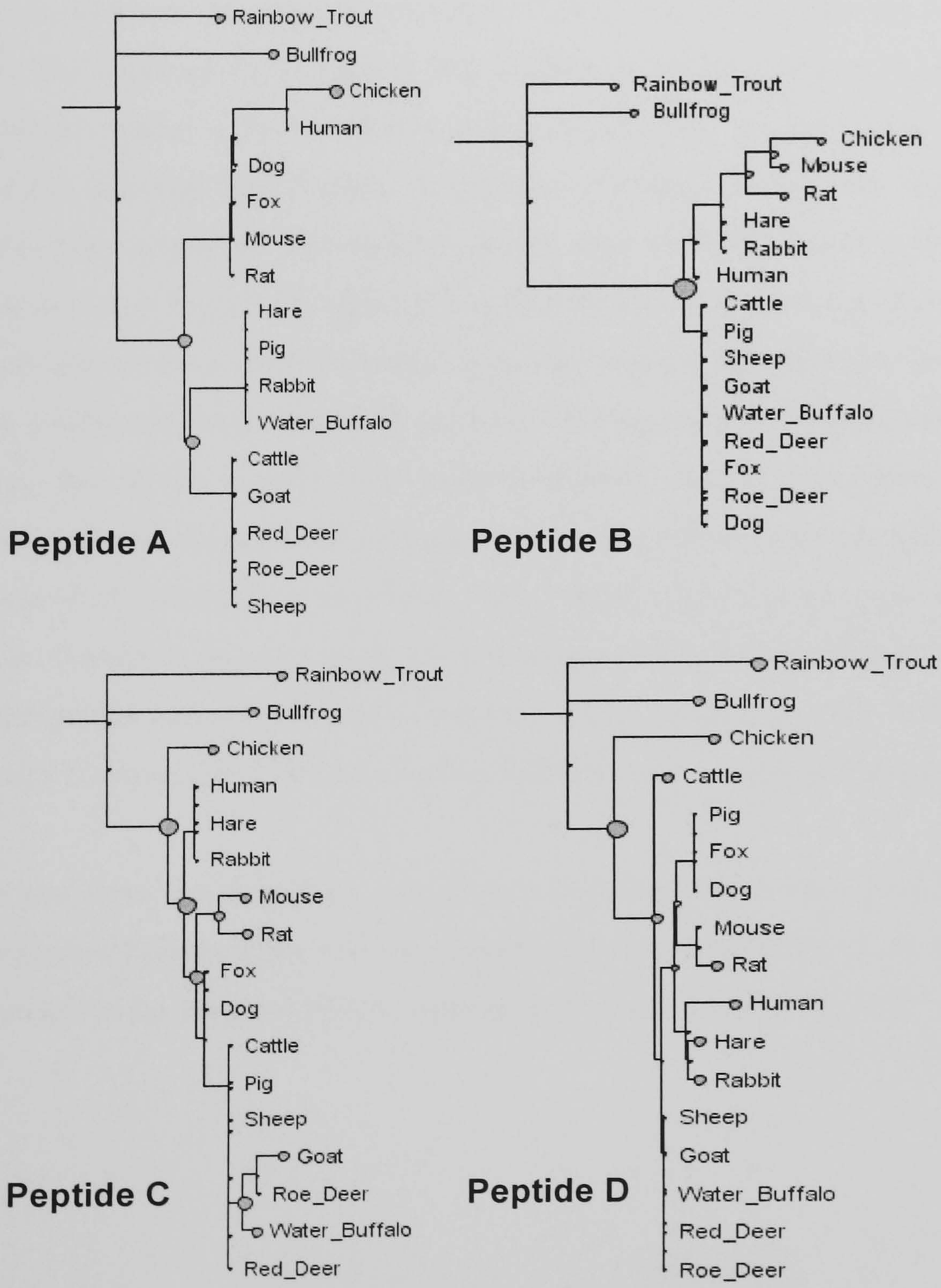


Figure 7.5 - Phylogenetic analyses of the sequences for each of the four collagen peptides for the 17 species analysed in this study (A-D)(created using Geneious 3.5.6).

The two most informative peptides of the four analysed in this study are Peptides C and D, where the mammalian taxa are divided into seven and six clades respectively. The limitations of Peptides A and B are indicated by the phylogenetic placement of the avian (chicken) with the mammals. The avian (chicken), the amphibian (bullfrog), and the fish (rainbow trout) have unique sequences (in the context of this subset of samples) for all four of the peptides, as expected regarding the proposed divergences of these taxa (see Murphy *et al.* 2001). Cattle, sheep, and red deer have the same sequences for each of the four peptides. The similarities and differences between the peptide sequences of the other species presented were more complicated. The sequence for Peptide A of pig, rabbit and hare appeared to be identical to the hydroxylated (where a Pro residue is hydroxylated) form of the sequences for dog, mouse, and rat, although the Peptide A sequence for fox matched the unhydroxylated variant of this same sequence; this sequence differs from the bovid sequence by one amino acid. The pig and dog Peptide B sequence was observed to be the same as the cattle, sheep and deer (red deer and roe deer) Peptide B sequence, but not as that of the rodents or the lagomorphs (rabbit and hare) Peptide B sequence (the lagomorph Peptide B sequence also differed from that of the rodents). The primate (human) Peptide B sequence was unique amongst the species investigated. All fox and dog peptide markers (inferred from the peptide mass) appeared to be identical to each other, with Peptide C and D sequences being specific to the canids. Peptide C appears to be of most interest to archaeozoologists as it is more variable amongst the artiodactyls. For example, roe deer, goat and water buffalo all have unique Peptide C sequences. The water buffalo Peptides A and D are also unique.

Although it was clear that Peptides C and D are more informative than Peptides A and B, the four sequences were amalgamated for a combined phylogenetic analysis in order to show the overall ability of the four peptides to distinguish between species (Fig. 7.6).

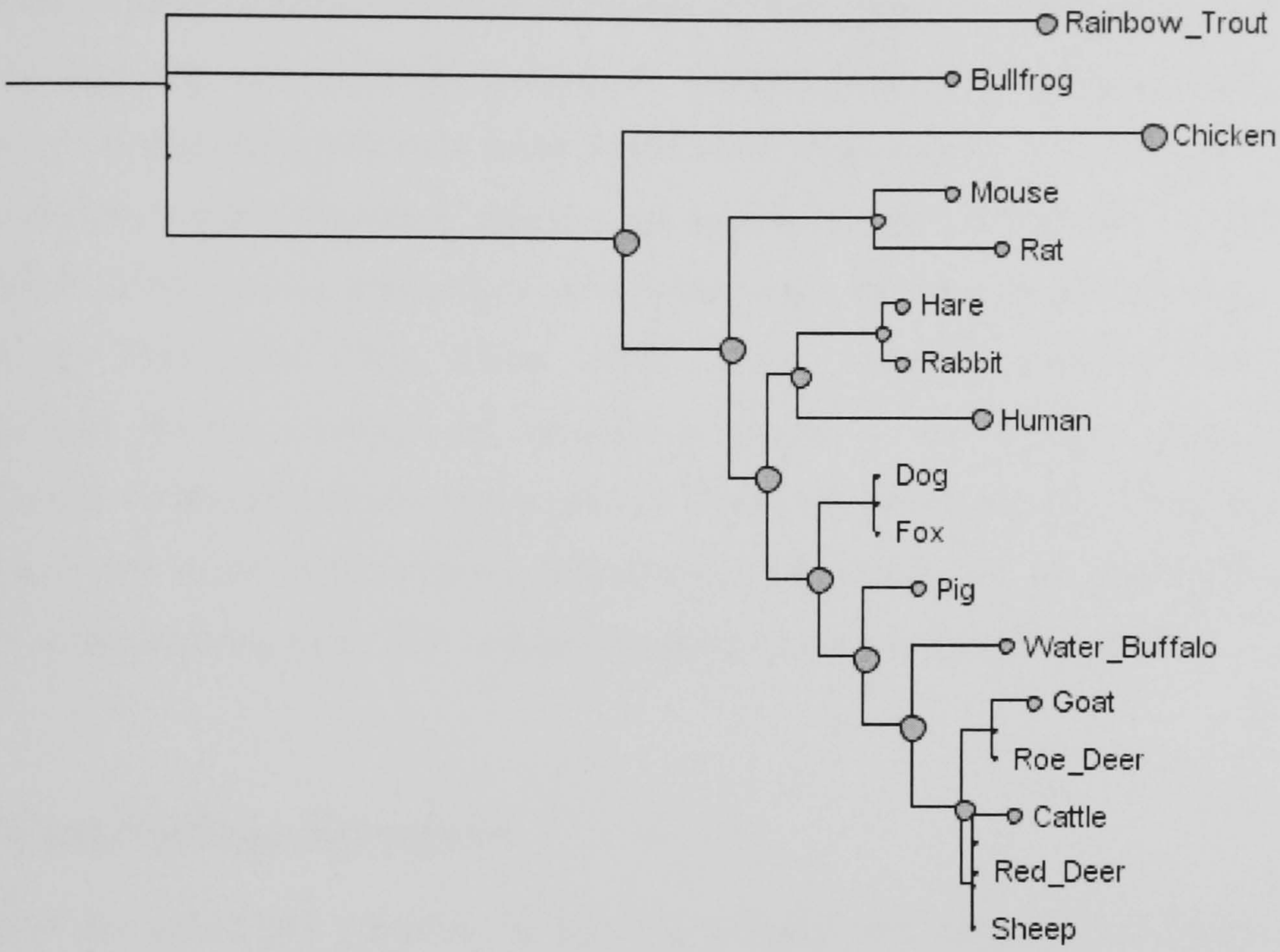


Figure 7.6 - Phylogenetic analyses to show ability to distinguish between species of the four peptides (A-D) combined (created using Geneious 3.5.6).

Using the four peptide markers collectively it is possible to distinguish between many of the species analysed in this study. The only artiodactyl species that could not be distinguished using these peptides are sheep and red deer whereas goat, roe deer and water buffalo peptides are more distinct than expected with regards to their proposed phylogenetic relationships amongst the artiodactyls (Fig. 7.6). It is also clear that fox could not be distinguished from dog using any of the peptide markers.

7.5 Discussion

All of the 17 species investigated could be distinguished using the chosen 'helix' peptide markers except red deer from sheep, and fox from dog. One of the peptides (Peptide D) contains a large proportion of the same peptide sequence as the carboxytelopeptide isolated following digestion with bacterial collagenase (section 5.2); this peptide sequence alone is variable enough to distinguish between many of the mammalian species studied. However, this relatively variable carboxytelopeptide sequence could not distinguish between any of the cervids, and some of the important bovids, i.e., between sheep and goat. It was apparent from this analysis that the Peptide C sequence supplements the use of the

carboxytelopeptide (whether obtained as Peptide D from digestion with trypsin (section 7.3) or from digestion with bacterial collagenase (section 5.2)) as a species-specific peptide marker by distinguishing between some bovid and cervid species. For example, roe deer, goat and water buffalo Peptide C markers are unique amongst the species studied. This is particularly useful because fragmented postcranial bones of these species are often difficult to identify (Boessneck 1969; Lister, 1996). Those artiodactyl species that are not distinguished by this method are considered simple to differentiate morphologically, primarily due to size differences in post-cranial bones. However, distinguishing between fox and dog, which could be problematic with fragmented material on morphological grounds, was not achieved using these four collagen peptides as species-specific markers.

7.6 Concluding Remarks

It is clear that alternative methods of digesting collagen and using SPE to isolate species-specific markers are possible. The method presented in this chapter is capable of distinguishing between species of morphological similarity which cause some difficulties for archaeozoologists, such as sheep, goat and roe deer. Further adaptation of these methods, whether using different SPE elutions or different enzymes to digest the sample, could improve the species-specificity of the collagen-peptide markers obtained and analysed, such as potentially being able to discriminate between sheep and red deer, between cattle and bison, or perhaps between other similar-sized deer species.

8 Distinguishing between archaeological sheep and goat bones using a single collagen marker

8.1 Abstract

A simple and cheap method of isolating and analysing a single collagen peptide able to distinguish between sheep and goat bones is used to test the osteological determination of 24 ovicaprid bones, and determine them for two immature specimens. The method described has advantages over other non-morphological methods of sheep/goat distinction because of its clarity and accuracy, and over DNA-based methods, due to the robustness and therefore reliability of collagen. This pilot study was carried out on archaeological bones from the 6th millennium site of Domuztepe, in south central Turkey. Amino acid racemisation analyses are also presented as a means of describing protein preservation and its effects on the survival of this marker. The results show that despite Asx D/L values >0.14 indicative of extensive biomolecular degradation, and consequently a low survival expectation for ancient DNA and osteocalcin, the targeted collagen-peptide marker was identified in all of the samples. The results also highlighted the problems in relying upon one of the morphological criteria, in this case on the distal radius, to distinguish between sheep and goat bones.

8.2 Introduction

Early evidence for the dominance of domesticated sheep (*Ovis aries*) and goats (*Capra hircus*) in the central Fertile Crescent region of the Near East is found at archaeological sites dating to the end of the Pre-Pottery Neolithic B (ca. 7500-7000 uncal. BC) (Clutton-Brock 1979; Legge 1996; Zohary *et al.* 1998). The management of sheep and goat herds will vary depending on a variety of factors, such as the needs of the communities keeping the animals, the availability of desired fodder for each taxon and the local environmental context. Although both sheep and goats provided a reliable source of meat and other primary products such as sinews and hides, they also may have been exploited for secondary products such as milk, fibers and dung (Sherrat 1981). It has been suggested that exploitation of certain secondary products would have emerged soon after domestication (Kohler-Rollefson 1992), but unequivocal archaeological evidence (in the form of carts and milking scenes, for example) does not appear until the 4th millennium BC (Sherratt 1983). A lack of unequivocal evidence is not sufficient, however, to dismiss the role of earlier exploitation of

secondary products, especially in light of the perishable nature of these products. In the absence of direct archaeological evidence for secondary products during the Late Neolithic in the Near East, investigators have drawn upon indirect evidence pertaining to differences in herd demographics.

One of the most common methods used to investigate which secondary products might have been the emphasis of particular prehistoric economies is through the investigation of mortality profiles (for example see Payne 1973). In essence, the mortality patterns are based on age curves of archaeological animal populations at a site, where age at death can be determined from bone fusion stages and tooth eruption and wear patterns (Grant 1982; Payne 1973; Zeder 2006). For example, it is believed that the most milk is available for human consumption if male lambs or kids are slaughtered soon after birth; the most meat is raised if males are killed as they reach adult body weight; and the most wool is secured by keeping castrated male sheep well into adulthood. Thus, in the archaeological record, an emphasis on dairy products, on meat, and on wool or hair should be characterised by heavy slaughter of males as infants, juveniles/subadults and adults respectively (Higham 1967; Payne 1973).

The main problem that impedes the investigation of differences in sheep and goat husbanding is that many sheep and goat skeletal elements are notoriously difficult to distinguish using morphological criteria (Boessneck *et al.* 1964; Clutton-Brock 1979; Payne 1985). This is particularly the case for bones of young individuals for which clear morphological markers have not yet developed, but also for bone fragments where few or no markers are present to distinguish the two taxa. Since most archaeological faunal assemblages contain fragmented bones, zooarchaeologists often neglect to differentiate data from sheep and goats, reporting them as 'sheep/goat', 'caprines' or more recently, 'ovicaprids' (Campbell *et al.* 1999; Clutton-Brock 1979; Croft 2003; Davis 1985; von den Driesch and Wodtke 1997). While combining the two taxa increases sample size and provides a general impression of ovicaprid exploitation, it is much more desirable to separate the sheep from the goats in order to investigate possible differences in herd management strategies of these two species. Archaeozoological analyses of mortality patterns for the

combined category "sheep/goat" thus risk masking important contrasts in management between these species or, worse still, creating an illusory composite picture which is valid for neither species. These problems are particularly acute in areas, such as the Mediterranean and Near East, where both sheep and goat have long been major constituents of livestock populations. The ability to distinguish reliably between the archaeological remains of domestic sheep and goats would have profound implications for a greater understanding of the role of primary and secondary animal products in prehistoric economies.

A range of non-morphological methods have been developed to separate sheep from goat specimens, including dietary signatures using stable carbon isotopes (Balasse and Ambrose 2005), tooth wear analyses (Grine *et al.* 1987) and aDNA based approaches (Bar-Gal *et al.* 2003; Loreille *et al.* 1997; Newman *et al.* 2002). However, all of these currently available non-morphological methods suffer from substantial limitations. The method described by Grine *et al.* (1987) based on tooth wear analysis is limited by the specific teeth which are amenable to the method. The method of Balasse and Ambrose (2005) using dietary signatures is limited by the extent of prior knowledge required about the archaeological site in question, where samples can only be reliably distinguished between sheep and goats that come from areas where potential access to C3 diet can be assessed. This is further hindered by the substantial overlap between the diet of ovicaprids, especially if human-induced foddering practices are considered, which must have some effect on the dietary signatures and different rates of tooth wear of sheep and goats (Balasse and Ambrose 2005; Wapnish and Hesse 1988).

Using the genetic code inherent in DNA is an unambiguous method and the biomolecule has been used as a powerful marker for distinguishing between sheep and goat samples of varying archaeological ages (Bar-Gal *et al.* 2003; Loreille *et al.* 1997). However, the survival of aDNA is known to be greatly effected by the thermal history of the environment it is deposited in (Poinar and Stankiewicz 1999; Waite *et al.* 1997). As a result, there are well-known difficulties in analysing prehistoric bones of Near Eastern origin, which are of primary interest to investigating the spread of agriculture. The severity of this problem has been highlighted in the reported success rates of aDNA analyses, where Newman *et al.*

(2002) only obtained 12 successful results from 27 samples of varying archaeological ages and locations and Bar-Gal *et al.* (2003) obtaining only one successful sequence amplification from two Neolithic samples of Near Eastern origin. Further to problems of survival, archaeological tissues (principally bone) suffer from difficulties regarding contamination with exogenous DNA, not only from excavators (Brown and Brown 1992) but from pre-excavation contamination such as that deriving from modern husbanding practices above archaeological sites (such as from permeating sheep urine) (Haile *et al.* 2007) or from contamination of laboratory reagents (Leonard *et al.* 2007).

Unlike DNA, proteins have long been known as the dominant biomolecules in both modern and ancient bone (Ambler and Daniel 1991), where collagen (I) is by far the most abundant, accounting for approximately 90% of the protein in bone (Millard 2001). Collagen (I) is a fibrous protein of which the basic unit is a triple helix made up of two α1 chains and a genetically distinct α2 chain. The previous chapters (Chapter 5, 6 & 7) describe methods using collagen sequences for speciation to the taxonomic level of genus in some instances, and capable of distinguishing between the two species of interest: sheep and goat (see Chapter 7). In this method a single collagen-peptide marker capable of distinguishing between the two species was reproducibly analysed by MALDI-MS. The absence of the sheep marker in collagen (I) from modern domestic goat bone was confirmed by LC-MS proteomics-based analyses, as was the absence of the goat marker in collagen (I) from modern domestic sheep bone (see Appendix 6.2). Once the described method had been developed using several long bones of known identity, it was applied to selected bones from the Domuztepe zooarchaeological assemblage due to its relatively high numbers of sheep/goat bones.

8.2.1 Domuztepe; an Archaeological Case Study

Domuztepe, located in south central Turkey (Fig. 8.1), is one of the largest known Neolithic sites in the Near East, covering ca. 20 hectares. The excavated areas of the site have been dated to ca. 5,800-5,450 cal. BC, which is considered a key period in the development of complex societies in the Near East. This is because it falls within the less-well-understood interim period between the beginning of sedentism in the Natufian (Later Epipalaeolithic in

the Levant) and fully urbanised societies of the 4th millennium BC (Campbell *et al.* 1999). The material culture characteristic of this period at Domuztepe includes stone architecture (both circular tholoi and rectilinear buildings), figurines, stamp seals, distinctive Halaf pottery, stone bowls, and obsidian implements (Campbell *et al.* 1999). It is currently believed that Domuztepe was occupied for at least 1000 years with an estimated population that peaked at over 2000 people.

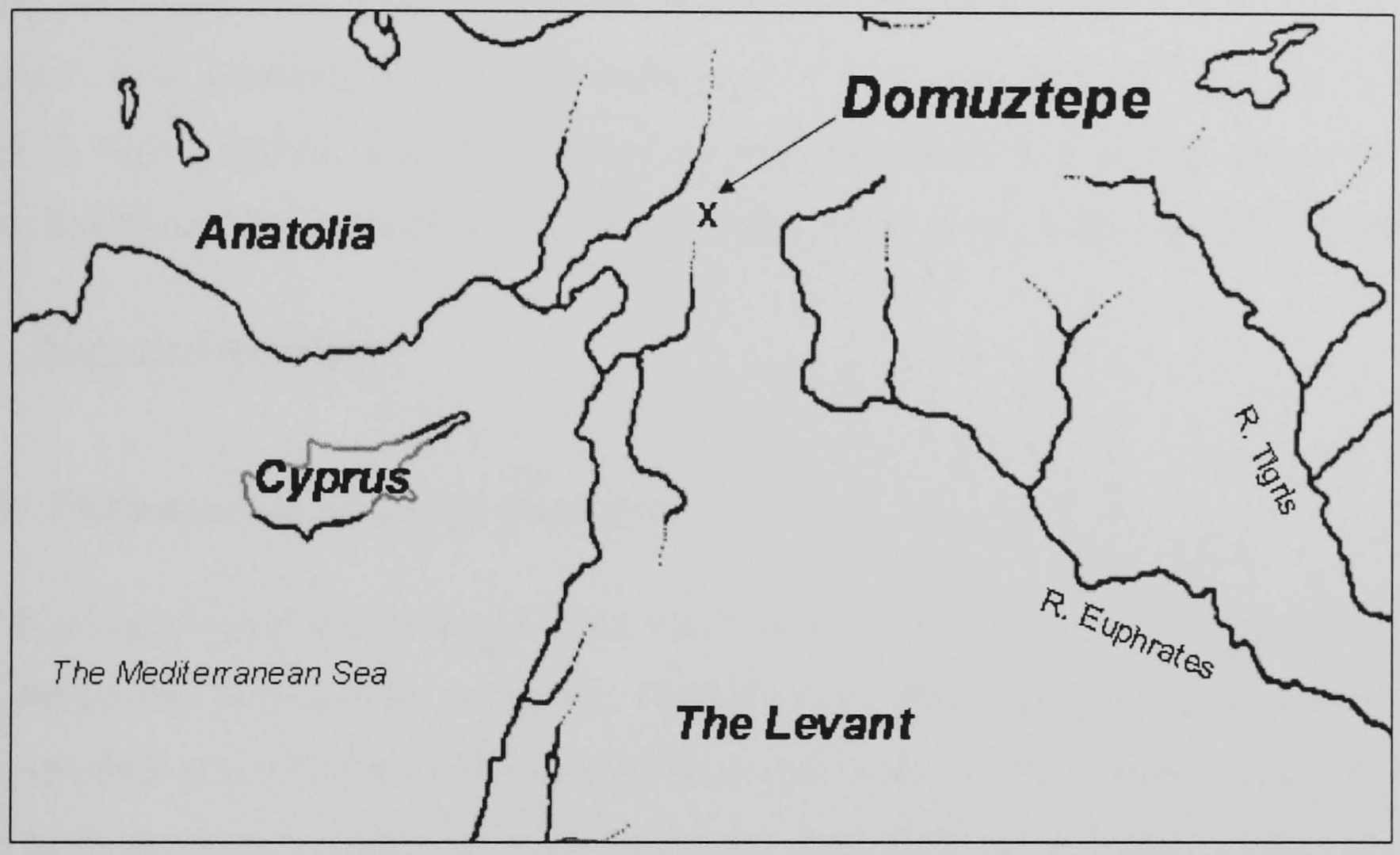


Figure 8.1 - Map of showing the location of the archaeological site of Domuztepe

The animal economy of Domuztepe was dominated by domestic sheep and goats making up 50% of the total analyzed assemblage to date. Where these can be identified, there are roughly equal numbers of sheep and goat bones in the assemblage and the sex ratio indicates that twice as many females survived to adulthood than males (Kansa, In prep.). The preponderance of adult females is a signature of domestic populations, where young males are typically killed and females are kept for breeding. The young males are often not recognised because their bones have not developed enough to determine their sex. Of the over 4000 bones identified as ovicaprids in the Domuztepe assemblage, about 15% could be assigned specifically to sheep or goat based on morphological criteria described by Boessneck *et al* (1964) and Boessneck (1969). Continuing analyses involving differences in age and sex of the sheep and goats from this site indicate that the people of Domuztepe preferrentially killed goats at a young age, presumably for their meat. Sheep, however, were kept alive well beyond the age of maximum growth, which could suggest that they valued

sheep for their secondary products such as wool (Kansa in prep.). The 26 bones chosen for the present study all come from the same context (Locus 3680) within a composite feature loosely referred to as the 'ditch'. In fact, this feature is a succession of repeatedly recut channels or hollows along an east-west alignment in Operation I at the site. The complex set of cuts and fills covers an area of about 25 m long and about 2 m wide and was probably built up over a long period with all the cuts having a distinctive fill with high levels of gleying, suggesting water-logged conditions in combination with organic material. The 'ditch' contained high concentrations of all categories of find, including well-preserved animal bones. A high proportion of these bones are morphologically identifiable sheep and goat bones, including those belonging to juveniles, making it an ideal assemblage for this study.

8.3 Experimental

8.3.1 Preparation of Bone Powder

The 26 archaeological bone samples from excavations at Domuztepe were hand selected and morphologically analysed by S. Kansa, Alexandria Archive Institute, California, prior to being shipped to the UK for collagen-peptide marker analysis. The 26 sheep and goat bones taken from the 'ditch' context for this study were selected based on their completeness and the presence of reliable morphological markers to distinguish between taxa. Table 8.1 lists the archaeological specimens selected and the criteria used to determine whether they were sheep or goat. Five samples (samples #27, #28, #41, #49 and #50) were tentatively identified as either sheep or goat as they came from immature specimens or displayed only weak features to point to a specific species. In addition two specimens (samples #26 and #32) could not be reliably distinguished because they came from immature individuals whose morphological markers were not yet prominent. For collagen-peptide marker analysis, samples of bone powder were drilled from the bones using a diamond-tipped dremel drill. Following the removal of periosteum, small pieces (approximately 10 x 10 mm) of modern long bones were cut out and defatted in chloroform/methanol (83/17%) over 30 min. After the chloroform/methanol was removed and the bone shards allowed to dry, they were ground to a fine powder with a liquid nitrogen grinding mill (Spexmill 6750). The species used as standards included domestic sheep (*Ovis aries*) and goat (*Capra hircus*).

*Table 8.1 - List of archaeological sheep and goat bone samples and morphological criteria used for identification. Morphological characteristic B1 is from Boessneck et al. (1964), B2 from Boessneck (1969); H from Hole et al (1969) and P from Prummel and Frisch (1986).*C = Complete, NC = nearly complete, PS = Proximal shaft, DS = Distal shaft.Taxon is a morphological identification.*

Sample Code	Taxon	Element	Part *	Fusion	Morphological Criteria
25	Goat	Astragalus	C		Lateral distal facet is round; medial-anterior protuberance is pronounced (P:574)
26	Sheep/ Goat	Phalanx 1	C	unfused proximal	No clear criteria present
27	Sheep?	Metatarsal	PS	fused proximal	Bone is long and narrow; proximal aspect is taller and narrow; possibly sheep
28	Goat?	Phalanx 1	C	unfused proximal	Distal articulation is v-shaped (B1:121)
29	Goat	Radius	PS	fused proximal	Long ulnar scar; small sulcus of lateral margin (B2:342)
30	Goat	Metacarpal	DS	fused distal	Bone is broad and flat; trochlear measurements indicate goat
31	Sheep	Radius	NC	fused proximal; unfused distal	Short ulnar scar; Somewhat pronounced sulcus of lateral margin (B2:342); small & young animal, both features not well-developed; bone long & thin; sheep tentative
32	Sheep/ Goat	Radius	DS	unfused proximal	No clear criteria present
33	Goat	Radius	C	fusing proximal; unfused distal	Long ulnar scar; small sulcus of lateral margin (B2:342); bone short & broad; animal very young, features not well developed; goat tentative
34	Sheep	Phalanx 1	C	fused proximal	Distal articulation is not V-shaped (B1:121)
35	Goat	Calcaneus	NC	unfused proximal	Two facets on processus anterior are attached (B1:105)
36	Goat	Humerus	DS	fused distal	Medial epicondyle is angled (B2:341); posterior portion of lateral condyle pinched
37	Sheep	Metacarpal	DS	fused distal	Relative size of trochlear condyles indicate sheep (H:271); shaft appears to be long
38	Goat	Metacarpal	PS	fused proximal	Shaft is short and wide; prox end is wide with no cranial tuberosity
39	Goat	Radius	PS	fused proximal	Long ulnar scar; small sulcus of lateral margin (B2:342)
40	Sheep	Radius	C	fused proximal, unfused distal	Short ulnar scar, but not well developed ; proximal eroded from gnawing - sulcus of lateral margin not visible; young animal; sheep tentative
41	Goat?	Metacarpal	PS	fused proximal	Proximal aspect is wide with rounded edges; goat tentative, no shaft present
42	Sheep	Calcaneus	C	unfused proximal	Two facets on processus anterior are not attached (B1:105)
45	Sheep	Astragalus	C		Lateral distal facet is elongated; medial-anterior protuberance is not pronounced (P:574)
47	Goat	Astragalus	C		Lateral distal facet is round; medial-anterior protuberance is pronounced (P:574)
48	Goat	Metacarpal	C	fused proximal, fused distal	Relative size of trochlear condyles indicates goat (H:271); bone short and broad; proximal aspect is wide
49	Goat?	Metatarsal	PS	fused proximal	Shaft is short; proximal end is round; young animal; goat tentative
50	Goat?	Phalanx 2	C	fused proximal	Distal articulation is slightly pointed with a hint of a ridge (B1:122)
51	Goat	Phalanx 2	C	fused proximal	Distal articulation is pointed with a ridge going up to the proximal end (B1:122)
52	Goat	Radius	DS	fused distal	Distal radial carpal facet is narrow and the medio-volar line is curved (B1:72-73)
53	Goat	Metacarpal	DS	fused distal	Condyles are angled towards each other; relative size of trochlear condyles indicates goat (H:271)

These 26 specimens, together with analytical notes, measurements and photos, can be accessed online in the Open Context archaeological database under the tag name "sheep/goat collagen study" (http://www.opencontext.org/).

8.3.2 Preparation of Collagen and the Isolation of the Collagen-Peptide Marker and Analysis by MALDI-MS

Collagen was isolated from the bone powders of modern and archaeological bone and particular collagen peptides were isolated following the method described in section 7.3.1 where only the second elution fraction (32% ACN) containing 'Peptide 32%C' is used (Fig. 8.2).

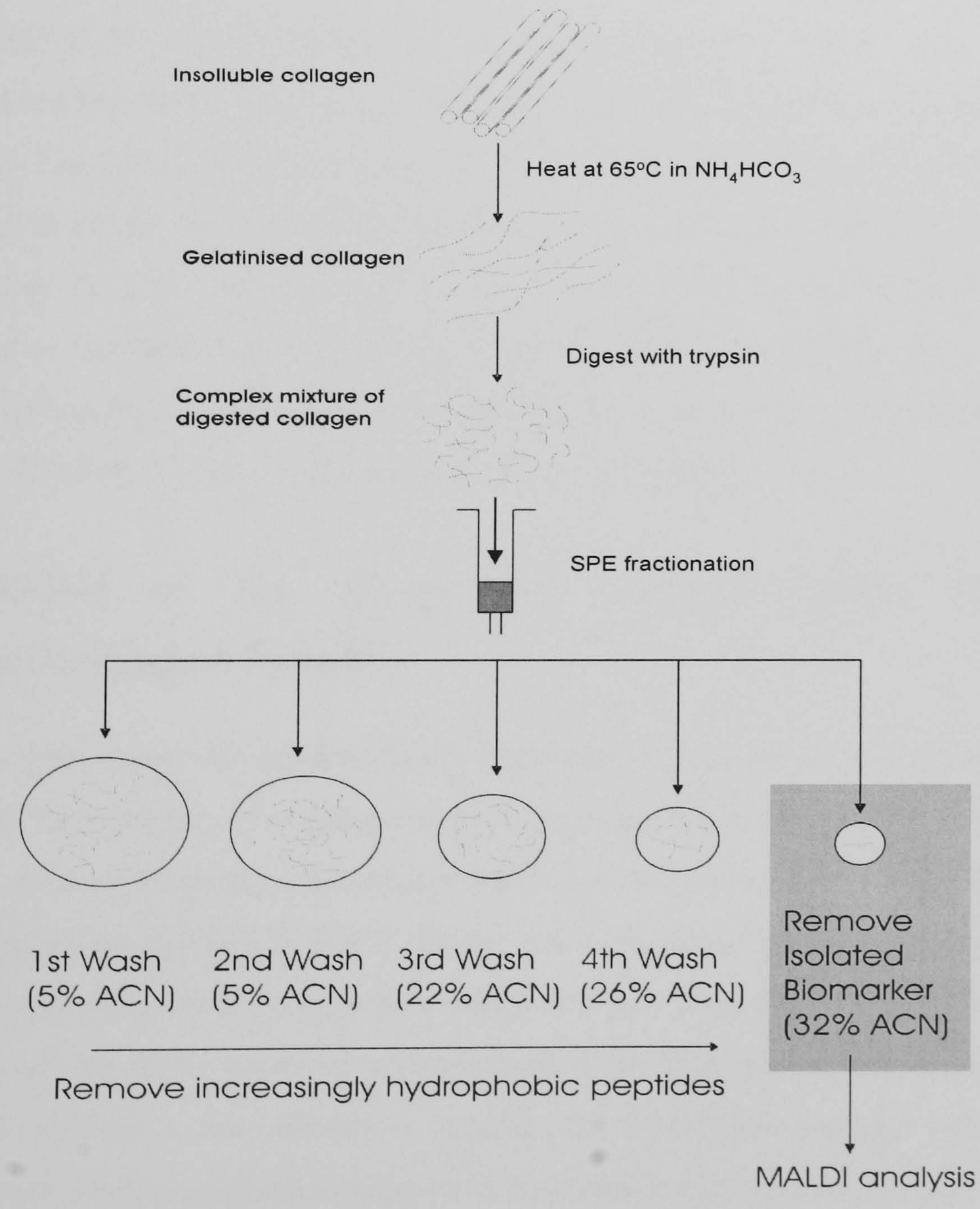


Figure 8.2 - Diagram showing the key stages in the extraction and isolation of the collagen-peptide marker.

8.3.3 Amino Acid Racemisation Analysis

Amino acid racemisation analyses were also carried out on the 26 archaeological bone samples following the methods described in section 2.2.4.

8.4 Results

8.4.1 Sequence Variation between the Sheep-Specific and Goat-Specific Collagen–Peptide Markers

The sequence obtained by *de novo* sequencing the isolated sheep peptide marker with an observed *m/z* of 3034.4 was identified as GPSGEOGTAGPOGTOGPQGLLGAOGFLGLOGSR (see section 7.4.1; Table 7.3; underlined residues differ from those of the goat peptide). The dominant peptide used as a marker was found to contain post-translational modifications (PTMs); a deamidated Glu (Q) residue and five hydroxylated Pro (O) residues. By comparison, the sequence of the isolated peptide from the goat samples, with an observed *m/z* 3094.7, was identified by *de novo* sequencing as GPSGEOGTAGPOGTOGPQGFLGPOGFLGLOGSR (see section 7.4.1; Table 7.3; underlined residues differ from those of the goat peptide). This peptide was found to contain the same PTMs; a deamidated Q and five O residues.

8.4.2 Survival of the Sheep/Goat Collagen-Peptide Marker in Archaeological Samples

The sheep/goat diagnostic peptide marker was isolated from 26 archaeological sheep/goat bones from Domuztepe, after being interpreted using morphological criteria (by S. Kansa, Alexandria Archive Institute, California)(Table 8.1), were analysed blind using the diagnostic collagen marker (see Table 8.3). The 'collagen' yields and Asx D/L values were also recorded for each sample where modern bone powders yield Asx concentrations of approximately 150 nmol/mg and Asx D/L values of approximately 0.07. The plot of Asx D/L against Asx concentration shows a clear difference between the Domuztepe samples and the younger archaeological samples analysed in Chapter 2; the Domuztepe samples appear to have much

higher Asx D/L values relative to their Asx concentrations with comparison to other archaeological samples (Fig. 8.3).

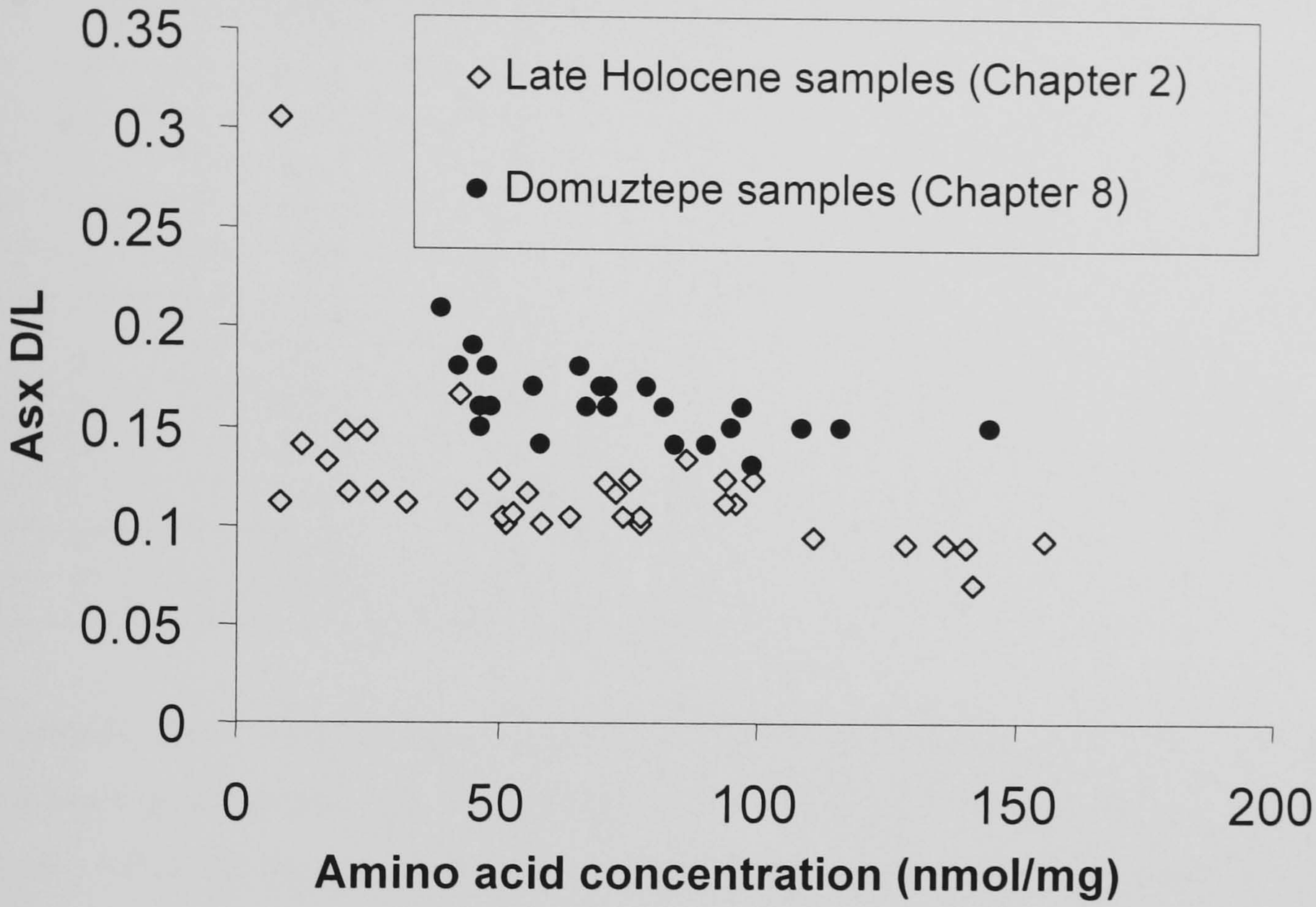


Figure 8.3 – Plot of Asx D/L vs Asx concentration of the 26 Domuztepe samples in comparision to the 36 archaeological samples from Chapter 2.

Table 8.2 – Collagen-peptide marker m/z values and distinction between sheep and goat. [a]Collagen-peptide marker analyses were carried out blind by MB without attempting morphological identifications. [b]Morphological identifications were carried out by SK but not revealed to MB until after the analyses. Asx conc. of modern bone approx. 140 nmol/mg.

Sample Code	Asx conc (nmol/mg)	Asx D/L	Peptide Marker m/z	Identification[a]	Morphological ID[b]
25	56.800	0.17	3094.2	Goat	Goat
26	48.486	0.16	3094.2	Goat	Sheep/Goat
27	46.477	0.15	3034.3	Sheep	Sheep (uncertain)
28	58.484	0.14	3094.3	Goat	Goat (uncertain)
29	46.246	0.15	3094.2	Goat	Goat
30	46.590	0.16	3094.2	Goat	Goat
31	82.081	0.16	3034.3	Sheep	Sheep
32	84.157	0.14	3094.2	Goat	Sheep/Goat
33	70.185	0.17	3094.1	Goat	Goat
34	42.196	0.18	3034.2	Sheep	Sheep
35	71.337	0.16	3094.2	Goat	Goat
36	71.357	0.17	3094.3	Goat	Goat
37	39.054	0.21	3034.1	Sheep	Sheep

Table 8.2 continued

Sample Code	Asx conc (nmol/mg)	Asx D/L	Peptide Marker m/z	Identification[a]	Morphological ID[b]
38	145.068	0.15	3094.2	Goat	Goat
39	97.4900	0.16	3094.3	Goat	Goat
40	96.991	0.16	3034.3	Sheep	Sheep
41	108.681	0.15	3094.3	Goat	Goat (uncertain)
42	65.740	0.18	3034.1	Sheep	Sheep
45	78.507	0.17	3034.3	Sheep	Sheep
47	45.514	0.19	3094.2	Goat	Goat
48	95.121	0.15	3094.3	Goat	Goat
49	47.922	0.18	3094.3	Goat	Goat (uncertain)
50	67.439	0.16	3094.3	Goat	Goat (uncertain)
51	99.588	0.13	3094.3	Goat	Goat
52	90.665	0.14	3034.1	Sheep	Goat
53	116.138	0.15	3094.3	Goat	Goat

The identifications using the species-specific collagen-peptide markers were successful in all 26 archaeological samples despite the majority of samples having Asx D/L >0.15. When analyzing ovicaprid samples, the eluting fraction usually contains only a single peak within the range of *m/z* 1000-4000. However, occasionally there are additional small peaks belonging to (un)modified forms of the same peptide (Fig. 8.4). The main example of this is the presence of the non-deamidated form of the peptide. Deamidation is a common post-translational modification that can occur during laboratory protocols, usually resulting in the conversion of an asparagine residue (N) or glutamine residue (Q) to their acid forms (D and E respectively) (Robinson and Rudd 1974). This modification alters the overall mass by +1 Da, where the calculated *m/z* of the non-deamidated sheep peptide marker is 3033.5 and the *m/z* of the deamidated sheep peptide marker is 3034.5. Likewise these are *m/z* 3093.5 and *m/z* 3094.5 for the non-deamidated and deamidated goat peptide markers respectively. The other albeit less prominent peak present in the spectra is approximately 16 Da less and indicates the presence of the same peptide with only four hydroxylated Pro residues instead of five. However, the presence of these additional small peaks or the small error in mass detection (~0.4 Da from calculated value), do not affect the simplicity of identifying sheep from goat samples as the described peptide markers are by far dominant in the spectra (Fig. 8.4) and differ by 60 Da. To confirm the identity of each recorded peak as belonging to the expected collagen sequence, and not some other contaminant peptide (such as from

bacteria), sequence information was obtained by MS/MS acquisition for each sample peptide marker.

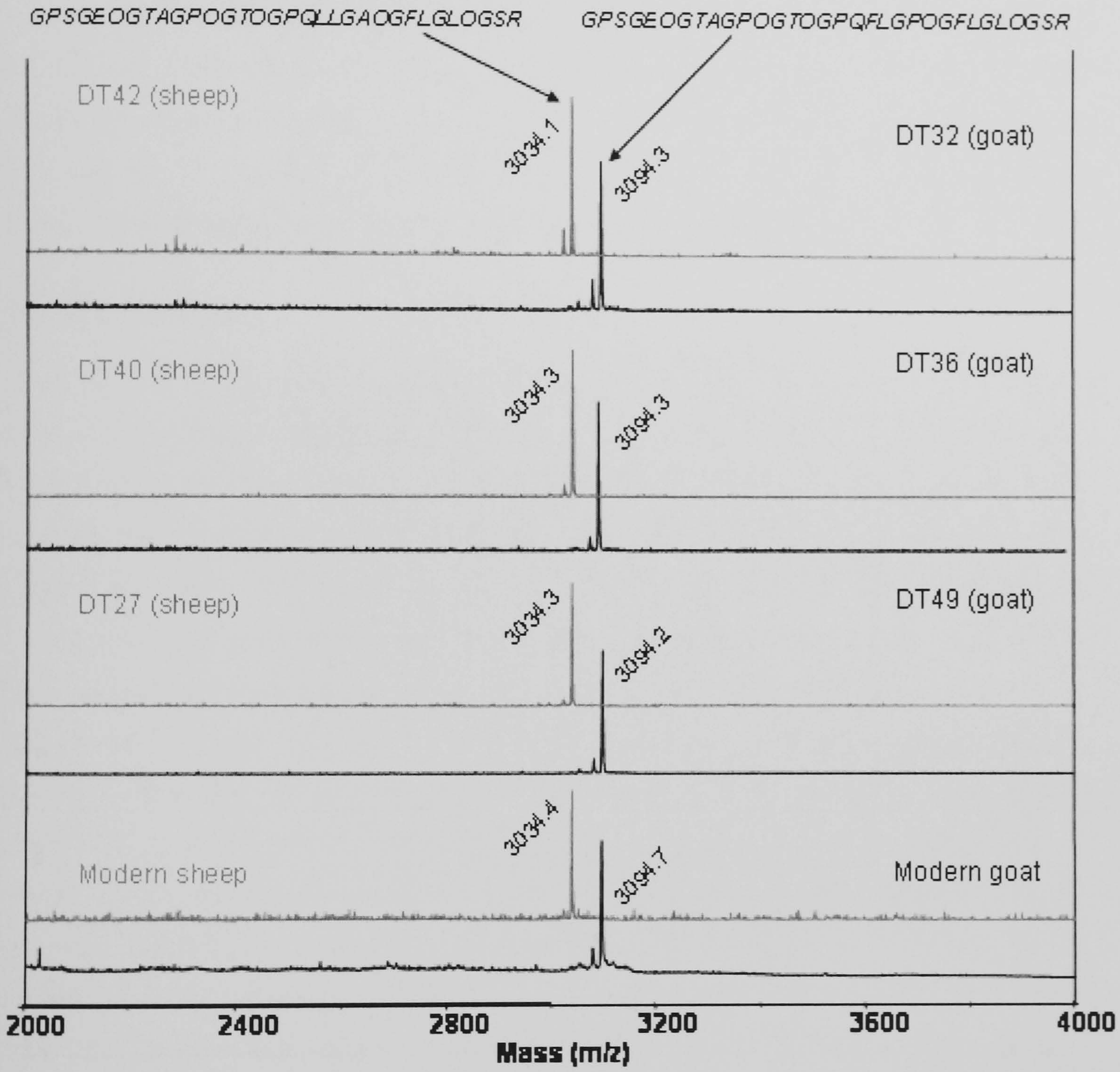


Figure 8.4 - Two modern standards and six examples of MALDI-MS reflectron mode mass spectra from Domuztepe sheep/goat samples, three archaeological samples are identified as sheep (shown in grey) and three are identified as goat (shown in black).

The blind collagen-peptide marker and morphological analyses on the 26 archaeological samples from Domuztepe matched in all but one of the specimens for which morphological criteria were present (23 of 24 specimens). The two morphologically undeterminable sheep/goat bones (both undeterminable because of their young age) were identified as goat using the described peptide markers. Six confidently identified sheep samples and one which

was ambiguous were all identified as sheep by the collagen-peptide markers. Likewise, the four weak goat assignments were all confirmed as goat by the collagen-peptide sequences. Of the 14 samples confidently identified as goat following various morphological criteria, 13 were confirmed by the collagen-peptide marker with only sample #52 returning conflicting identifications. This specimen, a distal radius fragment, provided only one morphological criterion (the shape of the distal articular surface of the radius) on which to base the visual distinction and its apparent misidentification reaffirms the need for identification using multiple morphological features rather than on any single criterion.

8.5 Discussion

Due to the loss of information regarding animal husbanding in the amalgamation of sheep and goats into a single category, demand for non-morphological methods of discriminating between sheep and goat bones had been tackled in a number of ways. These included sequential isotopic analysis of enamel (Balasse and Ambrose 2005), patterns of tooth wear (Wapnish and Hesse 1988) and DNA extraction from bone (Bar-Gal *et al.* 2003; Newman *et al.* 2002). Problems associated with the use of aDNA include (i) the degradation of DNA, which increases with temperature, detrimental to studies of bones from 'warm' sites, such as those of the Near East (Bar-Gal *et al.* 2003); (ii) contamination from husbanding practices above the archaeological deposits (Haile *et al.* 2007) as well as from laboratory reagents (Leonard *et al.* 2007); and (iii) the analytical costs, increased by poor success rates. In contrast, these problems associated with aDNA analyses are not as influential to the results of protein analyses. The proposed collagen-peptide marker method does not suffer such problems as, (i) the degradation of collagen occurs at much slower rates than DNA (Nielsen-Marsh 2002) and has been observed in lower palaeolithic and older samples (Armstrong *et al.* 1983); and (ii) because mass spectrometry is the direct measurement of present biomolecules (rather than the amplification of minute amounts) therefore not suffering from contamination problems to the same extent as ancient DNA. The use of protein sequences for phylogenetic information is somewhat more limited than DNA sequences, where resolution appears to be at the taxonomic level of genus for several protein sequences (see Chapter 1 and Chapter 2). This collagen-peptide marker method describes the isolation of only a small proportion of the collagen α2 (I) chain sequence (~3%), which is too small and conserved for meaningful phylogenetic analyses, but with which the distinction between

sheep (*Ovis sp.*) and goat (*Capra sp.*) can be clearly made. The major benefit to archaeological applications is that it represents an inexpensive, fast and efficient means of identifying species within small fragmented assemblages. In addition to these benefits, only small amounts of sample are required, as despite low collagen yields in many of the archaeological samples tested, only 40-50 mg of bone was needed for the analyses (whereas only ~10 mg is required for modern or well-preserved bone samples). Regarding the biomolecular preservation of the Domuztepe bone samples, of the 26 archaeological samples analysed in this study, the Asx D/L ranged from 0.13-0.21, and 22 specimens gave Asx D/L values >0.14, indicating that aDNA retrieval or osteocalcin isolation from these samples is unlikely (see Chapter 2).

In this pilot study we have focused on testing the described method against a set of morphological criteria using an archaeological assemblage with 19 of the 20 confident identifications made on morphological criteria supported by the described collagen-peptide marker. The one 'confidently' identified sample that was contradicted by the biochemical results was an identification based on only one criterion (the shape of the radial carpal facet on the distal articular surface of the radius, as described in Boessneck (1964:72-73)). As mentioned above, this error cautions against making identifications using only one criterion, in particular one on which considerable variation has been observed (see Boessneck 1964:73). The high level of correspondence between the morphological and chemical determinations for the rest of the samples, suggests that this may not be as reliable as a morphological determinant as features on other elements. In addition, the morphological criteria used to determine the four unconfident goat samples and one unconfident sheep sample were all supported by the collagen-peptide marker.

Despite the use of the key studies that present morphological distinctions between sheep and goat bones (Boessneck *et al.* 1964; Payne 1985), the reliability of an individual researcher's determination will largely be dependent on the preservation of the specimen, the age of the animal at death (with younger animals often lacking clear morphological markers) and the analyst's skill and confidence in making the determination. To avoid misrepresenting an assemblage, analysts should only make the determination between sheep and goat with

specimens exhibiting strong distinguishing features. In addition, certain morphological markers are more reliable than others, and ideally one should use a suite of markers on the same bone before making a final determination.

Of particular interest for the application of this method is the ability to objectively distinguish between sheep and goat bones belonging to immature animals overcoming problems regarding underdeveloped skeletal morphological criteria, especially on fragmented specimens. At the site of Domuztepe there has been speculation from preliminary analyses that the people may have preferred to kill goats at a very young age with more adult animals represented in the sheep population, possibly suggesting that they valued sheep for their wool. Since the 26 bones used in this study do not represent a random sample of elements from the 'ditch' context, the results cannot be applied to our current interpretations regarding the importance of sheep for wool at Domuztepe. However, it is of interest to note that the two specimens that were too young to morphologically determine the species both came from goats, which is consistent with the general observation that goats were killed younger than sheep. This method could be used to support conclusions about differential sheep and goat kill-off patterns because of the ability to reliably test immature ovicaprid bones and better understand culling practices involving immature individuals. Specifically, this method promises to be extremely informative in conjunction with age data gleaned from mandibular tooth eruption and wear, where it will provide incontrovertible distinction between sheep and goat mandibles from individuals of all ages. This will improve our understanding of the earlier sections of age curves, which are frequently obscured because of the increased bias of preservation against the less-mineralised young bones.

8.6 Conclusions

This pilot study has demonstrated an objective molecular method of distinguishing between degraded archaeological sheep and goat bones and has highlighted possible problems with one morphological criterion, the shape of the distal articular surface of the radius. In addition to testing the reliability of specific morphological criteria, the main applications of the method would be to identify morphologically undiagnostic bones. Many sites would find this method useful; either because material is too fragmented to distinguish between these

particular species or because there are immature animal remains. Using a set of 26 archaeological bones of varying degrees of morphological identity we have demonstrated that the method is accurate and 100% successful, even in samples of low collagen-yields. We have also proven that many of the widely used morphological criteria for distinguishing sheep and goat bones are reliable, though distinction should only be made with a high degree of confidence when more than one marker is present. There is clearly potential for the outlined method as a relatively easy, quick and low-cost way (analytical costs ~£5 per sample) of better understanding ovicaprid populations on any archaeological site. Although we have presented a detailed analysis of the sequence of each collagen-peptide marker for this pilot study, only the screening for the intact peptide is necessary and as a result large-scale analyses may become routine.

9 The survival of selected collagen α2 (I) chain species-specific peptide markers in rendered MBM

9.1 Abstract

The methods of isolating particular collagen a2 (I) chain peptides using SPE following two separate enzymatic digestions (one using bacterial collagenase, the other using trypsin) was carried out on cattle, sheep, pig and chicken 145°C-rendered MBM sediment samples and analysed by MALDI-MS. In addition, these collagen-peptide markers were analysed in cattle MBM sediment samples rendered at the range of temperatures 133°C, 137°C, 141°C, and 145°C in order to investigate the effects of increasing temperature on a sample of known collagen sequence. LC-MS analyses were carried out on the cattle, sheep, pig and chicken 145°C rendered MBM sediments using both LC-MALDI and LC-ESI and the data searched against a local 'Collagens' database that contains sequences for both pig and sheep (currently not available in public databases). The results indicated that using additional sequences of pig and sheep collagen (I), obtained from a combination of genomic data (aligned against the human genome for collagen sequence) and LC-MALDI and LC-ESI data, it was possible to identify the dominant species in each MBM sample. In order to standardise species identification approaches, nine peptides that were frequently observed throughout the LC-MS analyses of MBM sample digests were selected as potential species-specific markers. One mixed MBM sediment sample was also analysed by LC-ESI and the results indicated the ability to use these markers to identify a species present at the 15% level, but difficulty in identifying an additional species at the 5% level.

9.2 Introduction

Much of the research that has been carried out over the past decade in search of the most ideal methods of identifying the presence of species-specific markers in MBM has attempted to meet the ~0.1% (w:w) detection limits for contaminant MBM from one species in MBM targeted for possible consumption by another species (see section 1.2). Until such methods

have been proven successful (with a false positive rate of less than 5%), the ban on the use of all MBM in feedstuffs remains in place. Microscopy-based methods are widely accepted as sufficient to distinguish bone particles in MBM of marine from terrestrial animals (Baeten *et al.* 2001; Baeten *et al.* 2005; Perez-Marin *et al.* 2004), but in order to improve the taxonomic resolution of species identification in MBM, biomolecular methods have been investigated (Kim *et al.* 2004; Kim *et al.* 2005; Toyoda *et al.* 2004). Although there have been some reports of success with species identification at these levels (Baeten *et al.* 2004), they are all based on MBM rendered at ~133°C, which is the proposed minimum temperature required for sterilisation of MBM with intention to produce animal feed (Taylor 1998; Taylor and Woodgate 2003). However, to fully test the robustness of potential biomolecular methods several sets of standards were created for the STRATFEED project that range in rendering temperatures ~130-150°C (Garrido-Varo *et al.* 2005). Although some PCR and ELISA methods were shown to be successful at detecting contaminant species in the lower temperature rendered samples (~133°C) at the desired levels (~0.1%), none has yet been reported to be successful when applied to the higher temperature rendered samples (Baeten *et al.* 2004).

The research described in Chapter 3 has already shown that one potential target, the mineral-binding protein OC, was present intact in the 133°C-rendered MBM samples, but not in any of the higher temperature rendered (137°C, 141°C, or 145°C) MBM samples. Characterisation of the main proteinaceous constituents of MBM (Chapter 4) indicated that although several other proteins are present, collagen (I) is the dominant one. Two SPE-based methods of isolating particular species-specific collagen peptides by SPE and analysis by MALDI-MS (the carboxytelopeptide (following digestion with collagenase and isolation by SPE) and peptides from the helical region of collagen (following digestion with trypsin and isolation by SPE)) described in section 5.2 and section 7.3 proved to be capable of distinguishing between species of economic interest such as cattle, sheep, pig and chicken and 100% successful in the analysis of archaeological samples (Chapters 6 & 8). The aims of the research described in this chapter were to investigate the use of these SPE approaches (following either digestion with bacterial collagenase or digestion with trypsin) of isolating species-specific collagen-peptide markers on cattle, sheep, pig and chicken MBM samples rendered at the highest temperature (145°C) and sedimented and cattle MBM samples

rendered at the three lower temperatures (133°C, 137°C & 141°C) and sedimented. In addition, LC-MALDI and LC-ESI analyses were also carried out on cattle, sheep, pig and chicken MBM sediment samples without SPE. The problems with database searching are that peptides are only identified when the sequences already exist in the database. As a result, when sequences are not present, only homologous peptides present from other species are reported; those which maybe useful for species identification are not reported because there is no match in the database, e.g., in the analysis of collagen from the sheep MBM sample was matched to cattle collagen (see Table 4.1). Aligning collagen-specific genomic data for both sheep and pig against the human genome made it possible to piece together partial sequences of the collagen α1 (I) and α2 (I) chains derived from NCBi Expressed Sequence Tag information (Appendix 5.1). The potential to improve sequence coverage by using a locally produced 'Collagens' database containing pig and sheep collagen (I) sequences was investigated. Both LC-MALDI and LC-ESI were used in order to obtain improved protein sequence coverages. In order to standardise the LC-MS methods for species identification, species-specific peptides that were frequently observed in the analyses of MBM digests were identified. Because the LC-MS-based methods were the most successful at analysing collagen from the highest temperature rendered (145°C) MBM sediment samples, a mixed MBM sediment sample (created by VLA) of known species composition (although unknown until the analyses and interpretations were completed 'blind') was analysed only by LC-MS methods and not by the SPE-based methods.

9.3 Experimental

9.3.1 Preparation of 'Collagen' and Isolation of Collagen-Peptide Markers

The MBM sediment samples were supplied as a powder and defatted as described in section 4.3.1. Collagen α2 (I) chain carboxytelopeptide fragments and helix peptide markers were obtained following the methods described in section 5.2 and section 7.3 respectively.

9.3.2 LC-MALDI and LC-ESI Analyses of Bone and MBM Samples

LC-MALDI analyses were carried out as described in section 3.3.5. In addition to LC-MALDI analyses, LC-ESI analyses were carried out on cattle, sheep, pig and chicken bone (cattle in triplicate), cattle, sheep, pig and chicken MBM sediment samples (rendered at 145°C) and a mixed MBM sample as follows: 1 μL of the diluted sample digest (following sections 4.3.1 and 4.3.2) was applied to an Ultimate nanoLC (Dionex Ultimate capillary HPLC) system using a 0.1 mm x 50 mm PSDVB monolith column (Dionex) with a 1-50% (solvent B in A) continuous linear gradient over 7.5 min (solvent A = 2% ACN, 98% H_2O, 0.1% formic acid; solvent B = 100% ACN, 0.1% formic acid) at a flow rate of 1.2 μL/min (column temperature = 70°C) followed by a 5 min wash at 95% B and finally 11 min reconditioning in 1% B. For ESI analysis an Applied Biosystems/MDS Sciex QSTAR® API Pulsar i Hybrid LC-MS/MS system with a MicroIonSpray source (fitted with a 20 μm ID capillary). Positive ESI-MS & MS/MS spectra were acquired using Information Dependent Acquisition (IDA) with an ionspray voltage of 5200 V, and an m/z range 300 - 2000. IDA settings were 0.5 s for the acquisition of survey MS spectrum, 0.5 s for product ion spectra on the 1st and 2nd most abundant ions that meet the switch criteria (when ions m/z 300-2000 and a charge state 2 to 4 exceeds 20 counts, switch after 1 spectrum, excluding former target ions for 60 s), and a cycle time of 1.5 s with the collision energy (CE) calculated automatically from the IDA CE parameter table. Peptide MS and MS/MS data were obtained directly from IDA files using the vendor-provided Mascot script (version 1.6b21). The data were then submitted to Mascot and searched against the public UniProt and local 'Collagens' databases. Mascot's MudPIT scoring was employed and separate decoy databases were included in each search.

9.3.3 Database Search Parameters

The resulting product ion mass spectra were collectively converted to peak lists by Applied Biosystems 4000 Explorer (Applied Biosystems) with a S/N cut-off of 15, and searched against either the UniProt database (updated 31/01/07, 254609 sequences) or a local 'Collagens' database (114 partial protein sequences presented in Appendix 5.1) by Mascot version 2.2. Cleavage with the enzyme trypsin, with up to three possible missed cleavages, was chosen (except in the case of the analysis of digests from collagenase, in which 'no

enzyme' was selected). No fixed modifications were included, but the variable modifications of oxidation of methionine, amino-terminal pyro-glutamic acid, hydroxylation of proline and lysine, and the deamidation of glutamine and asparagine were allowed for. Mass tolerances of 0.5 Da were applied to both precursor and fragment ions. A significance threshold of 99% was employed and the presence of at least one top-ranking ('bold red' as determined by Mascot) peptide was required per protein match.

9.3.4 Compiling Unpublished Sequences and Creating the 'Collagens' Database

9.3.4.1 BLAST against Expressed Sequence Tags

Short sequences of cattle collagen α1(I) and α2 (I) chains were searched by BLAST searching against the NCBi Expressed Sequence Tag (EST) libraries for both pig (approx. 500,000 entries) and sheep (approx. 50,000 entries). Fragmentary sequences were then compiled using EST data alone (by M. Collins). Following this, the spectra obtained from LC-MS of pig and sheep bone collagen (I) were used to supplement sequences, where peptides confidently identified as cattle or dog collagen (I) were inserted into the EST-derived collagen sequence.

9.3.4.2 UCSC human genome alignments

The University of California Santa Cruz (UCSC) Bioinformatics group (B. Raney) carried out an alignment of the total genomic sequences from 28 species. The 28 genome sequences include two completed genome projects (human and mouse), with an estimated coverage of over 99% of the euchromatin, and an error rate of 1 in 100,000 (International Human Sequencing Consortium 2004). In addition there are 16 high quality draft sequences from different species based on whole-genome shotgun assemblies with five to eightfold coverage of the genome and ten lower quality draft sequences from different species based on whole-genome shotgun assemblies with two-fold or lower coverage. In theory, according to the Lander-Waterman statistic (Lander and Waterman 1988), assembly of the genome from two-fold coverage should include 87.5% of the bases in the genome, and assembly of the genome from a five-fold coverage should include 99.4%. Following genome alignment, the collagen

(I) sequences were obtained by comparison to the human genome collagen (I) loci (Appendix 5.1).

9.4 Results

9.4.1 Analysis of Collagen (I) Carboxytelopeptide Marker

The acid-insoluble 'collagen' residue from cattle 133°C, 137°C, 141°C and 145°C MBM sediment samples and sheep, pig and chicken 145°C MBM sediment samples were digested with collagenase and the α2 (I) chain carboxytelopeptide isolated using SPE and analysed as described in section 5.2. MALDI linear mode mass spectra for the four cattle MBM sediment samples are shown in Figure 9.1.

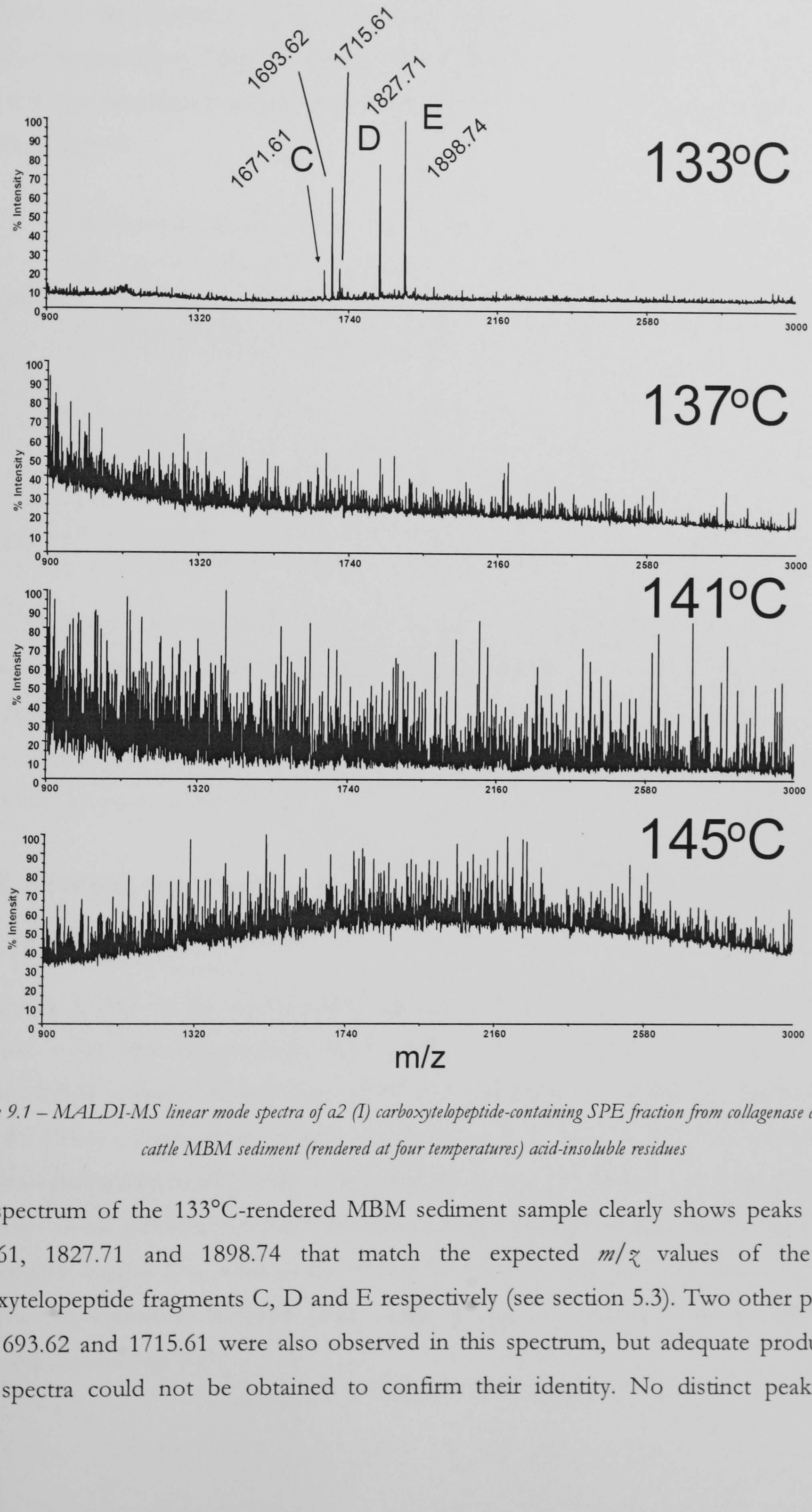


Figure 9.1 – MALDI-MS linear mode spectra of a2 (I) carboxytelopeptide-containing SPE fraction from collagenase digests of cattle MBM sediment (rendered at four temperatures) acid-insoluble residues

The spectrum of the 133°C-rendered MBM sediment sample clearly shows peaks at *m/z* 1671.61, 1827.71 and 1898.74 that match the expected *m/z* values of the cattle carboxytelopeptide fragments C, D and E respectively (see section 5.3). Two other peaks at *m/z* 1693.62 and 1715.61 were also observed in this spectrum, but adequate product ion mass spectra could not be obtained to confirm their identity. No distinct peaks were

observed in the spectra of any of the MBM sediment samples produced by higher temperature treatments. Table 9.1 lists the m/z and S/N values of the peaks observed in the four different temperature cattle MBM samples as well as the pig, sheep and chicken MBM samples analysed.

Table 9.1 - The m/z and S/N values (in parentheses) of observed telopeptide fragments (A-E) in MBM sediments of cattle 133-145°C, pig 145°C, sheep 145°C and chicken 145°C samples. X indicates no peak was observed.

Species	MBM Rendering Temperature (°C)	Fragment A	Fragment B	Fragment C	Fragment D	Fragment E.
Cattle	133	X	X	1671.6 (26.29)	1827.7 (70.72)	1898.7 (97.34)
Cattle	137	X	X	X	X	X
Cattle	141	X	X	X	X	X
Cattle	145	X	X	X	X	X
Pig	145	X	X	X	X	X
Sheep	145	X	X	X	X	X
Chicken	145	X	X	X	X	X

Table 9.1 shows that the only observable peaks were from the spectrum of the lowest temperature rendered MBM sediment sample (133°C) from cattle and these were predominantly for fragments C, D and E of the carboxytelopeptide. The lack of observable peaks in the highest temperature rendered samples (145°C) is consistent throughout the four animal species (spectra not shown).

9.4.2 Analysis of Peptide Markers from the Helical Region of the Collagen (I) Molecule

A separate aliquot of the acid-insoluble residue from the cattle MBM sediment samples rendered at the four temperatures 133°C, 137°C, 141°C and 145°C, and pig, sheep and chicken MBM sediment rendered at 145°C was gelatinised in ammonium bicarbonate at 65°C for 3 hours, digested with trypsin and particular α2 (I) chain 'helix' peptides isolated by SPE and analysed by MALDI-MS as described in section 7.3. Because only the 26% ACN fractions contain peptides potentially capable of distinguishing between cattle, sheep, pig and chicken (see section 7.4), whereas the 32% ACN fraction only yields Peptide C, a peptide that can not distinguish between cattle, sheep or pig, only spectra from this 26% ACN fraction is shown (Fig. 9.2).

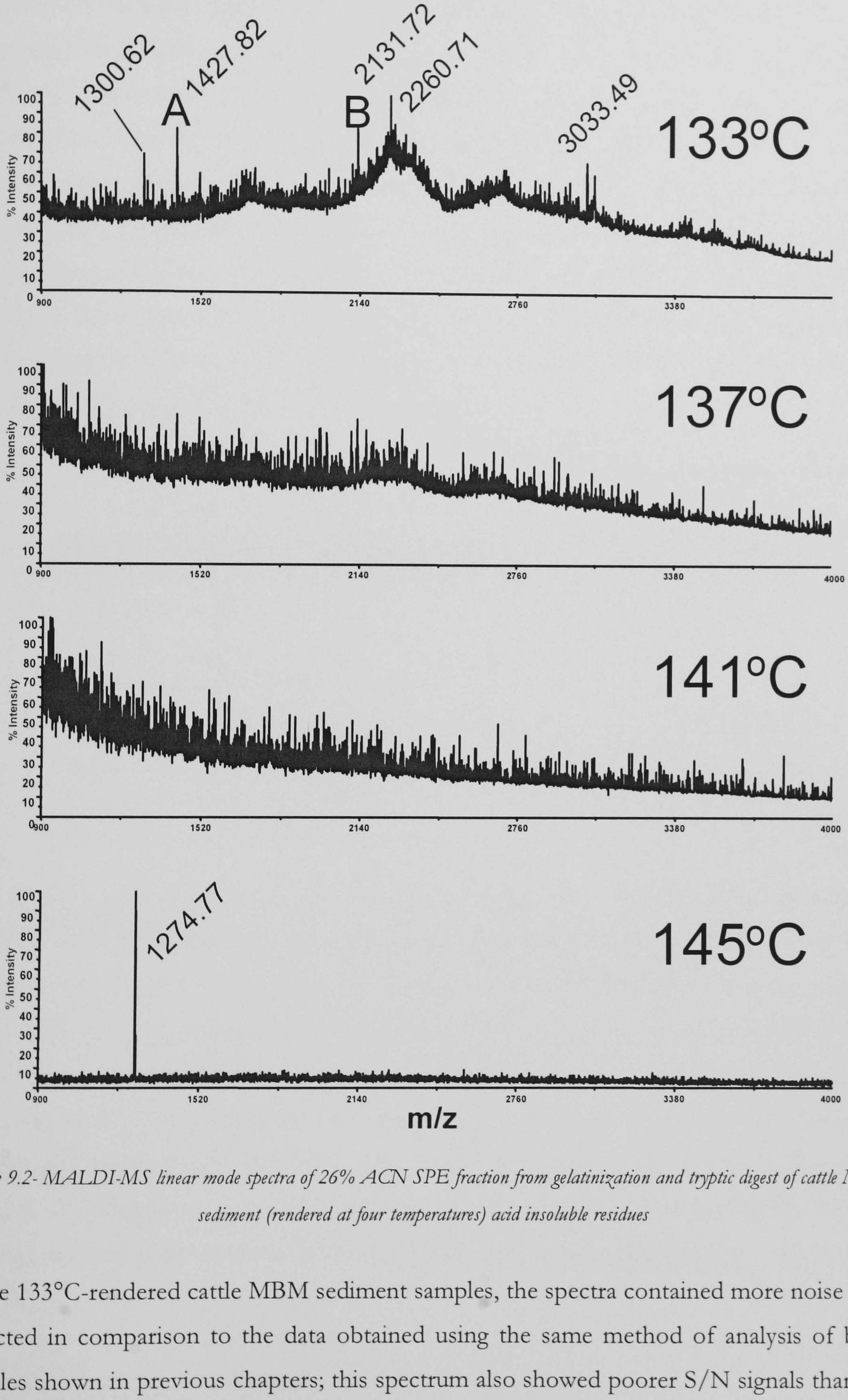


Figure 9.2- MALDI-MS linear mode spectra of 26% ACN SPE fraction from gelatinization and tryptic digest of cattle MBM sediment (rendered at four temperatures) acid insoluble residues

In the 133°C-rendered cattle MBM sediment samples, the spectra contained more noise than expected in comparison to the data obtained using the same method of analysis of bone samples shown in previous chapters; this spectrum also showed poorer S/N signals than the

spectrum for the same MBM sample analysed for carboxytelopeptide peaks (compare Fig. 9.2 to Fig. 9.1). Of the four peptides proposed as species-specific markers only Peptides A and B were observed as peaks at *m/z* 1427.8 and 2132.1 (Fig. 9.2). Three other peaks at *m/z* 1300.6, 2260.7 and 3033.5 were also observed but adequate product ion mass spectra could not be obtained to identify their identity. No distinct peaks were observed in the spectra of the two higher temperature rendered cattle MBM sediment samples (137°C & 141°C), and the spectrum of the highest temperature rendered (145°C) cattle MBM sediment sample had much less noise than the spectra of the three lower temperature rendered samples, and shows a single peak at *m/z* 1274.8. The *m/z* of the observed peaks are presented with their respective S/N values in Table 9.2.

Table 9.2 - The m/z and S/N values (in parentheses) of observed helix peptides (26%A-D) in MBM sediments of cattle 133-145°C, pig 145°C, sheep 145°C and chicken 145°C. X indicates no peak was observed.

Species	Temp. (°C)	*m/z* 26%A	*m/z* 26%B	*m/z* 26%C	*m/z* 26%D
Cattle	133	1427.8 (80.17)	2132.1 (78.77)	x	x
Cattle	137	X	X	x	x
Cattle	141	X	X	x	x
Cattle	145	1274.8 (84.95)	X	x	x
Pig	145	1274.8 (88.89)	X	x	x
Sheep	145	1274.8 (26.24)	X	x	x
Chicken	145	1288.8 (7.93)	X	x	x

Two of the expected peptide markers for cattle collagen were observed in the spectrum of the 26% ACN SPE fraction of the 133°C-rendered MBM sediment sample (Peptides A and B at *m/z* 1427.8 and 2132.1 respectively). Throughout the 26% ACN SPE fractions from the cattle MBM samples rendered at temperatures >133°C, only the spectrum from the highest temperature rendered MBM sediment sample possessed one observable peak at *m/z* 1274.8. This peak at *m/z* 1274.8 was also observed in the spectra for the other two mammalian samples and a peak at *m/z* 1288.8 was observed in the spectra for the chicken MBM sample. MALDI-MS analysis of the 137°C-rendered and the 141°C-rendered MBM sediment samples did not yield observable peaks, possibly due to the high levels of noise in the spectra.

In order to investigate this sole peptide dominating the spectra, product ion spectra were obtained (Fig. 9.3) for the precursor ion of m/z 1274.8 and searched against the UniProt protein database. When searched against UniProt ('no enzyme' was selected) using Mascot, the search resulted with the match to the sequence LLVVYPWTQR from the protein haemoglobin (subunit beta-1) and found in over 100 mammalian sepecies (in all haemoglobin subunit beta-1 sequences of mammal species currently in the UniProt database). Because the ion scores for each search were close to the ion score threshold, the identifications were supported with *de novo* sequencing (Fig. 9.3). Although most published mammalian sequences had the same sequence, the published avian sequences differed by only one amino acid; the mutation of valine to isoleucine at position 3 (LLIVYPWTQR). This matches the difference of +14 Da in the observed m/z of this peptide peak between the mammalian and avian samples analysed.

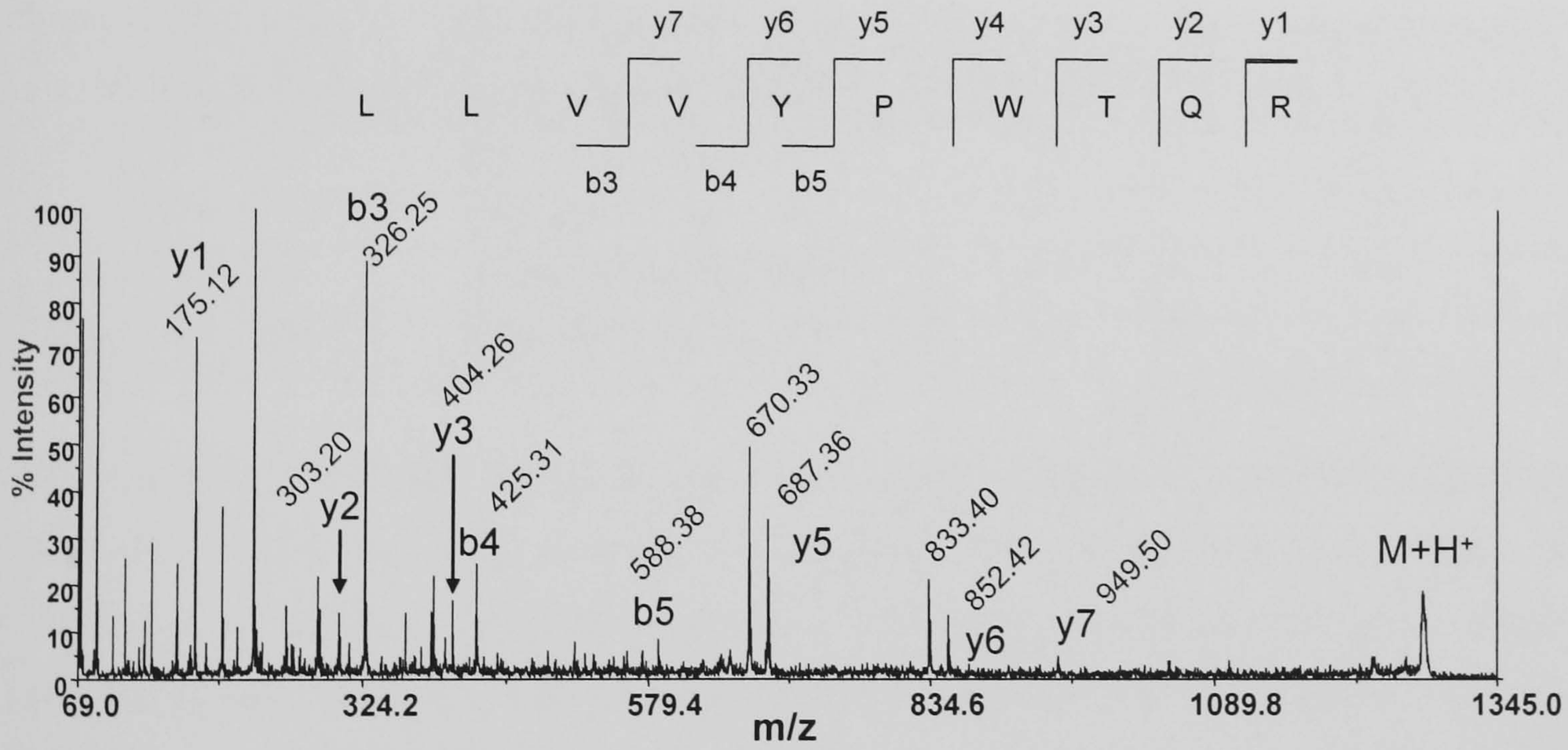


Figure 9.3 – MALDI-MS/MS spectrum of the only peptide observed in the cattle MBM sediment 145°C 'collagen' residue following gelatinisation, digestion with trypsin and isolation by SPE.

9.4.3 LC-MS Analysis of Sheep and Pig MBM Sediments

9.4.3.1 LC-MS of Pig and Sheep MBM Samples against UniProt Database

Because of the poor SPE-MS results from the analysis of target species-specific collagen peptides in the higher (>133°C) rendered sedimented MBM, LC-MS approaches on species of previously unknown collagen sequences were further investigated. LC-MS data was

obtained by LC-MALDI analyses for pig MBM sediment and sheep MBM sediment samples (rendered at 145°C) following gelatinisation and digestion with trypsin as described in section 4.3.2. LC-ESI analyses for both pig and sheep MBM sediment samples as well as pig and sheep bone were also carried out (LC-MALDI search results for pig and sheep bone samples are presented in Chapter 4). The data from each LC-MS analysis was searched against the UniProt database using Mascot (Table 9.3)(see Appendix 5.2 for data files and Appendix 5.3 for search results) and the sequence coverage maps investigated (Figs 9.4 & 9.5).

Table 9.3 - Collagen (I) sequence coverage (procollagen sequences not included), top matches, number of other matched peptides (peptide matches to other sequences), and number of unidentified peptides from LC-MALDI and LC-ESI Mascot searches against UniProt.

Sample + Analysis Method	Species Match + Sequence Coverage (α1/α2)	No. 'NCP' peptides	No. Unidentified peptides
Pig MBM -MALDI	Human (22%)/Cattle (12%)	15	99
Sheep MBM - MALDI	Cattle (46%)/Cattle (33%)	43	271
Pig MBM - ESI	Cattle (35%)/Dog (14%)	41	367
Sheep MBM - ESI	Cattle (35%)/Cattle (24%)	86	122

When analysed by LC-MS, the α1 (I) chain, (two copies of which are present in each triple helix,) had a higher sequence coverage than the α2 (I) chain, (which only occurs once) . In the analysis of the pig 145°C MBM sediment, LC-ESI search results matched more collagen (I) peptides than in the LC-MALDI analysis, but the opposite was observed for the sheep MBM sediment sample, in which the LC-MALDI coverages were greater than in the LC-ESI analyses; overall the sequence coverages were similar for both LC-MALDI and LC-ESI (see Table 9.3). The LC-MALDI and LC-ESI analyses of sheep MBM sediment samples resulted in top matches to cattle α1 (I) and α2 (I) collagen sequences. The LC-MALDI analysis of pig MBM sediment resulted in top matches to human (α1) and dog (α2) collagen (I) sequences and the LC-ESI analysis of pig MBM sediment resulted in matches to cattle (α1) and dog (α2) collagen (I) sequences, presumably due to the greater similarity between the actual pig collagen (I) sequence and the known dog collagen (I) sequence than to cattle collagen (I) sequences (at least for the α2 (I) chain sequences).

Pig sample human α1 (I) coverage

```
  1 MFSFVDLRLL LLLAATALLT HGQEEGQVEG QDEDIPPITC VQNGLRYHDR
 51 DVWKPEPCRI CVCDNGKVLC DDVICDETKN CPGAEVPEGE CCPVCPDGSE
101 SPTDQETTGV EGPKGDTGPR GPRGPAGPPG RDGIPGQPGL PGPPGPPGPP
151 GPPGLGGNFA PQLSYGYDEK STGGISVPGP MGPSGPRGLP GPPGAPGPQG
201 FQGPPGEPGE PGASGPMGPR GPPGPPGKNG DDGEAGKPGR PGERGPPGPQ
251 GARGLPGTAG LPGMKGHRGF SGLDGAKGDA GPAGPKGEPG SPGENGAPGQ
301 MGPRGLPGER GRPGAPGPAG ARGNDGATGA AGPPGPTGPA GPPGFPGAVG
351 AKGEAGPQGP RGSEGPQGVR GEPGPPGPAG AAGPAGNPGA DGQPGAKGAN
401 GAPGIAGAPG FPGARGPSGP QGPGGPPGPK GNSGEPGAPG SKGDTGAKGE
451 PGPVGVQGPP GPAGEEGKRG ARGEPGPTGL PGPPGERGGP GSRGFPGADG
501 VAGPKGPAGE RGSPGPAGPK GSPGEAGRPG EAGLPGAKGL TGSPGSPGPD
551 GKTGPPGPAG QDGRPGPPGP PGARGQAGVM GFPGPKGAAG EPGKAGERGV
601 PGPPGAVGPA GKDGEAGAQG PPGPAGPAGE RGEQGPAGSP GFQGLPGPAG
651 PPGEAGKPGE QGVPGDLGAP GPSGARGERG FPGERGVQGP PGPAGPRGAN
701 GAPGNDGAKG DAGAPGAPGS QGAPGLQGMP GERGAAGLPG PKGDRGDAGP
751 KGADGSPGKD GVRGLTGPIG PPGPAGAPGD KGESGPSGPA GPTGARGAPG
801 DRGEPGPPGP AGFAGPPGAD GQPGAKGEPG DAGAKGDAGP PGPAGPAGPP
851 GPIGNVGAPG AKGARGSAGP PGATGFPGAA GRVGPPGPSG NAGPPGPPGP
901 AGKEGGKGPR GETGPAGRPG EVGPPGPPGP AGEKGSPGAD GPAGAPGTPG
951 PQGIAGQRGV VGLPGQRGER GFPGLPGPSG EPGKQGPSGA SGERGPPGPM
1001 GPPGLAGPPG ESGREGAPGA EGSPGRDGSP GAKGDRGETG PAGPPGAPGA
1051 PGAPGPVGPA GKSGDRGETG PAGPAGPVGP VGARGPAGPQ GPRGDKGETG
1101 EQGDRGIKGH RGFSGLQGPP GPPGSPGEQG PSGASGPAGP RGPPGSAGAP
1151 GKDGLNGLPG PIGPPGPRGR TGDAGPVGPP GPPGPPGPPG PPSAGFDFSF
1201 LPQPPQEKAH DGGRYYRADD ANVVRDRDLE VDTTLKSLSQ QIENIRSPEG
1251 SRKNPARTCR DLKMCHSDWK SGEYWIDPNQ GCNLDAIKVF CNMETGETCV
1301 YPTQPSVAQK NWYISKNPKD KRHVWFGESM TDGFQFEYGG QGSDPADVAI
1351 QLTFLRLMST EASQNITYHC KNSVAYMDQQ TGNLKKALLL QGSNEIEIRA
1401 EGNSRFTYSV TVDGCTSHTG AWGKTVIEYK TTKTSRLPII DVAPLDVGAP
1451 DQEFGFDVGP VCFL
```

Pig sample cattle α2 (I) coverage

```
  1 MLSFVDTRTL LLLAVTSCLA TCQSLQEATA RKGPSGDRGP RGERGPPGPP
 51 GRDGDDGIPG PPGPPGPPGP PGLGGNFAAQ FDAKGGGPGP MGLMGPRGPP
101 GASGAPGPQG FQGPPGEPGE PGQTGPAGAR GPPGPPGKAG EDGHPGKPGR
151 PGERGVVGPQ GARGFPGTPG LPGFKGIRGH NGLDGLKGQP GAPGVKGEPG
201 APGENGTPGQ TGARGLPGER GRVGAPGPAG ARGSDGSVGP VGPAGPIGSA
251 GPPGFPGAPG PKGELGPVGN PGPAGPAGPR GEVGLPGLSG PVGPPGNPGA
301 NGLPGAKGAA GLPGVAGAPG LPGPRGIPGP VGAAGATGAR GLVGEPGPAG
351 SKGESGNKGE PGAVGQPGPP GPSGEEGKRG STGEIGPAGP PGPPGLRGNP
401 GSRGLPGADG RAGVMGPAGS RGATGPAGVR GPNGDSGRPG EPGLMGPRGF
451 PGSPGNIGPA GKEGPVGLPG IDGRPGPIGP AGARGEPGNI GFPGPKGPSG
501 DPGKAGEKGH AGLAGARGAP GPDGNNGAQG PPGLQGVQGG KGEQGPAGPP
551 GFQGLPGPAG TAGEAGKPGE RGIPGEFGLP GPAGARGERG PPGESGAAGP
601 TGPIGSRGPS GPPGPDGNKG EPGVVGAPGT AGPSGPSGLP GERGAAGIPG
651 GKGEKGETGL RGDIGSPGRD GARGAPGAIG APGPAGANGD RGEAGPAGPA
701 GPAGPRGSPG ERGEVGPAGP NGFAGPAGAA GQPGAKGERG TKGPKGENGP
751 VGPTGPVGAA GPSGPNGPPG PAGSRGDGGP PGATGFPGAA GRTGPPGPSG
801 ISGPPGPPGP AGKEGLRGPR GDQGPVGRSG ETGASGPPGF VGEKGPSGEP
851 GTAGPPGTPG PQGLLGAPGF LGLPGSRGER GLPGVAGSVG EPGPLGIAGP
901 PGARGPPGNV GNPGVNGAPG EAGRDGNPGN DGPPGRDGQP GHKGERGYPG
951 NAGPVGAAGA PGPQGPVGPV GKHGNRGEPG PAGAVGPAGA VGPRGPSGPQ
1001 GIRGDKGEPG DKGPRGLPGL KGHNGLQGLP GLAGHHGDQG APGAVGPAGP
1051 RGPAGPSGPA GKDGRIGQPG AVGPAGIRGS QGSQGPAGPP GPPGPPGPPG
1101 PSGGGYEFGF DGDFYRADQP RSPTSLRPKD YEVDATLKSL NNQIETLLTP
1151 EGSRKNPART CRDLRLSHPE WSSGYYWIDP NQGCTMDAIK VYCDFSTGET
1201 CIRAQPEDIP VKNWYRNSKA KKHVWVGETI NGGTQFEYNV EGVTTKEMAT
1251 QLAFMRLLAN HASQNITYHC KNSIAYMDEE TGNLKKAVIL QGSNDVELVA
1301 EGNSRFTYTV LVDGCSKKTN EWQKTIIEYK TNKPSRLPIL DIAPLDIGGA
1351 DQEIRLNIGP VCFK
```

Pig sample cattle α1 (I) coverage

```
  1 MFSFVDLRLL LLLAATALLT HGQEEGQEEG QEEDIPPVTC VQNGLRYHDR
 51 DVWKPVPCQI CVCDNGNVLC DDVICDELKD CPNAKVPTDE CCPVCPEGQE
101 SPTDQETTGV EGPKGDTGPR GPRGPAGPPG RDGIPGQPGL PGPPGPPGPP
151 GPPGLGGNFA PQLSYGYDEK STGISVPGPM GPSGPRGLPG PPGAPGPQGF
201 QGPPGEPGEP GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG
251 ARGLPGTAGL PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM
301 GPRGLPGERG RPGAPGPAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA
351 KGEGGPQGPR GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG
401 APGIAGAPGF PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP
451 GPTGIQGPPG PAGEEGKRGA RGEPGPAGLP GPPGERGGPG SRGFPGADGV
501 AGPKGPAGER GAPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG
551 KTGPPGPAGQ DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP
601 GPPGAVGPAG KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP
651 PGEAGKPGEQ GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG
701 APGNDGAKGD AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK
751 GADGAPGKDG VRGLTGPIGP PGPAGAPGDK GEAGPSGPAG PTGARGAPGD
801 RGEPGPPGPA GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPAGPPG
851 PIGNVGAPGP KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA
901 GKEGSKGPRG ETGPAGRPGE VGPPGPPGPA GEKGAPGADG PAGAPGTPGP
951 QGIAGQRGVV GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG
1001 PPGLAGPPGE SGREGAPGAE GSPGRDGSPG AKGDRGETGP AGPPGAPGAP
1051 GAPGPVGPAG KSGDRGETGP AGPAGPIGPV GARGPAGPQG PRGDKGETGE
1101 QGDRGIKGHR GFSGLQGPPG PPGSPGEQGP SGASGPAGPR GPPGSAGSPG
1151 KDGLNGLPGP IGPPGPRGRT GDAGPAGPPG PPGPPGPPGP PSGGYDLSFL
1201 PQPPQEKAHD GGRYYRADDA NVVRDRDLEV DTTLKSLSQQ IENIRSPEGS
1251 RKNPARTCRD LKMCHSDWKS GEYWIDPNQG CNLDAIKVFC NMETGETCVY
1301 PTQPSVAQKN WYISKNPKEK RHVWYGESMT GGFQFEYGGQ GSDPADVAIQ
1351 LTFLRLMSTE ASQNITYHCK NSVAYMDQQT GNLKKALLLQ GSNEIEIRAE
1401 GNSRFTYSVT YDGCTSHTGA WGKTVIEYKT TKTSRLPIID VAPLDVGAPD
1451 QEFGFDVGPA CFL
```

Pig sample dog α2 (I) coverage

```
  1 MLSFVDTRTL LLLAVTSCLA TCQSLQEATA RKGPTGDRGP RGERGPPGPP
 51 GRDGDDGIPG PPGPPGPPGP PGLGGNFAAQ YDGKGVGLGP GPMGLMGPRG
101 PPGASGAPGP QGFQGPAGEP GEPGQTGPAG ARGPPGPPGK AGEDGHPGKP
151 GRPGERGVVG PQGARGFPGT PGLPGFKGIR GHNGLDGLKG QPGAPGVKGE
201 PGAPGENGTP GQTGARGLPG ERGRVGAPGP AGARGSDGSV GPVGPAGPIG
251 SAGPPGFPGA PGPKGEIGPV GNPGPAGPAG PRGEVGLPGV SGPVGPPGNP
301 GANGLTGAKG AAGLPGVAGA PGLPGPRGIP GPVGAAGATG ARGIVGEPGP
351 AGSKGESGNK GEPGSAGAQG PPGPSGEEGK RGPNGEAGSA GPSGPPGLRG
401 SPGSRGLPGA DGPAGVMGPP GPRGATGPAG VRGPNGDSGR PGEPGLMGPR
451 GFPGAPGNVG PAGKEGPMGL PGIDGRPGPI GPAGARGEPG NIGFPGPKGP
501 TGDPGKNGDK GHAGLAGARG APGPDGNNGA QGPPGPQGVQ GGKGEQGPAG
551 PPGFQGLPGP AGTAGEVGKP GERGLPGEFG LPGPAGPRGE RGPPGESGAA
601 GPSGPIGSRG PSGPPGPDGN KGEPGVLGAP GTAGASGPGG LPGERGAAGI
651 PGGKGEKGET GLRGEIGNPG RDGARGAPGA MGAPGPAGAT GDRGEAGPAG
701 PAGPAGPRGT PGERGEVGPA GPNGFAGPAG AAGQPGAKGE RGTKGPKGEN
751 GPVGPTGPIG SAGPSGPNGP PGPAGSRGDG GPPGATGFPG AAGRTGPPGP
801 SGITGPPGPP GAAGKEGLRG PRGDQGPVGR TGETGASGPP GFTGEKGPSG
851 EPGTAGPPGT PGPQGLLGAP GILGLPGSRG ERGLPGVAGS VGEPGPLGIA
901 GPPGARGPPG AVGAPGVNGA PGEAGRDGNP GNDGPPGRDG QAGHKGERGY
951 PGNIGPVGAV GAPGPHGPVG PTGKHGNRGE PGPAGSVGPV GAVGPRGPSG
1001 PQGIRGDKGE PGEKGPRGLP GLKGHNGLQG LPGLAGQHGD QGAPGSVGPA
1051 GPRGPAGPSG PAGKDGRTGQ PGTVGPAGIR GSQGSQGPAG PPGPPGPPGP
1101 PGPSGGGYDF GYEGDFYRAD QPRSPPSLRP KDYEVDATLK SLNNQIETLL
1151 TPEGSRKNPA RTCRDLRLSH PEWSSGYYWI DPNQGCTMDA IKVYCDFSTG
1201 ETCIRAQPEN IPAKNWYRNS KVKKHIWLGE TINGGTQFEY NVEGVTTKEM
1251 ATQLAFMRLL ANHASQNITY HCKNSIAYMD EETGNLKKAV ILQGSNDVEL
1301 VAEGNSRFTY TVLVDGCSKK TNEWRKTIIE YKTNKPSRLP ILDIAPLDIG
1351 DADQEFRVDV GPVCFK
```

Figure 9.4 – Sequence coverage map to pig a1 (I) and a2 (I) chain collagen matches from pig MBM sediment rendered at 145°C and analysed by LC-MALDI and LC-ESI. Grey shading indicates procollagen sequence that is not part of the tropocollagen molecule, red text indicates peptide matches in LC-MALDI analyses.

In the pig α1 collagen (I) matches (to human collagen in the LC-MALDI analysis and cattle collagen in the LC-ESI analyses), the LC-MALDI search results matched four unique peptides (that were not matched in the LC-ESI search results) and the LC-ESI search results matched seven unique peptides. In the pig α2 (I) collagen matches (to dog collagen in both LC-MALDI and LC-ESI analyses), only two unique peptides were matched with LC-MALDI and four with LC-ESI.

Figure 9.5 – Sequence coverage map to sheep a1 (I) and a2 (I) chain collagen matches from sheep MBM sediment rendered at 145°C and analysed by LC-MALDI and LC-ESI. Grey shading indicates procollagen sequence that is not part of the tropocollagen molecule, red text indicates peptide matches in LC-MALDI analyses.

In the sheep α1 (I) collagen matches (to cattle collagen in both LC-MALDI and LC-ESI analyses), the LC-MALDI search results matched seven unique peptides and the LC-ESI search results matched eight unique peptides. In the sheep α2 (I) collagen matches (to cattle collagen using LC-MALDI and to dog collagen using LC-ESI), nine unique peptides were matched with LC-MALDI and eight were matched with LC-ESI. This is indicative of the complementarity of LC-MALDI and LC-ESI analyses in the study of collagen digests, which has been proposed in numerous other proteomic studies (Bodnar *et al.* 2003; Krutchinsky *et al.* 2000). However, the match to an α2 (I) chain procollagen peptide in the pig MBM sediment analysed by LC-ESI (see Fig. 9.4) is unexpected because following cleavage and secretion into the extracellular matrix these procollagen peptides are further degraded by

non-specific proteases into smaller peptides. An alternative explanation would be that this is a misidentification of a similar collagen peptide present within the triple helical region, but not present in the search database.

9.4.3.2 LC-MS Searches of Pig and Sheep Samples against 'Collagens' Database

The α1 (I) and α2 (I) chain sequences for both pig and sheep collagen were estimated from a combination of EST searches and results from previous LC-MALDI analyses of pig and sheep bone 'collagen' samples. These sequences were then compiled (along with numerous other collagen sequences derived from alignments with genomic sequence of other species) into a database called 'Collagens' (Appendix 5.1). This locally produced and held database of collagen sequences was then searched using the LC-MS data from pig and sheep MBM sediment samples (Table 9.4) and the resulting sequence coverage maps of the highest scoring collagen matches presented (Figs 9.6 & 9.7).

Table 9.4 - Summary of LC-MALDI and LC-ESI results of MBM sediment gelatine digested with trypsin and searched against 'Collagens' including collagen (I) sequence coverage, top matches, number of other matched peptides (bold red hits to other collagens and NCPs), and number of unidentified peptides.

Sample + Analysis Method	Species Match + Sequence Coverage (α1/α2)	No. NCP peptides	No. Unidentified peptides
Pig MBM - MALDI	Pig (40%)/Pig(28%)	28	26
Sheep MBM - MALDI	Sheep(75%)/Sheep(70%)	76	116
Pig MBM - ESI	Pig(45%)/Pig(37%)	41	110
Sheep MBM - ESI	Sheep(42%)/Sheep(29%)	47	264

Although slightly more of the α1 (I) chain was matched than the α2 (I) chain in each sample searched against the 'Collagens' database, the difference in sequence coverage between these two chains is not as great as was observed in the LC-MS searches against UniProt, particularly with the analyses of the pig MBM sediment samples. This indicates that the much greater sequence coverage for the α1 (I) sequence than the α2 (I) chain sequence using both LC-MALDI and LC-ESI (resulting from the searches against UniProt; Table 9.3) may not be due to the presence of two α1 (I) chain for every α2 (I) chain, but probably due to the

lack of any similar sequence (of a closely related species) in the UniProt database; the sequence coverage of the α1 (I) chain is less affected by the limited available sequences because it is much more conserved than the α2 (I) chain (see section 1.6.5). Similar to the UniProt search results of the pig MBM sediment sample analyses against the 'Collagens' database, the LC-ESI search results matched more collagen (I) peptides than in the LC-MALDI analysis; the opposite was again observed for the sheep MBM sediment sample, in which the LC-MALDI sequence coverages of collagen were greater than in the LC-ESI analyses; overall the sequence coverages of collagen were similar for both LC-MALDI and LC-ESI (see Table 9.4).

MALDI

Pig sample pig α1 (I) coverage

```
  1 QLSYGYDEKS AGKISVPGPM GPSGPRGLPG PPGAPGPQGF QGPPGEPGEP
 51 GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG ARGLPGTAGL
101 PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM GPRGLPGERG
151 RPGPPGTAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA KGEAGPQGAR
201 GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG APGIAGAPGF
251 PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP GPTGIQGPPG
301 PAGEEGKRGA RGEPGPSGLP GPPGERGGPG SRGFPGADGV AGPKGPAGER
351 GSPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG KTGPPGPAGQ
401 DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP GPPGAVGPAG
451 KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP PGEAGKPGEQ
501 GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG APGNDGAKGD
551 AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK GADGAPGKDG
601 VRGLTGPIGP PGPAGAPGDK GETGPSGPAG PTGARGAPGD RGEPGPPGPA
651 GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPTGPPG PIGSVGAPGP
701 KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA GKEGSKGPRG
751 ETGPAGRPGE AGPPGPPGPA GEKGSPGADG PAGAPGTPGP QGIAGQRGVV
801 GLPGQRGERG FPGLPGPSGE PGKQGPSGPS GERGPPGPMG PPGLAGPPGE
851 SGREGAPGAE GSPGRDGAPG PKGDRGESGP AGPPGAPGAP GAPGPVGPAG
901 KSGDRGETGP AGPAGPVGPV GARGPAGPQG PRGDKGETGE QGDRGIKGHR
951 GFSGLQGPPG PPGAPGEQGP SGASGPAGPR GPPGSAGSPG KDGLNGLPGP
1001 IGPPGPRGRT GDAGPVGPPG PPGPPGPPGP PSGGFDFSFL PQPPQEKAHD
1051 GGRYYRA
```

Pig sample pig α2 (I) coverage

```
  1 QYDGKGVGAG PGPMGLMGPR GPPGAVGAPG PQGFQGPAGE PGEPGQTGPA
 51 GARGPPGPPG KAGEDGHPGK PGRPGERGVV GPQGARGFPG TPGLPGFKGI
101 RGHNGLDGLK GQPGAPGVKG EPGAPGENGT PGQTGARGLP GERGRVGAPG
151 PAGARGNDGS VGPVGPAGPI GSAGPPGFPG APGPKGELGP VGNPGPAGPA
201 GPRGEVGLPG VSGPVGPPGN PGANGLPGAK GAAGLPGVAG APGLPGPRGI
251 PGPAGAAGAT GARGLVGEPG PAGSKGESGN KGEPGAAGPQ GPPGPSGEEG
301 KRGPNGEVGS AGPPGPPGLR GNPGSRGLPG ADGRAGVMGP PGSRGPTGPA
351 GVRGPNGDSG RPGEPGLMGP RGFPGSPGNV GPAGKEGPAG LPGIDGRPGP
401 IGPAGARGEP GNIGFPGPKG PTGDPGKNGE KGHAGLAGAR GAPGPDGNNG
451 AQGPPGPQGV QGGKGEQGPA GPPGFQGLPG PAGTAGEVGK PGERGIPGEF
501 GLPGLAGPRG ERGPPGESGA AGPAGLIGSR GPSGPPGLDG NKGELGVLGA
551 PGTAGPSGLS GLPGERGAAG IPGGKGEKGE TGLRGDVGSP GRDGARGAPG
601 AVGAPGPAGA NGDRGEAGPA GPAGPAGPRG SPGERGEVGP AGPNGFAGPA
651 GAAGQPGAKG ERGTKGPKGE NGPVGPTGPV GAAGPAGPNG PPGPAGSRGD
701 GGPPGATGFP GAAGRIGPPG PSGISGPPGP PGPAGKEGLR GPRGDQGPVG
751 RTGETGASGP PGFAGEKGPS GEPGTAGPPG TPGPQGILGA PGFLGLPGSR
801 GERGLPGVAG SVGEPGPLGI AGPPGARGPP GAVGNPGVNG APGEAGRDGN
851 PGSDGPPGRD GQAGHKGERG YPGNPGPAGA AGAPGPQGAV GPAGKHGNRG
901 EPGPAGSVGP AGAVGPRGPS GPQGIRGEKG EPGDKGPRGL PGLKGHNGLQ
951 GLPGLAGHHG DQGAPGPVGP AGPRGPAGPS GPAGKDGRTG QPGAVGPAGI
1001 RGSQGSQGPA GPPGPPGPPG PPGPSGGGYD FGYEGDFYRA
```

ESI

Pig sample pig α1 (I) coverage

```
  1 QLSYGYDEKS AGKISVPGPM GPSGPRGLPG PPGAPGPQGF QGPPGEPGEP
 51 GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG ARGLPGTAGL
101 PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM GPRGLPGERG
151 RPGPPGTAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA KGEAGPQGAR
201 GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG APGIAGAPGF
251 PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP GPTGIQGPPG
301 PAGEEGKRGA RGEPGPSGLP GPPGERGGPG SRGFPGADGV AGPKGPAGER
351 GSPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG KTGPPGPAGQ
401 DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP GPPGAVGPAG
451 KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP PGEAGKPGEQ
501 GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG APGNDGAKGD
551 AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK GADGAPGKDG
601 VRGLTGPIGP PGPAGAPGDK GETGPSGPAG PTGARGAPGD RGEPGPPGPA
651 GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPTGPPG PIGSVGAPGP
701 KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA GKEGSKGPRG
751 ETGPAGRPGE AGPPGPPGPA GEKGSPGADG PAGAPGTPGP QGIAGQRGVV
801 GLPGQRGERG FPGLPGPSGE PGKQGPSGPS GERGPPGPMG PPGLAGPPGE
851 SGREGAPGAE GSPGRDGAPG PKGDRGESGP AGPPGAPGAP GAPGPVGPAG
901 KSGDRGETGP AGPAGPVGPV GARGPAGPQG PRGDKGETGE QGDRGIKGHR
951 GFSGLQGPPG PPGAPGEQGP SGASGPAGPR GPPGSAGSPG KDGLNGLPGP
1001 IGPPGPRGRT GDAGPVGPPG PPGPPGPPGP PSGGFDFSFL PQPPQEKAHD
1051 GGRYYRA
```

Pig sample pig α2 (I) coverage

```
  1 QYDGKGVGAG PGPMGLMGPR GPPGAVGAPG PQGFQGPAGE PGEPGQTGPA
 51 GARGPPGPPG KAGEDGHPGK PGRPGERGVV GPQGARGFPG TPGLPGFKGI
101 RGHNGLDGLK GQPGAPGVKG EPGAPGENGT PGQTGARGLP GERGRVGAPG
151 PAGARGNDGS VGPVGPAGPI GSAGPPGFPG APGPKGELGP VGNPGPAGPA
201 GPRGEVGLPG VSGPVGPPGN PGANGLPGAK GAAGLPGVAG APGLPGPRGI
251 PGPAGAAGAT GARGLVGEPG PAGSKGESGN KGEPGAAGPQ GPPGPSGEEG
301 KRGPNGEVGS AGPPGPPGLR GNPGSRGLPG ADGRAGVMGP PGSRGPTGPA
351 GVRGPNGDSG RPGEPGLMGP RGFPGSPGNV GPAGKEGPAG LPGIDGRPGP
401 IGPAGARGEP GNIGFPGPKG PTGDPGKNGE KGHAGLAGAR GAPGPDGNNG
451 AQGPPGPQGV QGGKGEQGPA GPPGFQGLPG PAGTAGEVGK PGERGIPGEF
501 GLPGLAGPRG ERGPPGESGA AGPAGLIGSR GPSGPPGLDG NKGELGVLGA
551 PGTAGPSGLS GLPGERGAAG IPGGKGEKGE TGLRGDVGSP GRDGARGAPG
601 AVGAPGPAGA NGDRGEAGPA GPAGPAGPRG SPGERGEVGP AGPNGFAGPA
651 GAAGQPGAKG ERGTKGPKGE NGPVGPTGPV GAAGPAGPNG PPGPAGSRGD
701 GGPPGATGFP GAAGRIGPPG PSGISGPPGP PGPAGKEGLR GPRGDQGPVG
751 RTGETGASGP PGFAGEKGPS GEPGTAGPPG TPGPQGILGA PGFLGLPGSR
801 GERGLPGVAG SVGEPGPLGI AGPPGARGPP GAVGNPGVNG APGEAGRDGN
851 PGSDGPPGRD GQAGHKGERG YPGNPGPAGA AGAPGPQGAV GPAGKHGNRG
901 EPGPAGSVGP AGAVGPRGPS GPQGIRGEKG EPGDKGPRGL PGLKGHNGLQ
951 GLPGLAGHHG DQGAPGPVGP AGPRGPAGPS GPAGKDGRTG QPGAVGPAGI
1001 RGSQGSQGPA GPPGPPGPPG PPGPSGGGYD FGYEGDFYRA
```

Figure 9.6 – Sequence coverage map of best a1(I) and a2 (I) chain collagen matches to the 'Collagens' database from pig MBM sediment rendered at 145°C and analysed by LC-MALDI and LC-ESI. Red text indicates peptide matches in LC-MALDI analyses.

Sheep sample sheep α1 (I) coverage

Sheep sample sheep α2 (I) coverage

LC-MALDI

```
  1 QLSYGYDEKS TGKISVPGPM GPSGPRGLPG PPGAPGPQGF QGPPGEPGEP
 51 GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG ARGLPGTAGL
101 PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGTPGQM GPRGLPGERG
151 RPGAPGPAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA KGEAGPQGPR
201 GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG APGIAGAPGF
251 PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP GPTGIQGPPG
301 PAGEEGKRGA RGEPGPAGLP GPPGERGGPG SRGFPGSDGV AGPKGPAGER
351 GSPGEAGRPG EAGEAGRPGE AGLPGAKGLT GSPGSPGPDG KTGPPGPAGQ
401 DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP GPPGAVGPAG
451 KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP PGEAGKPGEQ
501 GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG APGNDGAKGD
551 AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK GADGAPGKDG
601 VRGLTGPIGP PGPAGAPGDK GETGPSGPAG PTGARGAPGD RGEPGPPGPA
651 GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPAGPPG PIGNVGAPGP
701 KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA GKEGSKGPRG
751 ETGPAGRAGE VGPPGPPGPA GEKGAPGADG PAGAPGTPGP QGIAGQRGVV
801 GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG PPGLAGPPGE
851 SGREGAPGAE GSPGRDGAPG AKGDRGETGP AGPPGAPGAP GAPGPVGPAG
901 KSGDRGETGP AGPAGPIGPV GARGPAGPQG PRGDKGETGE QGDRGIKGHR
951 GFSGLQGPPG PPGSPGEQGP SGASGPAGPR GPPGSAGTPG KDGLNGLPGP
1001 IGPPGPRGRT GDAGPAGPPG PPGPPGPPGP PSGGYDLSFL PQPPQEKAHD
1051 GGRYYRA
```

```
  1 QFDGKGGGPG PMGLMGPRGP PGASGAPGPQ GFQGPPGEPG EPGQTGPAGA
 51 RGPPGPPGKA GEDGHPGKPG RPGERGVVGP QGARGFPGTP GLPGFKGIRG
101 HNGLDGLKGQ PGAPGVKGEP GAPGENGTPG QTGARGLPGE RGRVGAPGPA
151 GARGSDGSVG PVGPAGPIGS AGPPGFPGAP GPKGELGPVG NPGPAGPAGP
201 RGEVGLPGLS GPVGPPGNPG ANGLPGAKGA AGLPGVAGAP GLPGPRGIPG
251 PVGAAGATGA RGLVGEPGPA GSKGESGNKG EPGAVGQPGP PGPSGEEGKR
301 GSTGEIGPAG PPGPPGLRGN PGSRGLPGAD GRAGVMGPAG SRGATGPAGV
351 RGPNGDSGRP GEPGLMGPRG FPGSPGNIGP AGKEGPAGLP GIDGRPGPIG
401 PAGARGEPGN IGFPGPKGPT GDPGKAGEKG HAGLAGPRGA PGPDGNNGAQ
451 GPPGLQGVQG GKGEQGPAGP PGFQGLPGPA GTAGEAGKPG ERGIPGEFGL
501 PGPAGARGER GPPGESGAAG PTGPIGSRGP SGPPGPDGNK GEPGVVGAPG
551 TAGPSGPSGL PGERGAAGIP GGKGEKGETG LRGDVGSPGR DGARGAPGAV
601 GAPGPAGANG DRGEAGPAGP AGPAGPRGSP GERGEVGPAG PNGFAGPAGA
651 AGQPGAKGER GTKGPKGENG PVGPTGPVGA AGPSGPNGPP GPAGSRGDGG
701 PPGATGFPGA AGRTGPPGPA GISGPPGPPG PAGKEGLRGP RGDQGPVGRT
751 GEPGAAGPPG FVGEKGPSGE PGTAGPPGTP GPQGLLGAPG FLGLPGSRGE
801 RGLPGVAGSV GEPGPLGIAG PPGARGPPGN VGNPGVNGAP GEAGRDGNPG
851 NDGPPGRDGQ PGHKGERGYP GNAGPVGAAG APGPQGPVGP TGKHGSRGEP
901 GPVGAVGPAG AVGPRGPSGP QGIRGDKGEP GDKGPRGLPG LKGHNGLQGL
951 PGLAGHHGDQ GAPGAVGPAG PRGPAGPTGP AGKDGRTGQP GAVGPAGIRG
1001 SQGSQGPAGP PGPPGPPGPP GPSGGGYDFG FDGDFYRA
```

ESI

```
  1 QLSYGYDEKS TGKISVPGPM GPSGPRGLPG PPGAPGPQGF QGPPGEPGEP
 51 GASGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG ARGLPGTAGL
101 PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGTPGQM GPRGLPGERG
151 RPGAPGPAGA RGNDGATGAA GPPGPTGPAG PPGFPGAVGA KGEAGPQGPR
201 GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG APGIAGAPGF
251 PGARGPSGPQ GPSGPPGPKG NSGEPGAPGS KGDTGAKGEP GPTGIQGPPG
301 PAGEEGKRGA RGEPGPAGLP GPPGERGGPG SRGFPGSDGV AGPKGPAGER
351 GSPGEAGRPG EAGEAGRPGE AGLPGAKGLT GSPGSPGPDG KTGPPGPAGQ
401 DGRPGPPGPP GARGQAGVMG FPGPKGAAGE PGKAGERGVP GPPGAVGPAG
451 KDGEAGAQGP PGPAGPAGER GEQGPAGSPG FQGLPGPAGP PGEAGKPGEQ
501 GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGANG APGNDGAKGD
551 AGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK GADGAPGKDG
601 VRGLTGPIGP PGPAGAPGDK GETGPSGPAG PTGARGAPGD RGEPGPPGPA
651 GFAGPPGADG QPGAKGEPGD AGAKGDAGPP GPAGPAGPPG PIGNVGAPGP
701 KGARGSAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPA GKEGSKGPRG
751 ETGPAGRAGE VGPPGPPGPA GEKGAPGADG PAGAPGTPGP QGIAGQRGVV
801 GLPGQRGERG FPGLPGPSGE PGKQGPSGAS GERGPPGPMG PPGLAGPPGE
851 SGREGAPGAE GSPGRDGAPG AKGDRGETGP AGPPGAPGAP GAPGPVGPAG
901 KSGDRGETGP AGPAGPIGPV GARGPAGPQG PRGDKGETGE QGDRGIKGHR
951 GFSGLQGPPG PPGSPGEQGP SGASGPAGPR GPPGSAGTPG KDGLNGLPGP
1001 IGPPGPRGRT GDAGPAGPPG PPGPPGPPGP PSGGYDLSFL PQPPQEKAHD
1051 GGRYYRA
```

```
  1 QFDGKGGGPG PMGLMGPRGP PGASGAPGPQ GFQGPPGEPG EPGQTGPAGA
 51 RGPPGPPGKA GEDGHPGKPG RPGERGVVGP QGARGFPGTP GLPGFKGIRG
101 HNGLDGLKGQ PGAPGVKGEP GAPGENGTPG QTGARGLPGE RGRVGAPGPA
151 GARGSDGSVG PVGPAGPIGS AGPPGFPGAP GPKGELGPVG NPGPAGPAGP
201 RGEVGLPGLS GPVGPPGNPG ANGLPGAKGA AGLPGVAGAP GLPGPRGIPG
251 PVGAAGATGA RGLVGEPGPA GSKGESGNKG EPGAVGQPGP PGPSGEEGKR
301 GSTGEIGPAG PPGPPGLRGN PGSRGLPGAD GRAGVMGPAG SRGATGPAGV
351 RGPNGDSGRP GEPGLMGPRG FPGSPGNIGP AGKEGPAGLP GIDGRPGPIG
401 PAGARGEPGN IGFPGPKGPT GDPGKAGEKG HAGLAGPRGA PGPDGNNGAQ
451 GPPGLQGVQG GKGEQGPAGP PGFQGLPGPA GTAGEAGKPG ERGIPGEFGL
501 PGPAGARGER GPPGESGAAG PTGPIGSRGP SGPPGPDGNK GEPGVVGAPG
551 TAGPSGPSGL PGERGAAGIP GGKGEKGETG LRGDVGSPGR DGARGAPGAV
601 GAPGPAGANG DRGEAGPAGP AGPAGPRGSP GERGEVGPAG PNGFAGPAGA
651 AGQPGAKGER GTKGPKGENG PVGPTGPVGA AGPSGPNGPP GPAGSRGDGG
701 PPGATGFPGA AGRTGPPGPA GISGPPGPPG PAGKEGLRGP RGDQGPVGRT
751 GEPGAAGPPG FVGEKGPSGE PGTAGPPGTP GPQGLLGAPG FLGLPGSRGE
801 RGLPGVAGSV GEPGPLGIAG PPGARGPPGN VGNPGVNGAP GEAGRDGNPG
851 NDGPPGRDGQ PGHKGERGYP GNAGPVGAAG APGPQGPVGP TGKHGSRGEP
901 GPVGAVGPAG AVGPRGPSGP QGIRGDKGEP GDKGPRGLPG LKGHNGLQGL
951 PGLAGHHGDQ GAPGAVGPAG PRGPAGPTGP AGKDGRTGQP GAVGPAGIRG
1001 SQGSQGPAGP PGPPGPPGPP GPSGGGYDFG FDGDFYRA
```

Figure 9.7 – Sequence coverage map of best a1(I) and a2 (I) chain collagen matches to the 'Collagens' database from sheep MBM sediment rendered at 145°C and analysed by LC-MALDI and LC-ESI. Red text indicates peptide matches in LC-MALDI analyses.

In the pig α1 collagen (I) matches (to pig collagen in both LC-MALDI and LC-ESI analyses), the LC-MALDI search results matched only two unique peptides and the LC-ESI search results matched thirteen unique peptides. In the pig α2 collagen (I) matches (to pig collagen in both LC-MALDI and LC-ESI analyses), only three unique peptides were matched with LC-MALDI and fourteen with LC-ESI. In the sheep α1 collagen (I) matches (to sheep collagen in both LC-MALDI and LC-ESI analyses), the LC-MALDI search results matched fifteen unique peptides and the LC-ESI search results matched eleven unique peptides. In the sheep α2 collagen matches (to sheep collagen in both LC-MALDI and LC-ESI analyses), seventeen unique peptides were matched with LC-MALDI and five were matched with LC-ESI. For this particular sheep MBM sediment sample, LC-ESI-based methods did not obtain as many unique matches as the LC-MALDI-based methods (contrary to the LC-MS analyses of the pig sample where LC-ESI-based analysis matched

many more unique peptides than the LC-MALDI-based analysis). The few peptides matched in the LC-MALDI analyses for both pig and sheep MBM samples that were not observed in LC-ESI analyses are predominantly arginine-terminated, whereas the majority of the large numbers of peptides matched in the LC-ESI analyses but not in the LC-MALDI analyses are predominantly lysine-terminated (see Appendix 5.4). This phenomenon, where the lysine-terminated peptides are less commonly observed in MALDI than in ESI owing to poor ionization and suppression of the signal by the dominant arginine-terminated peptides has been observed elsewhere (Krause *et al.* 1999). The other clearly observed trend in the peptides that are observed is that many of the peptides observed in LC-ESI are small peptides (<10 amino acids long), which could be expected due to the more limited peptide mass acquisition parameters used in the LC-MALDI method. Clear differences in the amino acid content of the peptides (and thus hydrophobicity) that could be expected to differ between the two methods (Olumee *et al.* 1995) were not observed in either LC-MALDI or LC-ESI analyses, most likely due to the homogenous repeating sequence (i.e., Gly-Pro-Xaa) throughout the molecule.

To compare the sequence coverages obtained from LC-MALDI and LC-ESI of both pig and sheep MBM sediment samples to those obtained from LC-MALDI and LC-ESI of pig and sheep bone samples, the results were plotted as a histogram (Fig. 9.8).

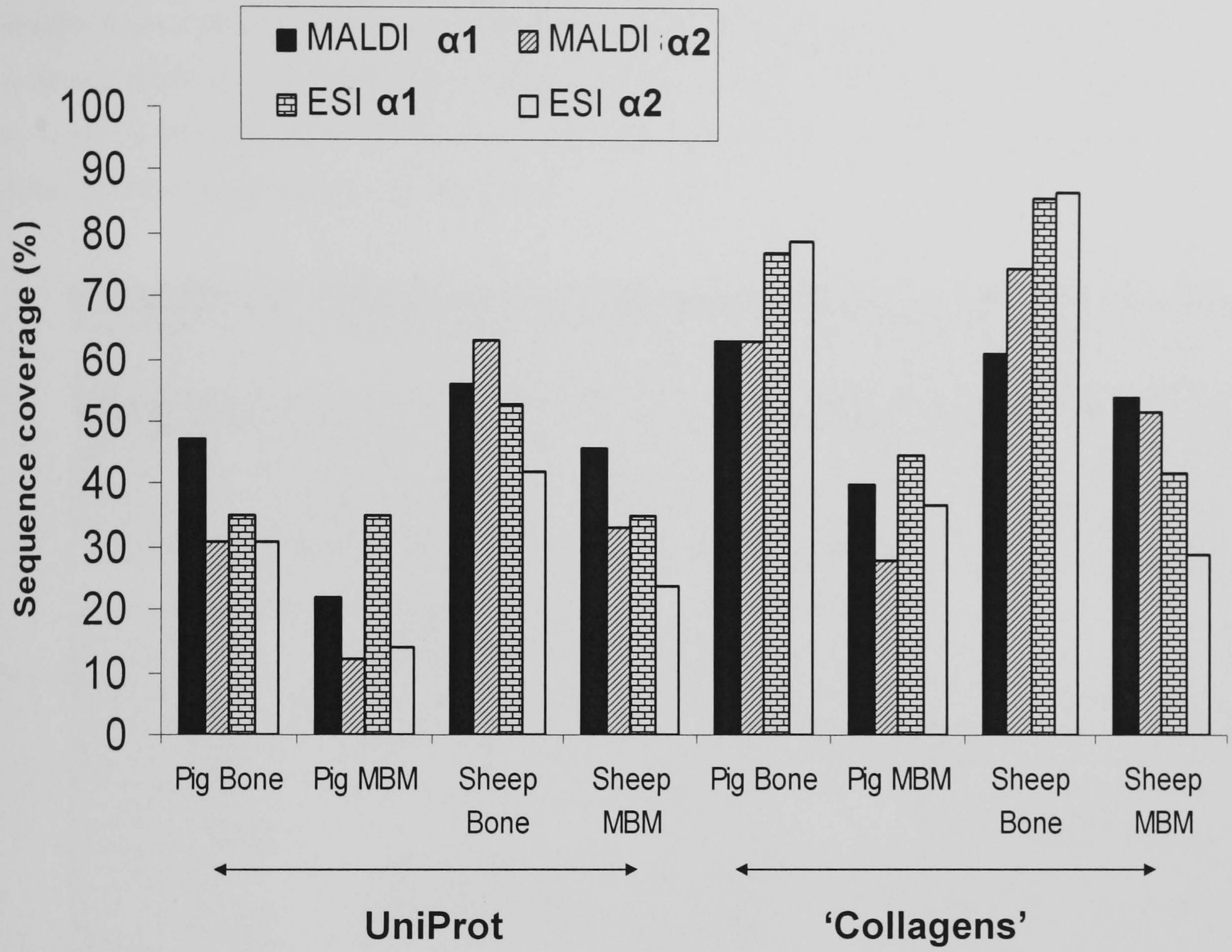


Figure 9.8 - Histogram showing the sequence coverages (in %) of both a1 (I) and a2 (I) chains for LC-MALDI and LC-ESI analyses. Modern untreated pig and sheep bone ('Pig Bone' and 'Sheep Bone' respectively) and sedimented pig and sheep MBM ('Pig MBM' and 'Sheep MBM' respectively) are shown.

The sequence coverages were observed to be higher in the searches against the 'Collagens' database than in the searches against the limited collagen sequences currently available in the public UniProt database (see Fig. 9.8). It is also clear that less sequence coverage was obtained for each alpha chain in each LC-MS analysis for both MBM sediment samples than was obtained from the same analyses on bone samples (of the same species).

9.4.4 Identifying Species-Specific Collagen (I) Peptides in the LC-MS Search Results

In order to view and select potentially useful peptides for distinguishing between cattle, sheep pig and chicken, the top collagen α1 (I) and α2 (I) chain sequences obtained for each

sample (three replicate cattle bone samples, cattle MBM sediment, sheep bone, sheep MBM sediment, pig bone, pig MBM sediment, chicken bone, chicken MBM sediment, all analysed by both LC-MALDI and LC-ESI) were separately studied in BioEdit Sequence Alignment Editor software version 7.0.0 (see Fig. 9.9).

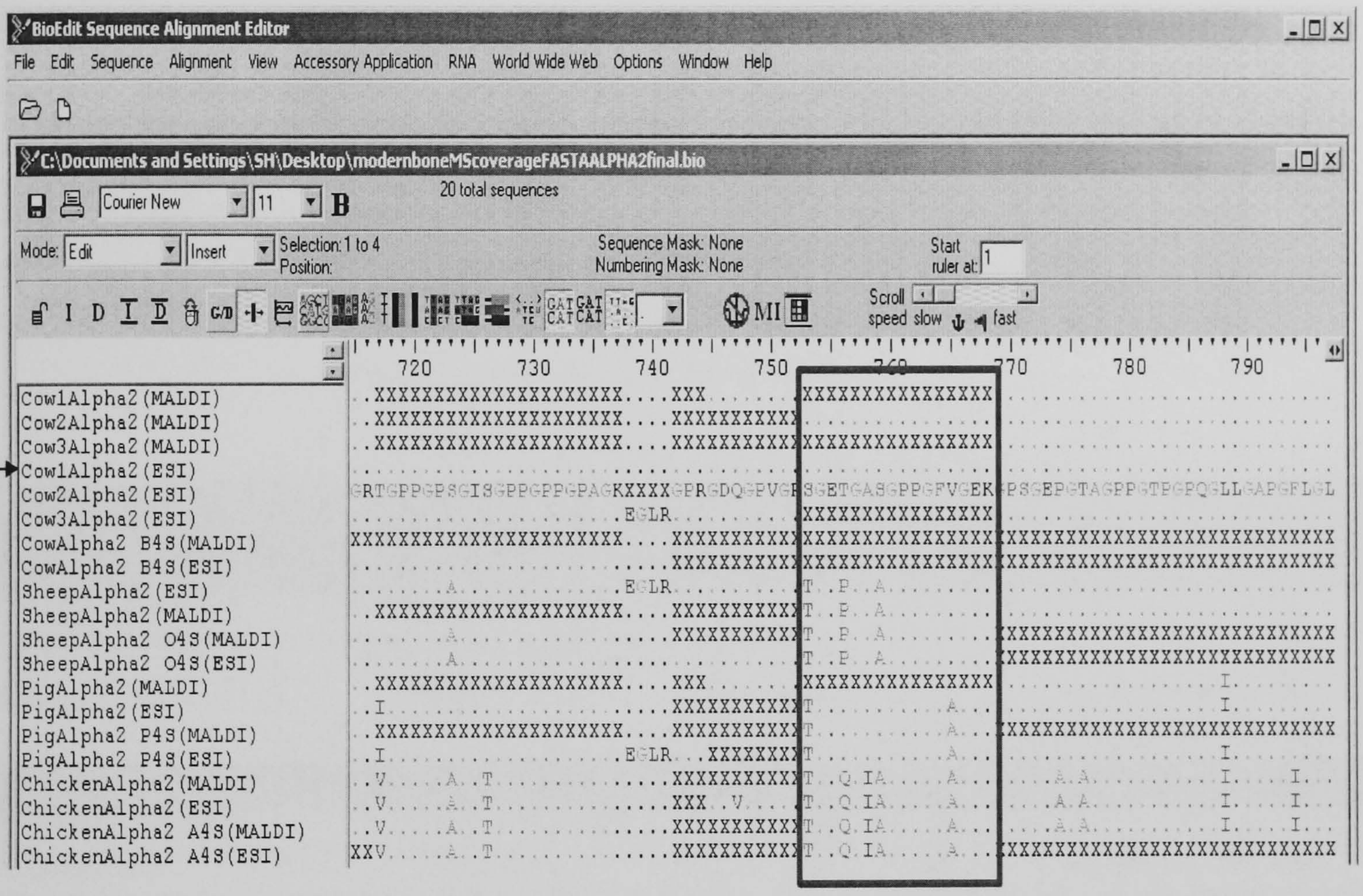


Figure 9.9 - Screenshot of BioEdit (Hall 1999) showing 20 sequences from a2(I) chain coverages of investigated samples. X represents sequence not identified by Mascot and dots represent conserved sequence to the one selected (arrow). The box highlights a single tryptic peptide capable of distinguishing between all four species.

Peptides that possess different amino acid sequences between cattle and sheep, or between cattle and pig, and observed in all LC-MS analyses of the bone samples were chosen as potential species-specific markers (Tables 9.5-9.8). Of these, only peptides that were matched in at least half of the eight MBM sample analyses (four species by LC-MALDI and LC-ESI) were selected as potential species-specific markers and are listed for cattle and sheep (shaded cells in Tables 9.5 and 9.7) and for cattle and pig (shaded cells in Tables 9.6 and 9.8) and presented for cattle, pig, sheep and chicken in Table 9.9. Additional observations, such as the frequency of occurences of each peptide match throughout the analyses, or the presence of Pro-Ser or Hyp-Ala motifs (that are isobaric and may be difficult

to differentiate in poor quality product ion mass spectra – particularly as fragmentation amino-terminal to proline in observed much more than would be expected in CID spectra (Breci *et al.* 2003; Schaaff *et al.* 2000), and fragmentation carboxy-terminal to proline is unusually less commonly observed (Vaisar and Urban 1996)). In order to show the peptides with most variable sequences between species, only mammalian peptide sequences are listed in Tables 9.5-9.8; amino acid differences between avian collagen (I) peptides and mammalian collagen peptides are much more common throughout the molecule.

Table 9.5 – Collagen a1 (I) chain peptide sequences matched against the 'Collagens' database that are potentially useful to distinguish between cattle and sheep. Peptides that were observed in at least half of the eight MBM runs (the four species analysed by LC-MALDI and LC-ESI), and not potentially obscured by the Hyp modification, are shaded and written in bold type. Amino acid differences are indicated by underlined type.

Cattle Peptide Sequence	Sheep Peptide Sequence	Occurence in 8 MBM	Comments
GEPGSPGENGAPGQMGPR	GEPGSPGENGTPGQMGPR	1/8	-
GEGGPQGPRGSEGPQGVR	GEAGPQGPRGSEGPQGVR	0/8	-
GFPGADGVAGPK	GFPGSDGVAGPK	4/8	Ala to Ser obscured by Hyp Cattle sequence same as pig sequence
GEAGPSGPAGPTGAR	**GETGPSGPAGPTGAR**	**6/8**	**Pig sequence same as Sheep sequence**
PGEVGPPGPPGPAGEK	AGEVGPPGPPGPAGEK	1/8	-
DGSPGAK	DGAPGAK	2/8	Ala to Ser obscured by Hyp
GPPGSAGSPGK	GPPGSAGTPGK	3/8	Pig sequence same as catte sequence

Of the seven α1 (I) chain tryptic peptides that vary in amino acid sequence between cattle and sheep (Table 9.5) that were observed in bone samples, only two were observed in at least half of the MBM samples. Of these two samples, only one was not complicated by the alanine to serine replacement (+16 Da) isobaric with the hydroxylation of proline (also +16 Da). However, this peptide (GEAGPSGPAGPTGAR in cattle) has an identical sequence (GETGPSGPAGPTGAR) in sheep and pig.

Table 9.6 – Collagen a1 (I) chain sequences matched against the 'Collagens' database that are potentially useful to distinguish between bovid (cattle shown) and pig. Peptides that were observed in at least half of the eight MBM runs (the four species analysed by LC-MALDI and LC-ESI), and not potentially obscured by the Hyp modification, are shaded and written in bold type. Amino acid differences are indicated by underlining.

Cattle Peptide Sequence	Pig Peptide Sequence	Occurence in 8 MBM	Comments
PGAPGPAGAR	PGPPGTAGAR	2/8	Sheep sequence same as cattle sequence
GEGGPQGPRGSEGPQGVR	GEAGPQGARGSEGPQGVR	0/8	-
GEPGPAGLPGPPGER	GEPGPSGLPGPPGER	8/8	Ala to Ser may be obscured by Hyp Cattle sequence same as sheep sequence
GAPGPAGPK	GSPGPAGPK	0/8	Ala to Ser may be obscured by Hyp
GEAGPSGPAGPTGAR	**GETGPSGPAGPTGAR**	**6/8**	**Sheep sequence same as Pig sequence**
GDAGPPGPAGPAGPPGPIGNVGAPGPK	GDAGPPGPAGPTGPPGPIGSVGAPGPK	2/8	Sheep same as cattle
GAPGADGPAGAPGTPGPQGIAGQR	GSPGADGPAGAPGTPGPQGIAGQR	3/8	Ala to Ser may be obscured by Hyp
QGPSGASGER	QGPSGPSGER	3/8	Cattle sequence same as sheep sequence
DGSPGAK	DGAPGPK	2/8	Small peptide
GETGPAGPPGAPGAPGAPGPVGPAGK	**GESGPAGPPGAPGAPGAPGPVGPAGK**	**6/8**	**Cattle sequence same as sheep sequence**
GETGPAGPAGPIGPVGAR	**GETGPAGPAGPVGPVGAR**	**6/8**	**Cattle sequence same as sheep sequence**
GFSGLQGPPGPPGSPGEQGPSGASGPAGPR	GFSGLQGPPGPPGAPGEQGPSGASGPAGPR	8/8	Ser to Ala may be obscured by Hyp Cattle sequence same as sheep sequence
TGDAGPAGPPGPPGPPGPPGPPSGGYDLSFLPQPPQEK	TGDAGPVGPPGPPGPPGPPGPPSGGFDFSFLPQPPQEK	1/8	Cattle sequence same as sheep sequence

Of the 13 α1 (I) chain peptides capable of distinguishing between cattle and pig (Table 6.6), only five peptides were observed in spectra from at least half of the MBM samples. Of these five samples, only three are not potentially complicated by the alanine to serine replacement. However, two of these have identical sequences in cattle and sheep, and the other has identical sequences in pig and sheep.

Table 9.7 – Collagen a2 (I) chain sequences matched against the 'Collagens' database that are potentially useful to distinguish between cattle and sheep. Peptides that are observed in at least half of the eight MBM runs (the four species analysed by LC-MALDI and LC-ESI), and not obscured by the Hyp modification, are shaded and written in bold type. Amino acid differences are indicated by underlining.

Cattle Peptide Sequence	Sheep Peptide Sequence	Occurence in 8 MBM	Comments
QFAKGGGPGPMGLMGPR	QFDGKGGGPGPMGLMGPR	2/8	Additional amino acid present in sheep sequence
EGPVGLPGIDGR	EGPAGLPGIDGR	1/8	Pig sequence same as sheep sequence
GHAGLAGAR	GHAGLAGPR	1/8	Pig sequence same as cattle sequence
GETGLRGDIGSPGR	GETGLRGDVGSPGR	0/8	
GAPGAIGAPGPAGANGDR	**GAPGAVGAPGPAGANGDR**	**4/8**	**Pig sequence same as sheep sequence**
TGPPGPAGISGPPGPPGPAGK	TGPPGPSGISGPPGPPGPAGK	6/8	Ala to Ser may be obscured by Hyp
SGETGASGPPGFVGEK	**TGEPGAAGPPGFVGEK**	**6/8**	
GYPGNAGPVGAAGAPGPQGPVGPVGK	GYPGNAGPVGAAGAPGPQGPVGPTGK	2/8	
GEPGPAGAVGPAGAVGPR	**GEPGPVGAVGPAGAVGPR**	**6/8**	
GPAGPSGPAGK	GPAGPTGPAGK	1/8	Pig sequence same as cattle sequence
IGQPGAVGPAGIR	**TGQPGAVGPAGIR**	**6/8**	**Sheep sequence same as pig sequence**
GSQGSQGPAGPPGPPGPPGPPGPSGGGYEFGFDGDFYR	GSQGSQGPAGPPGPPGPPGPPGPSGGGYDFGFDGDFYR	0/8	Carboxytelopeptide

Of the 12 α2 (I) chain peptides that are different in cattle and sheep (Table 9.7), only five were observed in at least half of the MBM sample analyses. Of these five peptides, only one is potentially complicated by the alanine to serine replacement (+16 Da) isobaric with hydroxylation of proline (also +16 Da). However, of the remaining four potentially diagnostic peptides, two have identical sequences in sheep and pig. The two remaining α2 (I) peptides capable of distinguishing between the cattle and sheep are SGETGASGPPGFVGEK and GEPGPAGAVGPAGAVGPR (in cattle). The carboxytelopeptide markers that were not observed following extraction and analysis using bacterial collagenase digestion and SPE isolation were also not detected in all LC-MS analyses of the MBM samples (Table 9.7).

Table 9.8 – Collagen a2 (I) chain sequences matched against the 'Collagens' database that are potentially useful to distinguish between bovid (cattle shown) and pig. Peptides that are observed in at least half of the eight MBM runs, and not obscured by the Hyp modification, are shaded and written in bold type. Amino acid differences are indicated by underlining.

Cattle Peptide Sequence	Pig Peptide Sequence	Occurence in 8 MBM	Comments
GPPGASGAPGPQGFQGPPGEPGEP GQTGPAGAR	GPPGAVGAPGPQGFQGPAGEPGEP GQTGPAGAR	3/8	Sheep sequence same as cattle sequence
GSDGSVGPVGPAGPIGSAGPPGFPG APGPK	GNDGSVGPVGPAGPIGSAGPPGFPG APGPK	0/8	Sheep sequence same as cattle sequence
GEVGLPGLSGPVGPPGNPGANGLP GAK	GEVGLPGVSGPVGPPGNPGANGLP GAK	0/8	Sheep sequence same as cattle sequence
GIPGPVGAAGATGAR	**GIPGPAGAAGATGAR**	**7/8**	**Sheep sequence same as cattle sequence**
GEPGAVGQPGPPGPSGEEGK	GEPGAAGPQGPPGPSGEEGK	3/8	Sheep sequence same as cattle sequence
AGVMGPAGSR	AGVMGPPGSR	1/8	Sheep sequence same as cattle sequence
GATGPAGVR	GPTGPAGVR	1/8	Sheep sequence same as cattle sequence
GFPGSPGNIGPAGK	GFPGSPGNVGPAGK	0/8	Sheep sequence same as cattle sequence
EGPVGLPGIDGR	EGPAGLPGIDGR	1/8	Sheep sequence same as pig sequence
GEQGPAGPPGFQGLPGPAGTAGEA GK	GEQGPAGPPGFQGLPGPAGTAGEV GK	1/8	Sheep sequence same as cattle sequence
GIPGEFGLPGPAGAR	GIPGEFGLPGLAGPR	8/8	Sheep sequence same as cattle sequence Pro to Leu may be obscured by Hyp
GPSGPPGPDGNK	GPSGPPGLDGNK	2/8	Sheep sequence same as cattle sequence
GEPGVVGAPGTAGPSGPSGLPGER	GELGVLGAPGTAGPSGLSGLPGER	2/8	Sheep sequence same as cattle sequence
GAPGAIGAPGPAGANGDR	**GAPGAVGAPGPAGANGDR**	**4/8**	**Sheep sequence same as pig sequence**
GENGPVGPTGPVGAAGPSGPNGPP GPAGSR	GENGPVGPTGPVGAAGPAGPNGPP GPAGSR	2/8	Ser to Ala may be obscured by Hyp
TGPPGPSGISGPPGPPGPAGK	**IGPPGPSGISGPPGPPGPAGK**	**6/8**	
SGETGASGPPGFVGEK	**TGETGASGPPGFAGEK**	**5/8**	
GPSGEPGTAGPPGTPGPQGLLGAPG FLGLPGSR	GPSGEPGTAGPPGTPGPQGILGAPG FLGLPGSR	3/8	L-I isobaric, cattle same as sheep
GPPGNVGNPGVNGAPGEAGR	GPPGAVGNPGVNGAPGEAGR	1/8	Sheep sequence same as cattle sequence
DGNPGNDGPPGR	DGNPGSDGPPGR	3/8	Sheep sequence same as cattle sequence
GYPGNAGPVGAAGAPGPQGPVGPV GK	GYPGNPGPAGAAGAPGPQGAVGPA GK	2/8	Sheep sequence same as cattle sequence
GEPGPAGAVGPAGAVGPR	GEPGPAGSVGPAGAVGPR	6/8	Ala to Ser may be obscured by Hyp
GHNGLQGLPGLAGHHGDQGAPGA VGPAGPR	GHNGLQGLPGLAGHHGDQGAPGP VGPAGPR	1/8	Sheep sequence same as cattle sequence
IGQPGAVGPAGIR	**TGQPGAVGPAGIR**	**6/8**	**Sheep sequence same as pig sequence**
GSQGSQGPAGPPGPPGPPGPPGPSG GGYEFGFDGDFYR	GSQGSQGPAGPPGPPGPPGPPGPSG GGYDFGYEGDFYR	0/8	Carboxytelopeptide

Of the 26 α2 (I) chain peptides capable of distinguishing cattle and pig (Table 9.8), only seven were observed in at least half of the MBM samples. Of these seven peptides, only two are potentially complicated by the alanine to serine replacement. However, of the remaining five potentially diagnostic peptides, two share sequences in sheep and pig, and two share sequences in cattle and sheep. The two remaining α2 (I) peptides capable of distinguishing between cattle and pig are SGETGASGPPGFVGEK and TGPPGPSGISGPPGPPGPAGK (in cattle). Thus the only peptide observed frequently in the analysis of MBM samples that has a unique amino acid sequence in cattle, pig and sheep is SGETGASGPPGFVGEK (in cattle).

9.4.5 Species Identification in an MBM Sample Containing Material from Several Different Species

To investigate the data obtained from LC-MS analysis of the mixed MBM sediment sample (which could contain non-mammalian species protein) the Mascot search results were filtered for matches to 'Collagens' proteins that include a top-ranking peptide match (as determined by Mascot 'bold red' peptide matches) (Table 9.9). The sequence coverage of each match was uploaded into BioEdit as FASTA files. Peptide matches from the same location in the protein and contain different amino acids between different species were used to infer the minimum possible species present (e.g., Fig. 9.10).

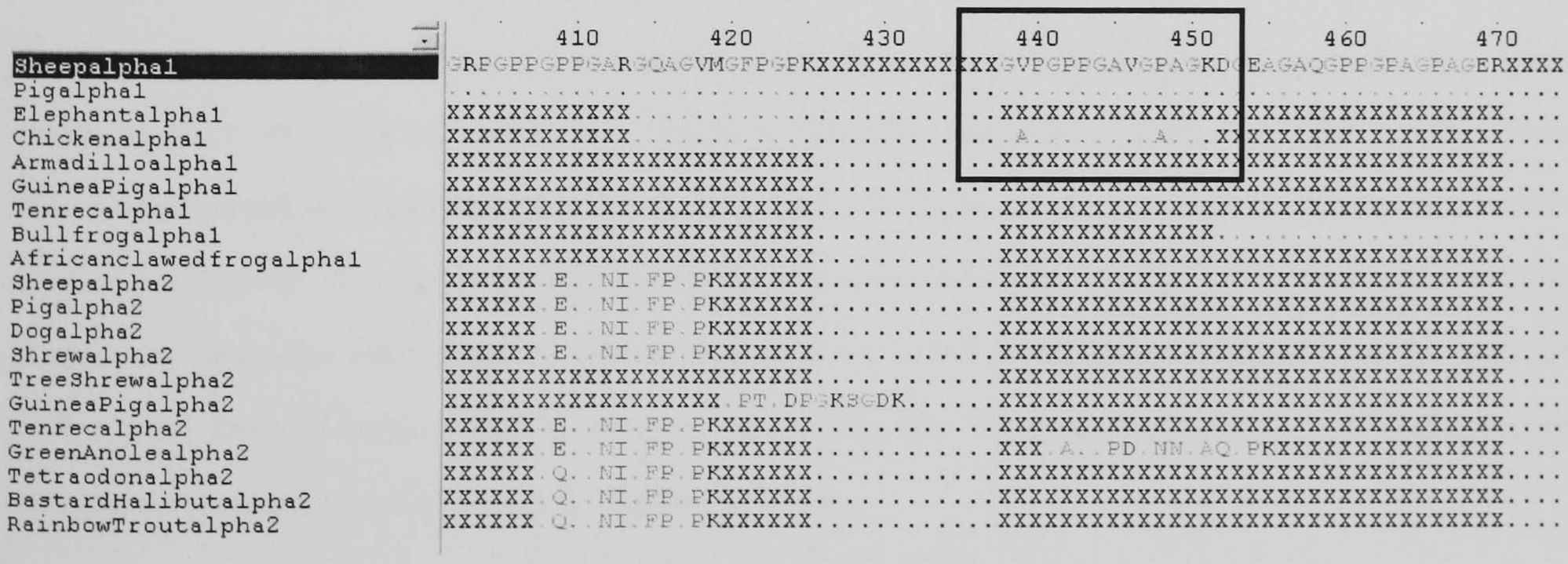


Figure 9.10 - Screenshot of the use of BioEdit in identifying peptide sequences matched in the LC-ESI analysis of the mixed MBM sample that are also matched as several different species. The box indicating the presence of both chicken and sheep/pig sequence from the same position. Dots indicate conserved amino acid sequence (from the sheep alpha 1 sequence) and X indicates that the peptide was not matched in the LC-MS analysis.

Although the results of searching the LC-ESI data against the 'Collagens' database matched nine α1 (I) sequences (see Appendix 5.5), only six of these (sheep, armadillo, elephant, pig, chicken and guinea pig) included peptide matches for the same position in the sequence. The only one of these six protein matches that could be misidentified due to Ala-Pro to Ser-Hyp replacements is the match to armadillo collagen. Thus, assuming a misidentification, the minimum possible species constituents of the mixed MBM sample as reported by a Mascot search against the 'Collagens' database is thus sheep, pig, chicken, elephant and guinea pig collagen. Of these, sheep, pig and chicken are realistic identifications whereas matches to elephant and guinea pig are not, as they are not added to the MBM. Of the 11 α2 (I) sequences matched, sheep, shrew, pig, tenrec, tree shrew, lizard (green anole), guinea pig and tetraodon are the eight matches that have peptide matches to identical positions in the protein. Of these, only the matches to lizard (green anole) and tetraodon collagen appear likely due to Ala-Pro to Ser-Hyp replacements. Of the remaining matches to sheep, pig, shrew, tree shrew, tenrec and guinea pig collagen, only the matches to sheep and pig collagen are likely, although the presence of shrew in MBM should not be dismissed. Searches using the 'Collagens' database thus result in several matches to collagen from species that are not present in the sample. These incorrect matches are probably caused by the presence of many similar sequences present in the 'Collagens' database, as described in section 4.4.

As the range of species present in the mixed MBM sediment sample is generally considered to be much more limited than those presented in the 'Collagens database' (the mixture most likely comprises the species of economic interest such as cattle, pig, sheep, chicken plus much smaller amounts of randomly disposed carcasses such as mouse, rat, rabbit, etc.), then the chosen peptide markers using LC-MS (Tables 9.5-9.8) could be used to determine which species are present (as determined by the presence of collagen). Table 9.9 lists the potentially diagnostic peptides identified in previously (Tables 9.5-9.8) that are detected on LC-MS of the trypsin treated mixed MBM sample, including the corresponding chicken collagen (I) sequences (not previously listed in Tables 9.5-9.8).

Table 9.9 - The chosen LC-MS collagen-peptide marker sequences and whether or not they were observed in the LC-ESI analysis of the mixed MBM sample. Merged cells indicate identical sequences. Matched peptides are in bold and underlined.

Protein Position	Cattle α1 sequence	Sheep α1 sequence	Pig α1 sequence	Chicken α1 sequence
621-635	GEAGPSGPAGPTGAR	**GETGPSGPAGPTGAR**		GEAGPPGPAGPTGAR
876-901	**GETGPAGPPGAPGAPGAPGPVGPAGK**		GESGPAGPPGAPGAPGAPGPVGPAGK	GESGPAGPPGAPGAPGAPGPVGPAGK
906-923	**GETGPAGPAGPIGPVGAR**		**GETGPAGPAGPVGPVGAR**	GETGPAGPAGPPGPAGAR
Protein Position	**Cattle α2 sequence**	**Sheep α2 sequence**	**Pig α2 sequence**	**Chicken α2 sequence**
247-261	GIPGPVGAAGATGAR		GIPGPAGAAGATGAR	GIPGPAGAAGATGAR
595-612	GAPGAIGAPGPAGANGDR	**GAPGAVGAPGPAGANGDR**		GLPGAIGAPGPAGGAGDR
713-734	TGPPGPSGISGPPGPPGPAGK	TGPPGPAGISGPPGPPGPAGK	IGPPGPSGISGPPGPPGPAGK	VGPPGPAGITGPPGPPGPAGK
750-765	SGETGASGPPGFVGEK	**TGEPGAAGPPGFVGEK**	TGETGASGPPGFAGEK	TGEQGIAGPPGFVGEK
898-915	GEPGPAGAVGPAGAVGPR	GEPGPVGAVGPAGAVGPR	GEPGPAGSVGPAGAVGPR	GDPGPVGPVGPAGAFGPR
987-999	IGQPGAVGPAGIR	**TGQPGAVGPAGIR**		TGQPGAVGPAGIR

Of cattle, sheep, pig and chicken peptide sequences, only a single species-specific peptide was matched in the LC-ESI analysis of the mixed MBM sample. This was to the sheep-specific peptide TGEPGAAGPPGFVGEK (Table 9.9). Of most interest to this sample was the identification of both the cattle/sheep peptide GETGPAGPAGPIGPVGAR and the pig-specific peptide GETGPAGPAGPVGPVGAR from the same position in the protein. This implies that both sheep and pig collagen is present in this mixture. When all peptide identifications are placed in order of S/N, the first (of greatest S/N) is the cattle/sheep peptide GETGPAGPAGPIGPVGAR (m/z 1559.8). The pig-specific peptide GETGPAGPAGPVGPVGAR (m/z 1545.8) ranks 61st (in terms of its S/N). Thus it was possible to infer that pig collagen is a contaminant of a dominantly sheep sample, where the relative peak intensity of the sheep/cattle peptide is 4.95 x greater than that of the peak for pig collagen. None of the nine chosen peptides (Table 9.9) was identified for chicken collagen, despite the indication that chicken collagen could be present (based on LC-MS Mascot search results); a single chicken α1 (I) peptide GAPGPPGAVGAAGK (m/z 1153.6) was observed in addition to the corresponding sheep/pig peptide GVPGPPGAVGPAGK (m/z 1191.6) where the relative S/N of the sheep/pig peptide was approximately 6 x greater than the chicken peptide. However, this peptide is not one of the nine selected peptides for species identification using LC-MS analyses implying the need to include more avian-specific

peptides that survive the MBM rendering processes. This would require MBM samples of alternate avian species, other than chicken, for such investigation.

9.5 Discussion

Attempts at isolating both the carboxytelopeptide (following digestion with collagenase and isolation by SPE) and peptides from the helical region of collagen (following digestion with trypsin and isolation by SPE) resulted in poor quality spectra in the lower temperature rendered MBM sediment samples (Figs 9.1 & 9.2); the only peptide observed in the highest temperature rendered samples derived from haemoglobin and has a highly conserved amino acid sequence. The implications are that the SPE-based methods that were successful in the analysis of archaeological bone samples (see Chapter 6 & 8) could not be adequately employed to distinguish between MBM sediment samples. Speculatively, as these two methods use different enzymes, the increased noise observed in the spectra of the 'helix' peptides could indicate that increasing amounts of signal-suppressing compounds such as sugars or lipids are present in the fractions containing isolated peptides, perhaps associated with the isolated collagen peptides via complex Maillard reactions. The reduction of this noise in the 145°C-rendered MBM sediment samples could indicate the breakdown of these signal-suppressing compounds. However, the much lower amounts of noise in the spectra following the carboxytelopeptide isolation method (by digestion with collagenase) could be due to the absence of lipid associated or sugar associated (cross-linked) peptides, as this method isolates only the non-cross-linked carboxytelopeptide (whereas the SPE fractions of peptides released with trypsin digestion potentially contain high molecular weight cross-linked peptides from the helical region of the molecule that are not observed in the spectra).

The lack of signal observed for the carboxytelopeptide in the analysis of the highest temperature rendered (145°C) MBM sediment sample using the SPE methods was consistent with its absence in all eight LC-MS analyses of the MBM sediment samples. Of the four 'helix' peptides that were not observed in the SPE analyses of the highest temperature rendered MBM sediment samples, Peptides A and B were observed in the LC-MALDI and LC-ESI analyses of all four species' (cattle, pig, sheep and chicken) MBM sediment samples. Peptide C was only observed in LC-MALDI analysis of the chicken MBM sediment sample and the LC-ESI analysis of the pig MBM sediment sample. Peptide D was

not observed in the LC-MALDI or LC-ESI of any MBM sediment sample. Because all of these peptides were observed in LC-MALDI and LC-ESI analyses of the bone samples (except Peptide D which was not observed in LC-MALDI or LC-ESI analysis of chicken bone), the lack of matches to Peptides C and D in the LC-MS analyses of the MBM sediment samples, as well as the lack of observation of any peptides of $m/z > 1290$ in the SPE analyses, could indicate extensive hydrolysis due to the rendering conditions.

It is interesting to note that of the three peptides capable of distinguishing between all four species (SGETGASGPPGFVGEK, GEPGPAGAVGPAGAVGPR and TGPPGPSGISGPPGPPGPAGK in cattle), the latter two of these are both complicated by the Ala to Ser replacement (from the cattle to pig sequences) and the Ser to Ala replacement (from the cattle to sheep sequences); the Ser to Ala or Ala to Ser replacement could be obscured in the mass spectral data by hydroxylation (or lack of) of nearby proline residues). However, these latter two peptides (GEPGPAGAVGPAGAVGPR and TGPPGPSGISGPPGPPGPAGK in cattle) were not observed for any of the four species of interest (cattle, sheep, pig and chicken) in the mixed sample (Table 9.9); only the peptide SGETGASGPPGFVGEK (in cattle) was present that is capable of distinguishing between cattle, sheep, pig and chicken.

Ideally, identification of a particular species within a mixed sample needs to be supported by a number of peptides and a level of confidence needs to be explored with a set of unknown samples before its application as a robust method of species identification. The major problems with relying on a number of different peptides, some of which can be conserved between the species of interest, are that the MBM samples could potentially be from a mixture of sources. In the one mixed sample that was analysed in this study, the dominance of sheep collagen was apparent with the likely contamination of pig collagen at a relatively low level. Using S/N values of the key peptides it was possible to speculate that sheep collagen was approximately 5 x more abundant than pig collagen. Following analysis and interpretation of the LC-MS data, the actual composition of the mixed MBM sample was revealed as being composed of 80% sheep, 15% pig, and 5% chicken MBM. This indicates that the use of relative peak intensity may have applications in estimating approximate

relative amounts, but improved by (more expensive) quantitative approaches such as iTRAQ (Wu *et al.* 2006).

Although six of the nine peptides considered useful for species discrimination among the mammals (Table 9.9) were observed, four were not observed. In the absence of the single species-specific peptide SGETGASGPPGFVGEK (in cattle), species identification using the remaining collagen-peptide markers described in Table 9.9 is much more difficult. For example, if it was known that only two species are present in a mixed sample, it may be possible to use a combination of particular 'semi-species-specific' collagen peptides, but when a third species may also be present, identification using these peptides may not be possible (e.g., a distinction of pig from sheep and cattle collagen cannot be made using a combination of pig/sheep and sheep/cattle peptide markers). Although this should not be as difficult when the third species is not mammalian (due to the much greater sequence variation), it was not possible to achieve identification of the 5% chicken contaminant in the mixed MBM sample analysed using the nine chosen peptide markers in Table 9.9, despite the match to the chicken α1 (I) sequence in the 'Collagens' search. However, the chicken α1 (I) peptide GAPGPPGAVGAAGK (m/z 1153.6) that was observed (and responsible for the protein match) was in addition to the sheep/pig peptide GVPGPPGAVGPAGK (m/z 1191.6) of the same position in the protein, where the relative peak intensity of the sheep/pig peptide was approximately 6 x greater than the chicken peptide. Despite the apparent accuracy (albeit with only one analysis of one mixed sample) of using relative peak intensity to estimate the relative amounts of pig contaminant in the predominantly sheep MBM sample, the estimate for the chicken contaminant was too high (estimated as 1/6 parts avian/mammal but was actually ~ 1/20 parts avian/mammalian). This indicated that the use of collagen (I) peptides for species composition analyses of mixed MBM samples may be possible, as observed in other studies (Old *et al.* 2005; Ono *et al.* 2006), but will require much further analysis of mixed MBM samples and that relative peak intensities should not be relied on without using appropriate quantitative methods.

Although the peptides potentially capable of distinguishing between relevant industrial species are described as being cattle/sheep, or pig-specific, the Mascot searches against the 'Collagens' database indicates that they also match several other species. For example, the

cattle/sheep peptide GETGPAGPAGPIGPVGAR also matches that of cat. Thus the species-specific collagen-peptide markers listed in Table 9.9 are solely for the analysis of samples limited to the species of economic interest, such as cattle, sheep, pig, and chicken, and may not be suited to the analysis of unknown palaeontological material.

9.6 Conclusions

The use of EST data clearly has potential for obtaining protein sequences from species lacking published information. The amalgamation of collagen (I) sequences, such as those proposed here for pig and sheep, is clearly advantageous for the LC-MS analyses, where the aim is to identity the species of origin of a particular sample. Knowledge of the collagen (I) sequences of all species potentially present is a great asset, as the improvements in the LC-MS peptide analyses for the pig and sheep MBM samples using the estimated sequences clearly showed. The two SPE-based collagen-peptide marker methods of isolating potentially diagnostic collagen peptides were not successful when applied to MBM sediment rendered at temperatures above 133°C. The absence of the carboxytelopeptide, one of the most useful collagen markers, as well as the helix Peptide C, was consistent with the LC-MS analyses. The potential for using LC-MS methods for the identification of multiple species in MBM samples was clearly shown with the successful identification of 15% pig collagen amongst 80% sheep collagen, however, the inability to observe 5% chicken collagen in this mixture suggests that this technique may not be as appropriate as ELISA or PCR methods capable of detection limits approaching 0.1%.

10 Collagen Peptide Analysis in Middle and Late Pleistocene Bones

10.1 Abstract

Selected peptides of the collagen α2 (I) chain were isolated from 40 archaeological and palaeontological bone samples dating back to the Cromerian (~650 ka) using SPE following two separate enzymatic digestions (one using bacterial collagenase; the other using trypsin) and analysed by MALDI-MS in order to investigate their 'archaeological' survival, particularly whether or not the collagen molecule is broken down via random hydrolysis along the protein chain or that the helical region is more protected than the non-helical telopeptides. This investigation was also supported by more in-depth LC-MS analyses of 12 selected ancient samples (six determined species; six morphologically indeterminable species) using two complementary methods: LC-ESI-qTOF-MS and LC-MALDI-TOF-MS. Although the long-term survival of ancient collagen is supported here by its identification in Cromerian (~650 ka) samples, analysis of the peptides present indicate that its survival is not adequately explained by the random cleavage hypothesis, but more consistent with the 'polymer-in-a-box' model of collagen stability. The limitations in species identification of using each of the proposed methods (SPE-MS or LC-MS) are more complex than initially anticipated and these approaches may require some prior knowledge of potential species to be successful. Amino acid concentrations and the extent of amino acid racemisation was also investigated with respect to the 'archaeological' survival of particular collagen peptides from 51 ancient samples.

10.2 Introduction

The alterations in the geography of Britain, notably its connection to the European mainland, arising as a result of climate-driven sea-level change, led to changes in the distribution of plants, animals and soils, with temperate woodland being driven out of Western Europe during glacial periods, the development of early man, and widespread migrations, colonisations and extinctions of plants and animals all related to environmental factors (Bradley 1985; Walker 2005). Species identification of animal bones contributes to

this understanding of environmental conditions and much research has focused on the British Quaternary period because of the extreme variations in climate history. Most well-preserved palaeontological bones come from cave deposits, primarily due to the exceptional preservation conditions that caves can offer. Cave deposits are also particularly informative as they can represent the cumulative collection of faunal remains via a number of different processes, mainly from habitation, natural traps, and carnivore accumulation. The fragmentation of bone via either trampling or carnivore activity often results in a complete loss of morphologically-diagnostic criteria for species identification, and thus large collections of unidentified material remain unable to confer environmental information. Although the complexity of this relationship cannot be understood by species identifications alone, there is potential to improve species lists (see Currant and Jacobi 2001) and thus current knowledge regarding the climatic history of the last 2 million years.

The aims of the research described in this chapter were to investigate the methods of collagen peptide isolation and analysis so far proposed in this thesis in application to British archaeological and palaeontological samples dating back to the Cromerian (~650 ka). Three methods of analysing collagen (I) peptides by protein mass spectrometry for purposes of species discrimination were developed during the work described in this thesis: an LC-MS peptide analysis approach (see section 4.3.2) analysing the complex mixtures of all peptides present in the digested 'collagen' samples, and two SPE-based collagen-peptide marker approaches specifically targeting the carboxytelopeptide (hereafter Peptide T) (see section 5.2), and four peptides from the helical region of the collagen molecule (I) (Peptides A-D) (see section 7.3.1). The SPE-based collagen-peptide marker approaches have the advantage that they are much simpler and quicker to carry out and less costly, but have the disadvantage that the information obtained is limited, being only useful to discriminate particular taxa. However, the two SPE-based collagen-peptide marker methods are complementary to each other in terms of the species-specific information they can produce. These SPE-based collagen-peptide marker methods are here carried out for the 40 ancient bone samples that yielded an acid-insoluble 'collagen' residue in order to assess the survival rates of collagen over long periods of time (see Table 10.1). The LC-MS approach has the advantage that much more information is obtained in each analysis, but the disadvantage that it is less repeatable, ~50-100 times more costly (per sample), and because of the

increased amount of data it is much more time-consuming to carry out and analyse, so that only 12 palaeontological samples were analysed in this way (using both LC-MS methods; LC-MALDI and LC-ESI). Although the survival of collagen (including its α2 (I) chain carboxytelopeptide) in samples devoid of aDNA was demonstrated previously (Chapter 5), these four approaches for collagen peptide analysis (the two SPE-based and two LC-MS-based methods) have not yet been compared to identify which methods are better suited to species discrimination in archaeological and palaeontological bone. Both types of SPE-based and LC-MS-based collagen-peptide analysis can potentially aid investigations into the diagenesis of the collagen molecule. For example, there have been several different models proposed for the break-down of collagen; one model by Collins *et al.* (1995) suggests that random cleavage along the peptide chain occurs, whereas the 'polymer-in-a-box' model (Miles and Ghelashvili 1999) suggests that the helical part of the collagen molecule is protected by the bone mineral, whereas the telopeptides are not. In the former scenario, we could assume that in progressively more degraded archaeological bone, the peptides observed in MS analyses would decrease in mass as the larger peptides are more likely to get hydrolysed than smaller peptides, and would not be influenced by their relative position in the molecule; in the latter scenario, the position (ie., within the helical region, or telopeptide) would be a key factor, where the telopeptides could be expected to hydrolyse (and be leached out of archaeological bones) much earlier than peptides from the 'protected' helical region. In this study, a total of 51 palaeontological specimens were analysed. All were analysed for amino acid concentrations and extent of racemisation. The 40 of these samples that yielded an acid-insoluble 'collagen' residue (using up to 1 gram of bone sample) were investigated by both SPE-based collagen-peptide marker methods (isolating Peptide T (see section 5.2) and Peptides A-D (see section 7.3.1)), and a subset of 12 of these specimens were further investigated by complementary LC-MS approaches (LC-MALDI and LC-ESI), along with six modern standards. LC-MS results were analysed using Mascot searches as described in section 4.3.3 and the nine most useful LC-ESI species-specific collagen-peptide markers identified in section 9.4.5 (Table 9.9) were also manually investigated to see if they differ between a wider taxonomic range of species than previously investigated.

10.3 Experimental

In order to cover a wide temporal range, samples from many different archaeological and palaeontological sites (see Fig. 10.1) were obtained, ranging from the Neolithic (Carsington Pasture Cave and Totty Pot), Late-Glacial (Etches Cave & Fox Hole Cave), Middle Devensian (Uphill Cave), Early Devensian (Steetley Wood Cave and Stump Cross Caverns), Ipswichian (Kirkdale Cave, Lavenham, Barrington, Barnwell Gravel Pit), to the Cromerian (Kessingland Forest Bed and West Runton Forest Bed) as well as numerous samples of uncertain age dredged from the North Sea. In addition to the 49 British samples, to further investigate species variability and the effects of sampling distantly-related unknown species on the interpretations of the results obtained by LC-MS, samples of dodo and giant tortoise from Mauritius were also analysed. A brief description of the samples and the methods by which they were each analysed are listed in Table 10.1.

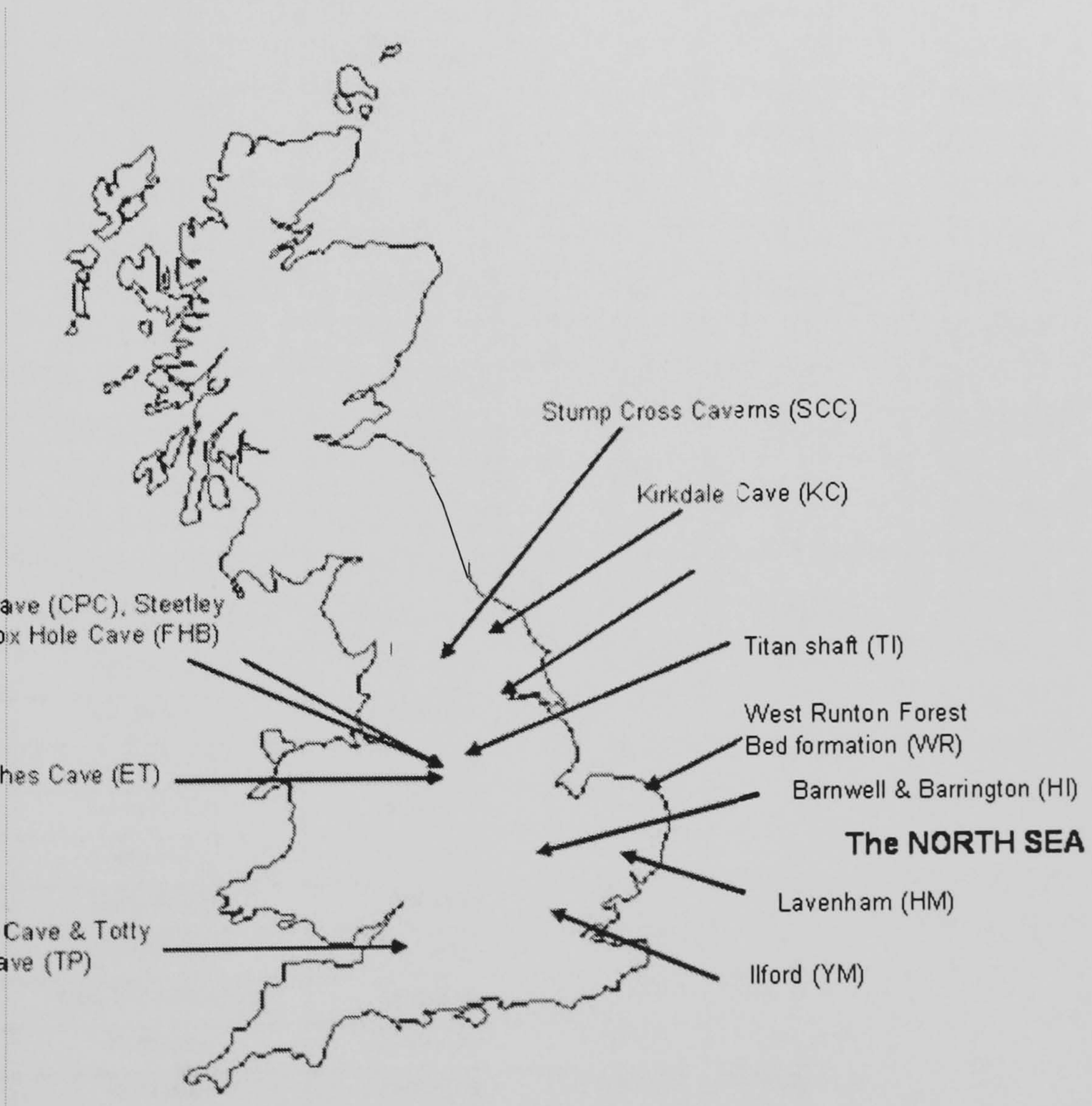


Figure 10.1 - Map of British sites of samples used in this research chapter.

Table 10.1 - Sample information including sample code, site, approximate age, species (if known) and the analytical approaches used on each sample, where AA = amino acid analysis (concentrations and racemisation), T = telopeptide marker, H = helix markers and P = LC-MS proteomics-based methods.

Sample	Site	Age (ka)	Age (ka)	Species	Methods
CPC1	Carsington Pasture Cave	Neolithic	4	Auroch	AA+T+H
CPC2	Carsington Pasture Cave	Neolithic	4	Caprine	AA+T+H
CPC3	Carsington Pasture Cave	Neolithic	4	Pig	AA+T+H
CPC4	Carsington Pasture Cave	Neolithic	4	Horse	AA+T+H
CPC5	Carsington Pasture Cave	Neolithic	4	Roe Deer	AA+T+H
GT	Mauritius	Holocene	4	Giant Tortoise	AA+T+H+P
DODO	Mauritius	Holocene	4	Dodo	AA+T+H+P
TP1	Totty Pot Cave	Holocene	~5	Aurochs	AA+T+H
FHB	Fox Hole	Late/Post Glavial	12	?	AA
ET1	Etches Cave	Mid-Late Devensian	28	Bear	AA+T+H
ET3	Etches Cave	Mid-Late Devensian	28	Bear	AA+T+H
ET4	Etches Cave	Mid-Late Devensian	28	Bear	AA+T+H
ET5	Etches Cave	Mid-Late Devensian	28	Smaller cat	AA+T+H
ET6	Etches Cave	Mid-Late Devensian	28	Large cat (Lynx)	AA+T+H
UP1	Uphill Cave	Mid Devensian	~30	Canid?	AA+T+H
UP3	Uphill Cave	Mid Devensian	~30	Bear	AA+T+H
TI1	Titan shaft	Mid Devensian	45.8	Bison?	AA+T+H
ST11	Steetely Wood Cave	Early Devensian	~66	Wolf?	AA+T+H
ST12	Steetely Wood Cave	Early Devensian	~66	Wolf	AA+T+H
ST16	Steetely Wood Cave	Early Devensian	~66	Reindeer	AA+T+H
ST18	Steetely Wood Cave	Early Devensian	~66	?	AA+T+H
ST23	Steetely Wood Cave	Early Devensian	~66	?	AA+T+H
ST33	Steetely Wood Cave	Early Devensian	~66	Bovid	AA+T+H+P
ST36	Steetely Wood Cave	Early Devensian	~66	?	AA+T+H
ST38	Steetely Wood Cave	Early Devensian	~66	?	AA+T+H
ST43	Steetely Wood Cave	Early Devensian	~66	?	AA+T+H
SCC1	Stump Cross Caverns	Early Devensian	79	Reindeer	AA+T+H
KC1	Kirkdale Cave	Ipswichian	114-135	Rhino?	AA+T+H
KC2	Kirkdale Cave	Ipswichian	114-135	?	AA+T+H
KC3	Kirkdale Cave	Ipswichian	114-135	Cattle	AA+T+H+P
KC4	Kirkdale Cave	Ipswichian	114-135	Cattle	AA+T+H+P
KC5	Kirkdale Cave	Ipswichian	114-135	Bison	AA+T+H+P
KC6	Kirkdale Cave	Ipswichian	114-135	Bovine	AA+T+H
KC7	Kirkdale Cave	Ipswichian	114-135	Bovine	AA+T+H
HI4	Barnwell Gravel Pit	Ipswichian	114-135	Hippopotamus	AA
HI5	Barrington	Ipswichian	114-135	Hippopotamus	AA
HM1	Lavenham	Ipswichian	114-135	Hippopotamus	AA
HM2	Lavenham	Ipswichian	114-135	Hippopotamus	AA
YM2	Ilford	Post-Hoxnian	200	Bovid?	AA
WR5	West Runton Forest Bed	Cromerian	~650	Deer	AA+T+H+P
WR6	West Runton Forest Bed	Cromerian	~650	Large mammal	AA

Table 10.1 continued

Sample	Site	Age	Age	Species	Methods
WR7	West Runton Forest	Cromerian	~650	?	AA
WR8	West Runton Forest	Cromerian	~650	?	AA+T+H+P
WR9	West Runton Forest	Cromerian	~650	?	AA+T+H+P
WR12	West Runton Forest	Cromerian	~650	Artiodactyl	AA
WR13	West Runton Forest	Cromerian	~650	Large mammal	AA
WR14	West Runton Forest	Cromerian	~650	Large mammal	AA+T+H+P
WRE2	West Runton Forest	Cromerian	~650	Steppe Mammoth	AA+T+H+P
KS1	Kessingland	Cromerian	~650	Cervid?	AA
DR1	North Sea	?	?	Bovid?	AA+T+H+P
DR2	North Sea	?	?	?	AA+T+H

10.3.1 Preparation of 'Collagen' and the Isolation and Analysis of Collagen-Peptide Markers

The archaeological and palaeontological samples were supplied as a powder and the collagen α2 (I) chain carboxytelopeptide fragments (Peptide T) and helix peptide markers (Peptides 26%A-D and 32%C) were obtained following the methods described in section 5.2 and section 7.3.1 respectively (see Fig. 10.2), and S/N values are given for each.

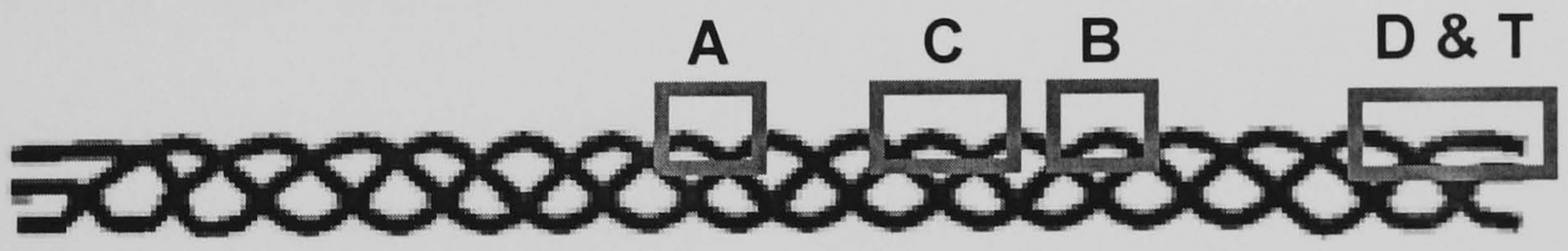


Figure 10.2 - Approximate positions of collagen-peptide markers in the tropocollagen molecule (carboxytelopeptide, labeled T, helix peptides labeled A-D).

Note that Peptide D, from gelatinisation of the 'collagen' residue and digestion with trypsin includes the amino-terminal 17 amino acids of Peptide T (the entire carboxytelopeptide sequence minus the carboxy-terminal alanine), obtained from digestion of the 'collagen' residue with bacterial collagenase.

10.3.2 LC-MALDI-MS and LC-ESI-MS Analyses

Both LC-MALDI-MS and LC-ESI-MS analyses were carried out on trypsin digested 'collagen' samples from selected palaeontological specimens as described (section 4.3.2 and

section 9.3.2 respectively). Database searches were carried out against the UniProt and the local 'Collagens' databases (section 9.3.3) but using Mascot's standard scoring instead of Mascot's MudPIT scoring. Bone samples of six modern bones (cattle, pig, sheep, goat, Indian elephant and chicken), six ancient samples of determined species and six archaeological samples of indeterminate species were analysed by LC-MALDI-MS and LC-ESI-MS (the LC-MALDI-MS analyses of modern cattle, pig, sheep and chicken are from Chapter 4 and the LC-ESI-MS analyses of modern cattle, pig, sheep and chicken from Chapter 9) in order to compare the ability to discrimate between known and unknown species to the previously described SPE-based approaches. The ancient samples of morphologically-determined species included cattle (KC3 and KC4), bison (KC5), steppe mammoth (WRE2), dodo and giant tortoise. The latter two had previously failed aDNA screening (Shapiro *et al.* unpublished data) following the method of Shapiro *et al.* (2004). The six 'indeterminate' specimens included a large deer metapodial (WR5), two very small fragments (WR8 and WR9) and a large mammal (WR14) from West Runton, a bovid (ST33) from early devensian Steetley Wood Cave and a bovid (DR1) dredged from the North Sea.

10.3.3 Amino Acid Analyses

Amino acid composition and racemisation analyses were also carried out on the 51 archaeological bone samples as described (section 2.2.4).

10.4 Results

10.4.1 Collagen (I) Peptide Markers in Palaeontological Samples Using SPE-MS-Based Approaches

Following the demineralisation of <1 gram bone mineral from 51 ancient samples, 40 yielded an acid-insoluble 'collagen' residue. The collagen residues from these 40 ancient samples were analysed for the survival of the collagen α2 (I) chain carboxytelopeptide (Peptide T) and particular peptides from the helical region of the tropocollagen molecule (Peptides A-D). The S/N value of the intact carboxytelopeptide (from digestion with bacterial collagenase) and the four 'helix' peptide (from gelatinisation and digestion with trypsin) peaks (Peptides 26%A-D and Peptide 32%C; see section 7.3) are listed collectively

in Table 10.2. The actual m/z values for each of these species-specific peptide markers, as well as for the lower mass fragments of the carboxytelopeptide are presented in Appendix 6.3. An example of spectra quality is shown in Figure 10.3 and the resulting S/N values are presented in Table 10.2 (see Appendix 6.3 for summary tables including the m/z value and S/N value for each species-specific collagen-peptide marker peak).

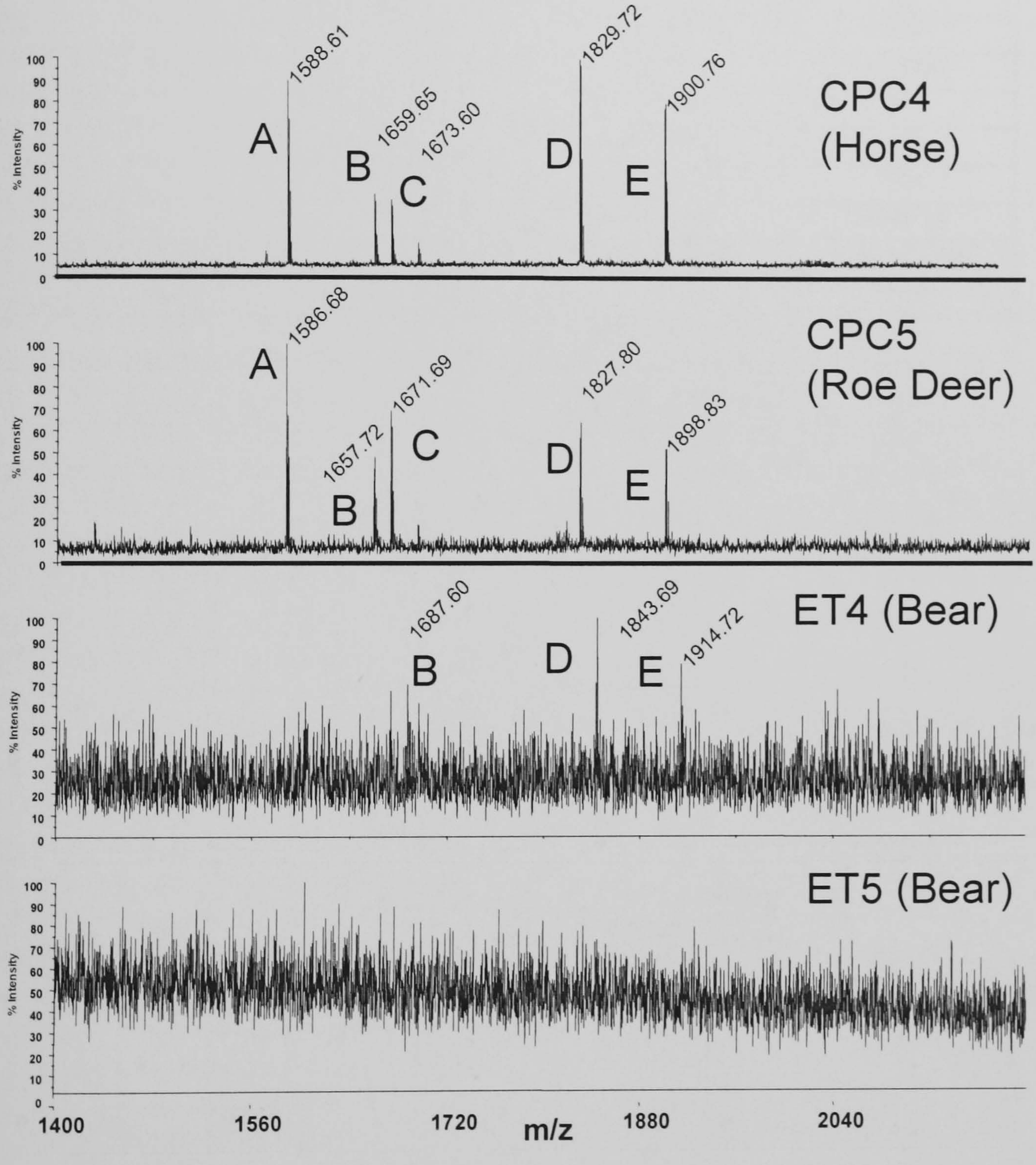


Figure 10.3 - Decreasing quality of carboxytelopeptide peak S/N in MALDI-MS reflectron mode spectra: (from top to bottom) a spectrum showing all expected carboxytelopeptide peaks of high quality S/N from analysis of CPC4, a spectrum showing all expected peaks with a small amount of noise from analysis of CPC5, a spectrum showing some of the expected peaks but at poor S/N ratios from analysis of ET4 and a spectrum showing no peaks observed from the analysis of ET1. Telopeptide fragments are labeled A-E as described in section 5.2.

Table 10.2 - S/N of SPE-isolated peptides and the extent of species information that can be extrapolated from their m/z values in comparison to their morphological identifications. Species markers predominantly derive from Chapters 5-7 but horse, roe deer, bear and reindeer peptide m/z values were determined from samples analysed herein. X indicates that the peptide was not observed.

Sample	S/N of Peptide Marker						Peptide Marker Species Identification	Morphological Species Identification
	26%	26%B	26%C	26%D	32%C	T		
CPC1	X	X	X	X	X	5.11	Bovine	Aurochs
CPC2	X	X	7.49	X	41.72	52.29	Sheep/Red Deer	Caprine
CPC3	30.49	X	X	X	35.09	25.15	Pig	Pig
CPC4	X	34.92	10.15	X	30.39	90.48	Horse	Horse
CPC5	X	20.34	21.06	X	40.90	14.26	Roe Deer	Roe Deer
ET1	144.6	50.11	6.54	X	12.12	X	Bear	Bear
ET3	19.52	X	10.87	X	7.88	17.79	Bear/Dog?	Bear
ET4	X	X	X	X	8.99	8.01	Bear/Dog?	Bear
ET5	10.67	X	X	X	X	X	Pig/Dog	Smaller cat
ET6	282.7	174.74	9.07	X	9.26	X	Hare/Bear/Dog?	Large cat (Lynx)
KC1	72.3	266.77	15.2	X	16.87	X	?	Rhino?
KC2	X	249.72	95.25	X	183.18	X	Bovine/Sheep/Red Deer	?
KC3	X	480.24	80.35	X	253.38	X	Bovine/Sheep/Red Deer	Cattle
KC4	X	412.16	82.77	X	204.48	X	Bovine/Sheep/Red Deer	Cattle
KC5	X	12.64	15.35	X	62.54	X	Bovine/Sheep/Red Deer	Bison
KC6	55.63	28.28	10.82	X	148.71	X	Bovine/Sheep/Red Deer	Bovine
KC7	X	X	X	X	158.55	X	Bovine/Sheep/Red Deer	Bovine
WR14	X	32.18	21.00	X	X	X	Bovine/Sheep/Red Deer	Large mammal
WR5	12.87	27.00	6.01	X	42.91	X	Bovine/Sheep/Red Deer	Deer
WR8	X	10.93	X	X	7.87	X	Bovine/Sheep/Red Deer	?
WR9	X	X	X	X	9.58	X	Goat/Reindeer	?
DR2	11.81	19.36	7.91	X	10.66	11.25	Bovine	?
SCC1	X	234.78	59.14	X	54.82	X	Goat	Reindeer
ST11	96.05	54.37	22.19	X	X	X	Canid	Wolf?
ST12	37.55	379.01	219.39	X	224.33	X	Canid	Wolf
ST16	X	16.36	13.87	X	22.81	X	Goat	Reindeer
ST18	21.28	9.13	X	X	19.50	X	Bovine/Sheep/Red Deer	?
ST23	7.51	6.28	X	X	89.19	X	Bovine/Sheep/Red Deer	?
ST33	X	201.14	62.71	X	73.12	6.19	Bovine	Bovid
ST36	91.06	21.78	11.55	X	8.85	X	Bear	Artiodactyl
ST38	75.41	X	5.16	X	62.83	X	Hare/Badger/Bear	?
ST43	183.2		7.3	X	25.36	X	Goat/Reindeer	?
TI1	143.2		14.44	X	27.47	X	Sheep/Red Deer	Bison?
TP1	27.85	17.56	X	X	24.94	7.55	Bovine	Aurochs
UP1	6.6	X	X	X	5.43	X	Fox/Dog	Canid?
UP3	11.68	8.14	X	X	18.42	X	Badger/Bear?	Bear
DR1	5.37	34.07	6.38	X	19.57	35.38	Sheep/Red Deer	Bovid?
WRE2	X	X	X	X	11.13	X	Many mammals	Mammoth
Dodo	98.33	48.49	23.78	X	53.31	46.83	Unique avian	Dodo
GTo	85.37	X	29.44	X	63.55	33.78	Not-mammal or avian	Giant Tortoise

The spectra obtained from all five of the Carsington Pasture Cave samples contained a peak for the intact carboxytelopeptide. Only the spectra from two of the Etches Cave samples contained an intact carboxytelopeptide peak, although peaks for the intact carboxytelopeptide were unexpectedly not observed in some of the other Etches Cave samples. This was unexpected because they are post-glacial and thus relatively young in terms of expected collagen survival. Of the spectra from the ten early Devensian samples from Steetley Wood Quarry and Stump Cross Caverns, peaks for the intact carboxytelopeptide were only observed in one sample (sample ST33). Of the spectra from the three late Devensian samples, only the Titan shaft sample (sample TI1) spectra contained carboxytelopeptide peaks, whereas the spectra from both of the Uphill Cave samples did not have any peaks present. The spectra from all seven of the Ipswichian Kirkdale Cave samples and all five Cromerian West Runton samples also failed to show any peaks for the carboxytelopeptide.

It is interesting to note that the carboxytelopeptide-containing helix Peptide D was not observed in any of the spectra from any of the ancient (archaeological or palaeontological) samples. Of the spectra from the five Carsington Pasture Cave samples, Peptide 32%C was observed in all but one (CPC1) of the samples, Peptide 26%A was only observed in the spectrum of one of the Carsington Pasture Cave samples (CPC3), Peptide 26%B in spectra of two of the five samples (CPC4 and CPC5). Of the spectra from the five Etches Cave samples, Peptide 32%C was observed in four of the samples with only the spectra from the ET4 sample lacking signals for Peptides A-C. In the spectra from the early Devensian Steetley Wood Quarry samples and the late Devensian Uphill Cave samples, all 10 samples yielded at least two of the four targeted peptides (Peptides 26%A-D) and all except ST11 showed a signal for Peptide 32%C. Of the spectra from the seven Ipswichian Kirkdale Cave samples, most yielded three of the four targeted peptides (not including Peptide 26%D which is absent throughout); only the spectra from samples KC5 and KC7 lacked Peptide 26%A peaks, and none of the Peptide 26%A-C peaks were observed in the spectrum of the KC7 sample. The spectra of the five Cromerian samples showed signals for Peptide 32%C, but were less successful with the 26% ACN SPE fraction peaks, as the spectrum of the WR9 sample lacked signals for Peptides 26%A, B and C, the spectrum of the WR8 sample lacked

signals for Peptides 26%A and C, and the spectrum of the WR14 sample lacked a signal for Peptide 26%A.

10.4.2 LC-MS Analyses

10.4.2.1 Protein Identification in Palaeontological 'Collagen' Digests

In order to study the potential survival of other non-collagenous proteins, LC-MS data was obtained for 12 of the 51 ancient samples (see Table 10.1) and searched against the public UniProt protein database and the top protein matches compared to the LC-MS results obtained from the six modern bone samples (Table 10.3). The deconvoluted data files are included in Appendix 6.1 and the Mascot search results in Appendix 6.2.

Table 10.3 - Protein Identifications from Mascot searches of LC-MS data searched against UniProt public database.Data for modern cattle, pig sheep and chicken are from analyses in Chapter 4.

Specimen	LC-MALDI protein matches	LC-ESI protein matches
Modern Cattle	Collagen (I) α1, α2, A2HS, Biglycan, OC, SPARC, Prothrombin.	Collagen α2(I), α1(I), α1(XI), α1(V), A2HS, Biglycan
Modern Pig	Collagen (I) α1, α2, (II) α1, (XI) α1 & (III) α1, SPARC, Albumin.	Collagen α1(I), α2(I), α1(II), α1(III)
Modern Sheep	Collagen (I) α1, α2, A2HS, Collagen (II) α1, Collagen (V) α2, Biglycan, Collagen (VII) α1, Albumin, SPARC	Collagen α1(I), α2(I), A2HS, Collagen α1(II), α1(III), α2(V), Biglycan, Collagen α1(V)
Modern Goat	Collagen α1(I), α2(I), α1(IV), α1(III), α1(XII), Titin	Collagen α1(I), α2(I), A2HS, Collagen α1(II), α1(III), α1(IX), α1(V)
Modern Indian Elephant	Collagen (I) α1, α2, Collagen (V) α2	Collagen α1(I), α2(I), α1(III)
Modern Chicken	Collagen α1(I), α2(I), α1(II), α1(III), Haemoglobin, Apolipoprotein, Collagen α2(V)	Collagen α1(I), α2(I), α1(II), α1(III), Apolipoprotein, Albumin
KC3	Collagen α1(I), α2(I)	Collagen α1(I), α2(I), α1(II), α1(IV)
KC4	Collagen α2(I), α1(I)	Collagen α1(I), α2(I), α1(II), α1(VII), α1(III), α2(IV), α1(IX)
KC5	Collagen (I), α2, α1	Collagen α1(I), α2(I), α1(XI), α1(VII), α2(IX)
ST33	Collagen α2(I), α1(I)	Collagen α1(I), α2(I), α1(III), α1(VII)
WR5	Collagen (I) α1, α2, (II) α1	Collagen α1(I), α2(I), (II) α1, α1(V), α1(III)
WR8	Collagen α1(I), α2(I)	Collagen α1(I), α2(I), α1(II), α2(IX)
WR9	Collagen α1(I), α2(I), α1(II)	Collagen α1(I), α2(I), α1(II), α1(V), α1(IX), α1(VII)
WR14	Collagen α1(I), α2(I), α1(II)	Collagen α1(I), α2(I), α1(II), α2(IV), α1(V)
WRE2	Collagen α1(I), α2(I)	Collagen α1(I), α2(I), α1(III)
DR1	Collagen α2(I), α1(I)	Collagen α1(I), α2(I), α1(II)
Giant Tortoise	Collagen α1(I), α2(I), α1(II)	Collagen α1(I), α2(I), α1(II), α1(V), α1(III)
Dodo	Collagen α1(I), α2(I), α1(V)	Collagen α1(I), α2(I), α1(II), α2(XI), α2(IV), α1(III), α2(V), α1(IV), α1(XI)

Collagen (I) is the dominant protein identified in all 36 LC-MS analyses (LC-MALDI and LC-ESI; 18 samples). Numerous NCPs were matched in most of the modern specimens (all except goat and Indian elephant) including proteins such as alpha-2-HS-glycoprotein, albumin, SPARC, OC and biglycan as discussed previously in section 4.4.1. Interestingly, none of the ancient samples exhibit peptide matches to any of these NCPs, suggesting that the globular NCPs do not survive in the burial environment as well as the fibrous collagens. Unexpectedly, several other collagen types were also identified in both modern samples (types II, III, V, VII, IX, and XI) and ancient samples (types II, III, IV, V, VII, IX, and XI). There appears to be more of these other collagen types identified in the LC-ESI analyses than the LC-MALDI-MS analyses, but this is likely a result of the increased number of peptides searched in LC-ESI-MS analyses. It is plausible that the matches to these other collagen types do not reflect accurate identifications, but more likely misidentifications of similar proteins (from the same protein (collagen) family) due to either poor quality spectra (note the unusual fragmentation of proline-containing peptides upon CID described in section 9.4.4) or limited sequence information present in the UniProt database and the high levels of similarity between these proteins (as described in section 4.4.1).

10.4.2.2 Sequence Coverages Obtained by LC-MS

The acid insoluble 'collagen' from each of the 18 samples was gelatinised and digested with trypsin for LC-MALDI and LC-ESI analyses. Text files containing peak mass and intensity values were obtained for the resulting product ion tandem mass spectra of each sample (presented in Appendix 6.1) and searched against the UniProt database by Mascot (Appendix 6.2). Collagen (I) sequence coverages of the α1 and α2 chains for several known and indeterminate species from Mascot searches against the 'Collagens' database (for sequences included in the 'Collagens' database see Appendix 5.1), with Mascot search results are presented in Appendix 6.2. The species-identity of the collagen sequences matched and their respective percentage sequence coverages, along with the number of unmatched peptides for all samples analysed by LC-MS are shown in Table 10.4 (see Appendix 6.2 for detailed search results).

Table 10.4 - LC-MS Mascot search results against the 'Collagens' database from LC-MALDI and LC-ESI of 12 ancient samples in comparison to modern cattle, sheep, pig, goat, elephant and chicken.

Sample	Species Identified in LC-MALDI (α1/α2)	α1/α2 sequence coverage (%)	Species Identified in LC-ESI (α1/ α2)	α1/α2 sequence coverage (%)
Modern Cattle	Cattle/Cattle	61/69	Cattle/Cattle	82/86
Modern Sheep	Sheep/Sheep	61/75	Sheep/Sheep	86/87
Modern Goat	Sheep/Sheep	49/56	Sheep/Sheep	84/84
Modern Pig	Pig/Pig	61/58	Pig/Pig	77/79
Modern Indian Elephant	Elephant/Elephant	50/57	Elephant/Elephant	68/63
Modern Chicken	Chicken/Chicken	54/50	Chicken/Chicken	70/60
Dodo	Chicken/Chicken	37/20	Chicken/Chicken	64/55
Giant Tortoise	Chicken/Dog	23/11	Chicken/Dog	50/25
ST33	Cattle/Cattle	18/19	Cattle/Cattle	63/65
KC3	Sheep/Cattle	37/31	Cattle/Cattle	76/65
KC4	Cattle/Cattle	17/23	Cattle/Cattle	75/73
KC5	Cattle/Cattle	20/18	Cattle/Cattle	67/67
WR5	Sheep/Cattle	25/13	Sheep/Sheep	77/62
WR8	Sheep/Cattle	33/19	Sheep/Sheep	66/51
WR9	Sheep/Cattle	24/24	Cattle/Cattle	83/64
WR14	Elephant/Elephant	33/30	Elephant/Elephant	55/53
WRE2	Elephant/Elephant	30/24	Elephant/Elephant	55/55
DR1	Horse/Sheep	5/11	Sheep/Sheep	72/58

As expected, the top match for each collagen (I) alpha chain of the modern cattle, pig, sheep, elephant and chicken samples were cattle, pig, sheep, elephant and chicken respectively (Table 10.4). The goat sample was identified as sheep, as expected, because there is no more closely-related species' sequence present in the 'Collagens' database (at the time of the search). Likewise, KC5, the only known bison sample was also matched to cattle for both alpha chains throughout both LC-MS analyses (as did the Ipwsichian cattle samples KC3 and KC4, except where KC3 α1 (I) chain matched sheep), most likely also due to the lack of a more closely-related species' sequence. As expected, the steppe mammoth from the West Runton forest bed (WRE2) closely matched the only elephantid sequence present in the Collagens database. Interestingly, another Cromerian sample (WR14) of limited species

information (large mammal?) was also closely matched with the elephant sequence for both chains. The Cromerian deer sample (WR5) was identified as sheep/cattle (α1/α2) for LC-MALDI and sheep/sheep (α1/α2) for LC-ESI analyses. The other two unidentified Cromerian samples, WR8 and WR9 were both identified as sheep/cattle in all LC-MALDI searches and sheep/sheep (α1/α2) and cattle/cattle (α1/α2) in both LC-ESI searches respectively. Because of their similarity to WR5, they could both be possible deer specimens, but further research to confirm this would be required (although see section 10.4.2.3). The added comparative samples of non-mammalian archaeological species, dodo and giant tortoise, were matched with chicken/chicken (α1/α2) and chicken/dog (α1/α2) respectively throughout all analyses. Although the matches for dodo collagen to chicken collagen make sense as there is no more closely-related species' sequence present in the database, the giant tortoise matches to chicken and dog do not make sense because there are other more closely-related species's sequences present in the database, namely that of the green anole.

The collagen sequence coverages from LC-MALDI and collagen sequence coverages from LC-ESI obtained from each sample was plotted as a bar chart (Fig. 10.4). The relative amounts of α1:α2 chain sequence coverage did not appear to be consistent (Fig. 10.4) where some of the archaeologically younger specimens have resulted in less sequence coverage than some of the older specimens, even within specimens of the same taxonomic family (e.g. the bovids).

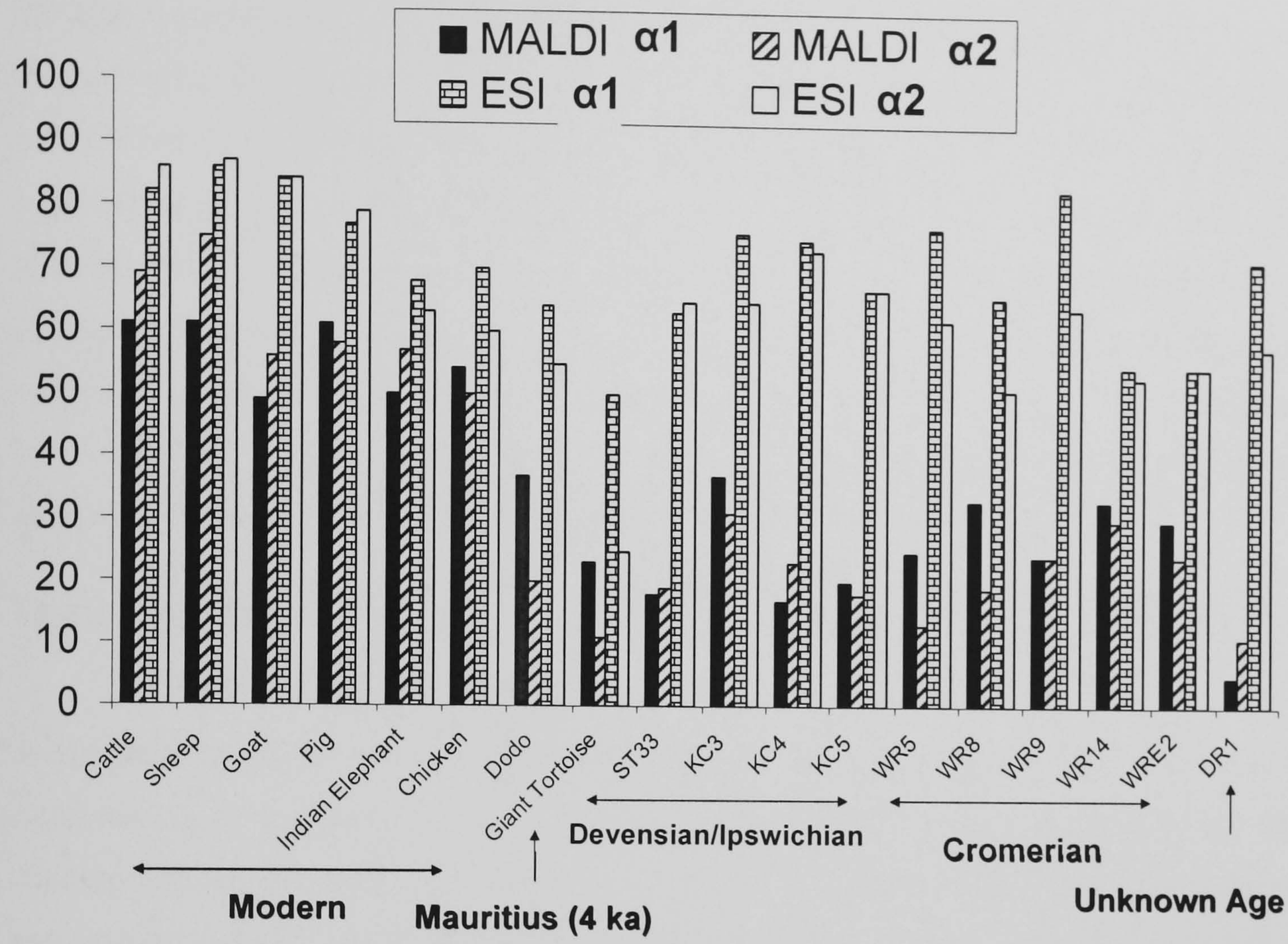


Figure 10.4 - Bar chart showing a comparison between LC-MALDI and LC-ESI data for both a1 (I) and a2 (I) chain sequence coverages (Appendix 6.5) using Mascot searches against the 'Collagens' database.

Although LC-ESI analyses resulted in slightly improved sequence coverages for both alpha chains in each of the modern samples, it resulted in much greater collagen (I) sequences coverages in all of the ancient samples (Fig. 10.4). This difference was even observed in palaeontological specimens of known species's sequence, such as KC3 and KC4. The large increase could be in part due to the greater number of peptides analysed by LC-ESI, which could adequately analyse peptides of m/z <900 (LC-MALDI analyses were restricted at this lower m/z range due to the potential presence of matrix peaks). The observed increase in sequence coverage does not appear to be only due to differences in the types of peptides that were matched in both techniques because not only are many more lysine-terminated peptides and small (<10 amino acids long) peptides observed in the LC-ESI and not in the LC-MALDI analyses (as in section 9.4.3.2 and references therein), but many of the arginine-terminated and larger (>10 amino acids long) peptides are also observed throughout all of the LC-ESI analyses of the ancient samples that are not observed in the LC-MALDI analyses. Although this implies that it is unlikely that diagenesis is selectively affecting

MALDI-type peptides such as the shorter peptides and arginine-terminated peptides, there is the possibility that diagenesis is affecting more hydrophilic peptides, which are favoured by MALDI than the hydrophobic peptides that are favoured by ESI. However, the greater sequence coverage observed for the less hydrophobic α1 (I) chain over the more hydrophobic α2 (I) chain sequence does not support this speculation. Because this effect of decreasing sequence coverage was not observed with the LC-ESI analyses of modern bone samples, its distinct observation in the LC-MALDI analyses of ancient samples could be due to suppression of the signal in owing to the presence of other components of the diagenetically altered collagen, such as high molecular weight humic acids.

10.4.2.3 LC-ESI Species-Specific Collagen-Peptide Markers in Palaeontological Samples

In further investigation of the collagen peptide peaks observed in the LC-ESI analyses of the palaeontological samples, the matches to the nine particular collagen peptides that were observed in modern cattle, sheep, pig and chicken bone samples following LC-ESI analyses and selected as being capable of differentiating between these four species (see section 9.4.5; Table 9.10) are shown in Tables 10.5 and 10.6.

Table 10.5 – Peptide sequences of nine selected collagen peptides observed in the LC-ESI analysis of palaeontological artiodactyl samples in comparison to modern cattle, sheep, goat and pig samples. Unidentified peptides are marked with a '?'.

Protein Position	Cattle, KC3, KC4, KC5 & ST33 α1 sequence	WR5, WR8, WR9 & DR1 α1 sequence	Sheep & Goat α1 sequence	Pig α1 sequence
621-635	GEAGPSGPAGPTGAR	GETGPSGPAGPTGAR		
876-901	GETGPAGPPGAPGAPGAPGPVGPAGK			GESGPAGPPGAPGAPGAPGPVGPAGK
906-923	GETGPAGPAGPIGPVGAR			GETGPAGPAGPVGPVGAR
Protein Position	**Cattle, KC3, KC4, KC5 & ST33 α2 sequence**	**WR5, WR8, WR9 & DR1 α2 sequence**	**Sheep & Goat α2 sequence**	**Pig α2 sequence**
247-261	GIPGPVGAAGATGAR			GIPGPAGAAGATGAR
595-612	GAPGAIGAPGPAGANGDR		GAPGAVGAPGPAGANGDR	
713-734	TGPPGPSGISGPPGPPGPAGK		TGPPGPAGISGPPGPPGPAGK	IGPPGPSGISGPPGPPGPAGK
750-765	SGETGASGPPGFVGEK		TGEPGAAGPPGFVGEK	TGETGASGPPGFAGEK
898-915	GEPGPAGAVGPAGAVGPR		GEPGPVGAVGPAGAVGPR	GEPGPAGSVGPAGAVGPR
987-999	IGQPGAVGPAGIR	TGQPGAVGPAGIR		

Table 10.6 - Peptide sequences of nine selected collagen peptides observed in the LC-ESI analysis of palaeontological elephantid, giant tortoise, chicken and dodo samples

Protein Position	WRE2 α1 sequence	Indian Elephant α1 sequence	WR14 α1 sequence	Giant Tortoise α1 sequence	Dodo α1 sequence	Chicken α1 sequence
621-635	GEAGPSGPAGPTGAR				GEAGPPGPAGPTGAR	
876-901	GETGPAGPPGAPGAPGAPGPVGPAGK				GESGPAGPPGAPGAPGAPGPVGPAGK	
906-923	GETGPAGPAGPAGPAGVR			?	GETGPAGPAGPPGPAGAR	

Protein Position	WRE2 α2 sequence	Indian Elephant α2 sequence	WR14 α2 sequence	Giant Tortoise α2 sequence	Dodo α2 sequence	Chicken α2 sequence
247-261	GIPGPVGAAGATGAR			GIPGPPGPAGPSGAR	GIPGPAGAAGATGAR	
595-612	GAPGAVGAPGPAGATGDR			GAPGAMGAPGPAGATGD RGEAGPAGPAGPAGPR	GLPGAIGAPGPAGGAGDR	
713-734	TGPPGPAGITGPPGPPGAAGK			?	VGPPGPAGITGPPGPPGPAGK	
750-765	TGETGASGPPGFAGEK			?	TGEQGIAGPPGFVGEK	
898-915	GEPGPAGSVGPVGAVGPR			?	?	GDPGPVGPVGPA GAFGPR
987-999	SGHPGAVGPAGVR			?	TGQPGAVGPAGIR	

Although the goat sample could not be distinguished from the sheep sample using these nine selected LC-isolated peptides, the large deer specimen (WR5) exhibited a combination of peptides that matched (from the Mascot search results against the 'Collagens' database) either cattle or sheep/goat/pig peptides (Table 10.5), which implies that although each individual peptide marker is not species-specific, the combination of peptide markers could be species-specific. Also matching the same set of peptides were samples WR8, WR9 and DR1. WR8, identified by SPE-isolated peptide markers as being cattle/bison/sheep/red deer, could be presumed to be red deer according to the LC-MS peptide markers that indicate that it is not from cattle, bison or sheep. Similarly, the WR9 sample was identified as goat/reindeer by SPE-isolated peptide markers and could be presumably reindeer as the LC-MS biomarkers rule out the possibility of it being from a goat (Table 10.5). DR1, identified as sheep/red deer by the SPE-biomarkers is potentially from red deer due to the LC-MS peptide markers because the peptides present are distinguishable from the known sheep peptides (Table 10.5). However, these identifications are limited to the species analysed throughout this thesis and so a greater understanding of potential species present at the archaeological sites being analysed is clearly needed (for example, if fallow deer or giant Irish elk could also be present, then it is inappropriate to simply label an identification as red

deer). In combination, these peptides indicate that deer species (such as red deer) that could not be distinguished using the SPE-isolated peptide marker approaches, can be distinguished from other artiodactyl species using collagen peptides observed in LC-MS analyses. The nine LC-MS peptide markers for the ST33 sample matches the cattle peptide markers, but the bison (KC5) sample could not be distinguished from the cattle (modern cattle bone, KC3 & KC4) samples (Table 10.5). Thus ST33 can only be identified as cattle/bison (herein 'bovine'). The data from the known elephantid specimens (modern Indian elephant bone and WRE2) had identical peptide sequences for the peptides that could be identified. The nine selected collagen-peptide markers from WR14, the previously unknown sample identified as elephantid from the LC-MS Mascot search, are identical to those of the modern Indian elephant and WRE2 (the West Runton steppe mammoth. Two of the α1 (I) chain peptides matched the elephantid and cattle peptide markers and one matched only elephantid sequence. All six α2 (I) chain peptide markers were matched to elephantid peptide sequences. In the giant tortoise sample, two of the α1 (I) chain peptides matched the elephantid and cattle peptide markers and one was not identified in the Mascot search, whereas three of the α2 (I) chain peptide markers were matched to unique sequences (from other species' collagen sequences in the Collagens database) and three could not be identified. All three of the α1 (I) chain peptides of the dodo specimen matched the chicken collagen-peptide markers, whereas three of the α2 (I) chain peptide markers matched the chicken peptide markers and three could not be confirmed from the Mascot search results.

10.4.3 Amino Acid Analyses

Amino acid concentration and racemisation analyses were carried out on all 51 ancient samples, including those that did not yield an acid-insoluble residue on demineralisation of up to 1 g of starting material. The overall amino acid compositions of the samples are shown (Fig. 10.5) where samples are separated into generalised archaeological age categories. The relative amino acid amounts were plotted as principal components in order to investigate slight differences in amino acid composition that could be observed between archaeological and palaeontological samples from different sites of different ages (Fig. 10.6). In order to make comparisons with the amino acid composition data of Hare (1980), where modern

bone contained about 2500 nmol/mg amino acids and below about 50 nmol/mg of fossil bone, the amino acid composition is considered to be unlike collagen (see section 1.4.3), the glycine concentrations are listed for each sample (Appendix 6.4). Asx D/L values are plotted against Asx concentration (Fig. 10.7; also listed in Appendix 6.4) (Collins and Riley 2000), where values for Asx D/L >0.14 indicate poor survival of mineral-binding molecules such as DNA and OC (see Chapter 2). Ala D/L values are also presented in Appendix 6.4 for comparison to Asx D/L values in order to evaluate possible contamination as described by Bada *et al.* (1973), where if the more slowly racemising amino acid (Ala) is observed to be more racemised than the faster racemising amino acids (Asx), contamination can be presumed present in the sample.

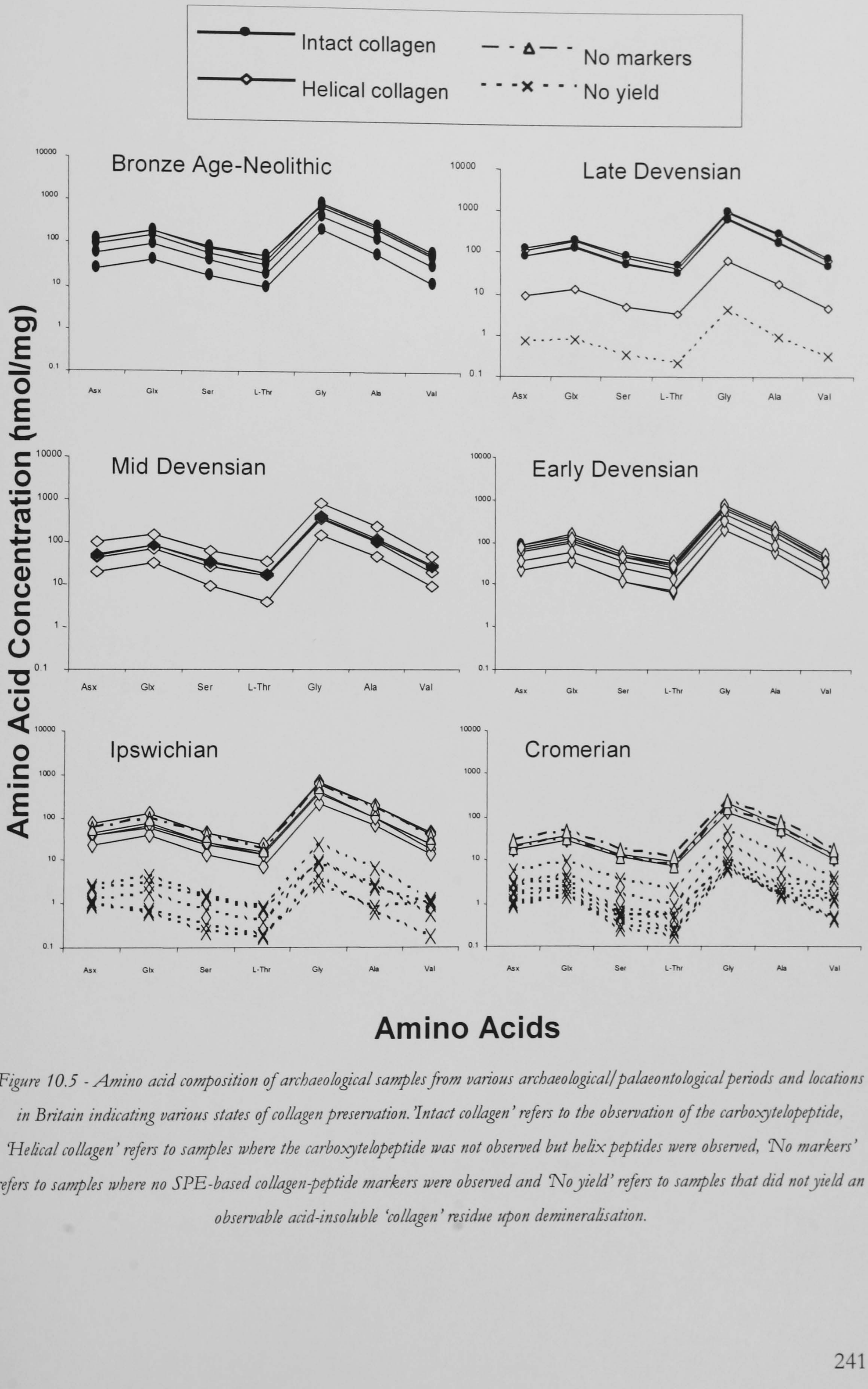


Figure 10.5 - Amino acid composition of archaeological samples from various archaeological/palaeontological periods and locations in Britain indicating various states of collagen preservation. 'Intact collagen' refers to the observation of the carboxytelopeptide, 'Helical collagen' refers to samples where the carboxytelopeptide was not observed but helix peptides were observed, 'No markers' refers to samples where no SPE-based collagen-peptide markers were observed and 'No yield' refers to samples that did not yield an observable acid-insoluble 'collagen' residue upon demineralisation.

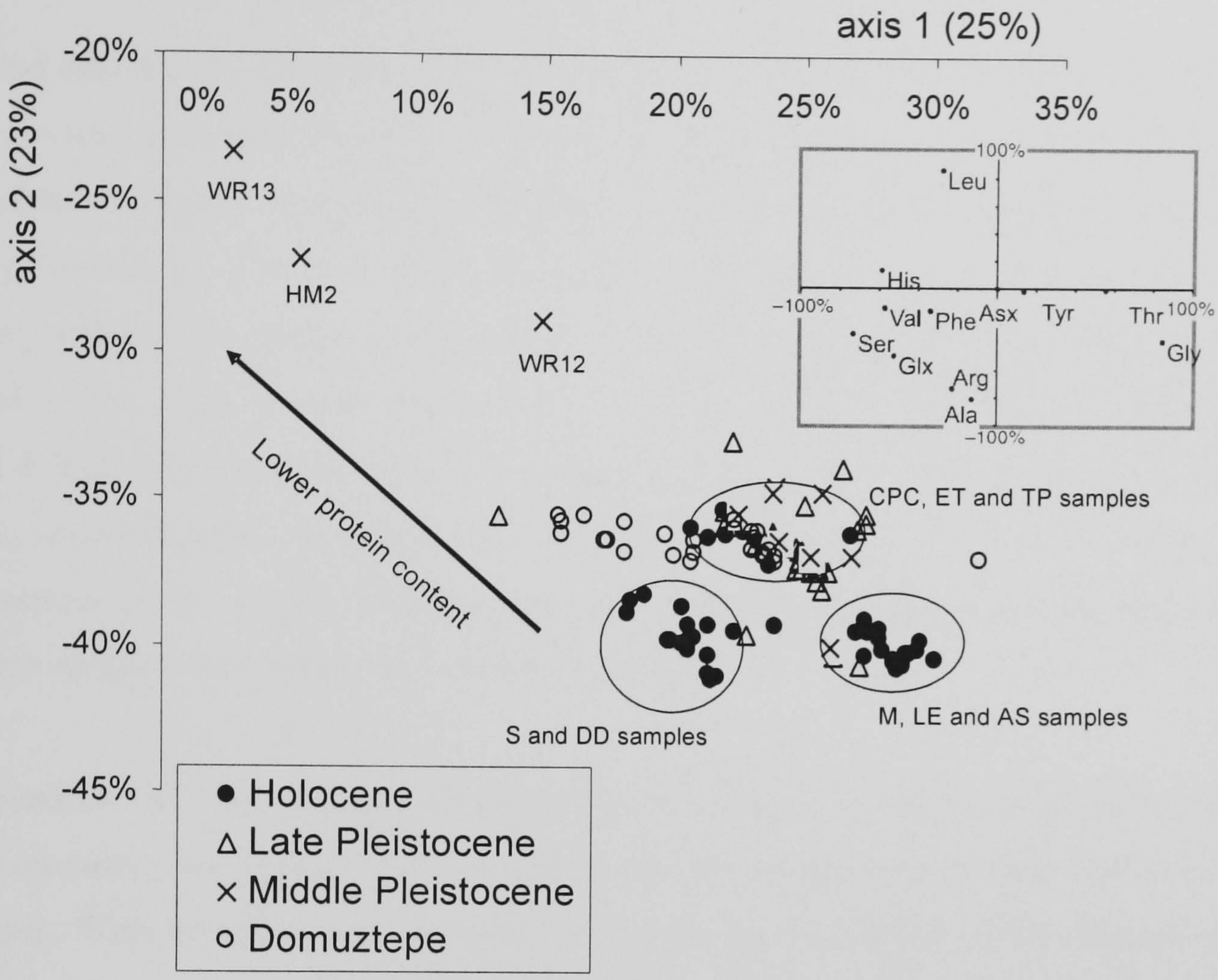


Figure 10.6 - Principal component analysis showing samples of different archaeological/palaeontological periods (samples from the archaeological site of Domuztepe analysed in Chapter 8 are also included) and the loadings of each amino acid presented (inset).

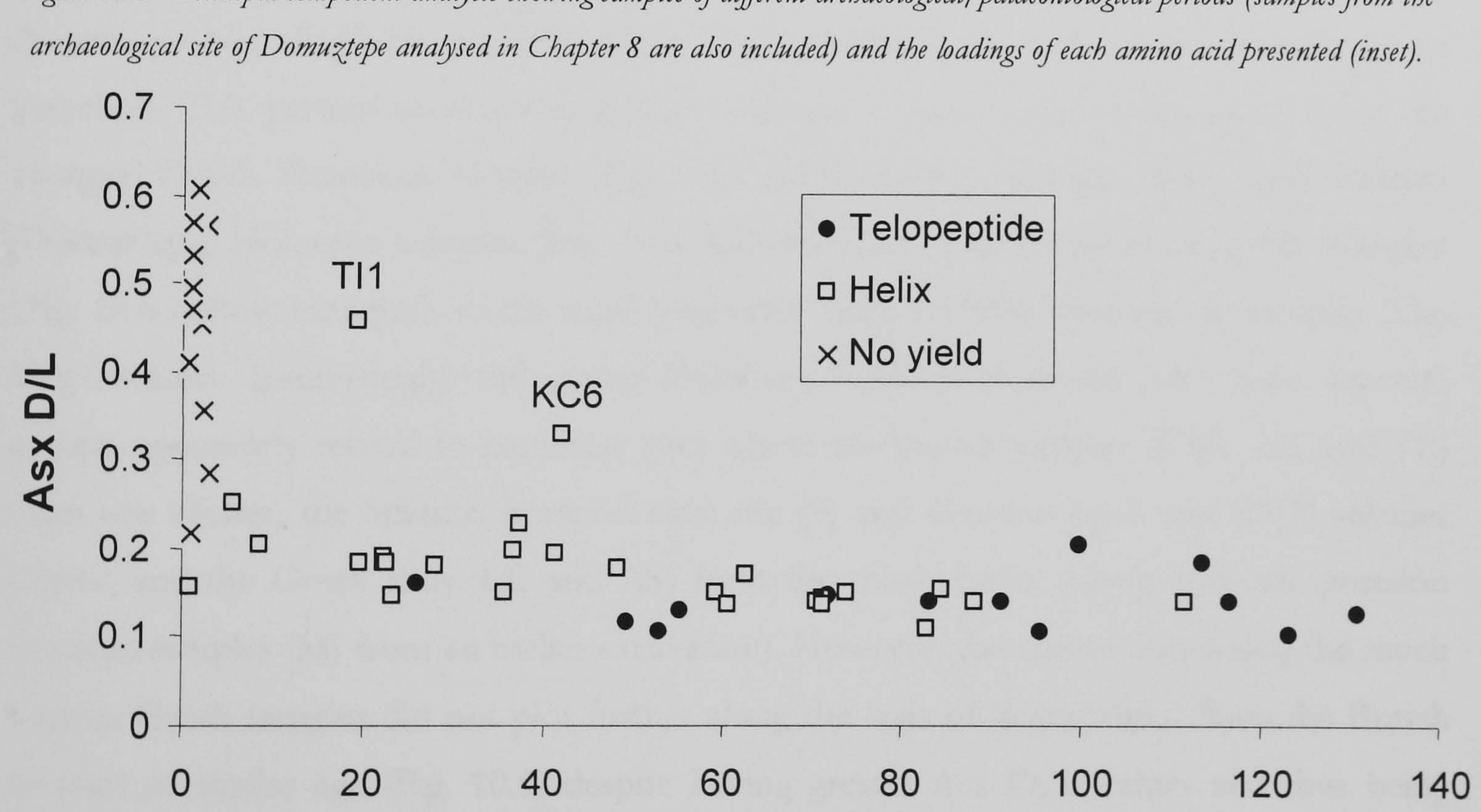


Figure 10.7 – A plot showing the Asx concentration as a function of Asx D/L for all 51 samples in relation to whether or not they yielded the carboxytelopeptide (Telopeptide), exhibited the presence of helical collagen (Helix), or yielded no acid-insoluble residue (No yield).

The amino acid composition plots (Fig. 10.5) indicate that even with the oldest (Cromerian) samples, including samples that did not yield SPE-isolated collagen-peptide markers or even acid-insoluble 'collagen' residues, a collagen-like profile (e.g. Gly to Asp ratio 5.5) (Elster *et al.* 1991) can still be observed. However, a few of the samples with very low yields and apparently collagen-like amino acid profiles did exhibit some compositional differences as identified in the PCA analyses (Fig. 10.6). It is interesting to note that, as observed in Chapter 4 (Fig. 4.5), the concentration of amino acids decreases with decreasing quality of collagen, assessed from the increasing Asx D/L values (Fig. 10.7). The amino acid concentrations of the samples that did not yield 'collagen' appear to be significantly lower than those samples that yielded the 'collagen' residues.

Plotting amino acid compositional data on the first two eigen vectors (accounting for 48% of the total variance), the youngest archaeological samples cluster with modern standards (data not shown). With increasing age/degradation there is a trend towards more negative values on axis 1 and more positive values on axis 2 (this trend appears to be dominated by a decrease in overall glycine content which loads positively on vector 1 and negatively on vector 2). This pattern reveals a progressive change in amino acid composition from the younger British Holocene samples (Fig. 10.6 filled circles), through older Near Eastern (Domuztepe) Holocene samples (Fig. 10.6 hollow circles), Late Pleistocene British samples (Fig. 10.6 hollow triangles), to the most 'degraded' British Middle Pleistocene samples (Fig. 10.6 crosses). Interestingly, the young Holocene samples clustered into three separate groups, apparently related to particular sites where the British samples (CPC, ET and TP) form one cluster, the Spanish portalón cave site (S) and German open site (DD) another cluster, and the Greek sites (LE and AS) form the third cluster (along with the portalón museum samples (M) from an earlier excavation). However, the cluster containing the much warmer Greek samples did not plot further along the 'axis of degradation' from the British samples of similar age (Fig. 10.6) despite having greater Asx D/L values and thus being effectively more degraded.

The range of glycine concentrations observed in these 51 palaeontological samples (see Appendix 6.4) was 2.68-1044.95 nmol/mg and Asx D/L values was 0.103-0.605. The range

of glycine concentrations observed in the seven samples that did not yield a 'collagen' residue was 2.684-22.201 nmol/mg and the range of Asx D/L values was 0.218-0.605. As indicated by Figure 10.7, amino acid concentrations were significantly different between those specimens that did yield a collagen residue and those that did not (Mann-Whitney U, p=0.000, N=51). Of the samples yielding a collagenous residue, the amino acid concentrations of those specimens not showing a signal for the carboxytelopeptide were also significantly different (lower) than those yielding the carboxytelopeptide (Mann-Whitney U, p=0.004, N=40). There also appeared to be a significant difference in the Asx D/L values of those samples that did yield at least some of the collagen-peptide markers, and those samples that did not yield an acid-insoluble 'collagen' residue (Mann-Whitney U, p=0.000, N=51), and of the samples that yielded the peptide markers (Fig. 10.7), the Asx D/L of those not showing a signal for the carboxytelopeptide was also significantly different from those that did (Mann-Whitney U, p=0.026, N=40). Based on the known relative rates of racemisation of particular amino acids, for example Asx>Ala=Glx>Ile=Leu, a greater Ala D/L value than Asx D/L value is considered an indicator of contamination (Bada *et al.* 1973). In the case of the 51 samples analysed in this study, even amongst the most racemised samples, such as HI5 and HM1, all have Asx D/L values higher than their respective Ala D/L values (see Appendix 6.4) and so no evidence of contamination is evident on this basis.

10.5 Discussion

10.5.1 Limitations of Collagen Survival – Telopeptide vs Helix

In comparing the survival of the two types of SPE-isolated species-specific collagen-peptide markers from the 40 ancient bone samples from which an acid-inolsuble 'collagen' residue was isolated (Table 10.2) it is clear that the 'helix' peptide markers survive in more samples of greater age than the carboxytelopeptide marker. There does not appear to be a clear relationship between the age of the assemblage and the survival of the selected collagen-peptide markers other than the survival of the carboxytelopeptide being confined to the youngest of samples (Carsington Pasture Cave, Etches Cave and North Sea samples). The almost complete absence of the carboxytelopeptide marker (Peptide T) in the older material such as the Devensian (Steetley Quarry Cave), Ipswichian (Kirkdale Cave), and Cromerian

(West Runton Forest Bed) samples indicates its relatively poor survival in comparison to that of the helix peptide markers, despite its greater survival than aDNA as shown in Chapter 6. This was supported by the complete absence of the carboxytelopeptide-containing Peptide D in all samples analysed. Unfortunately, as the proposed method of species identification by selected biomolecules relies upon both sets of peptides for certain species, the identifications of many samples were not satisfactory using SPE-based methods alone. In the absence of the carboxytelopeptide information in the older samples, the 'helix' Peptides A-C peaks are necessary to distinguish between most species investigated in this study. However, the observations of Peptides 26%A-C in the samples analysed was not as reproducible as the observations of Peptide 32%C, as evidenced by the poor success rates in the youngest (Neolithic) samples.

10.5.2 Survival of Particular Collagen Peptides – Supporting the 'Polymer-In-A-Box' Model of Collagen Degradation?

The random hydrolysis of collagen proposed by Collins *et al.* (1995) is not supported by the observations of the particular collagen peptides analysed in this study. For instance, Peptide A and/or Peptide B were less consistently observed than Peptide C, despite both peptides being much smaller and thus less likely to be cleaved if hydrolysis occurred at random positions in the collagen molecule. The peptides matched in the LC-MS search results against the 'Collagens' database (see Appendix 6.2) do not show a decrease in peptide length observed with progressively older samples. There were fewer observations of the carboxytelopeptide in progressively older (in chronological age) samples (Fig. 10.8). In support of the SPE-isolated marker methods and the 'polymer-in-a-box' model of collagen degradation, the α2 (I) carboxytelopeptide, that was observed in the LC-MALDI and LC-ESI analyses of modern specimens (Chapter 4), were not observed from any of the ancient specimens analysed by LC-MS (Appendix 6.5). This is consistent with the 'polymer-in-a-box' model of collagen stability, whereby it was proposed that degradation of the helical region of collagen is inhibited by the lack of space to collapse into (Miles & Ghelashvili 1999) but the telopeptides are not.

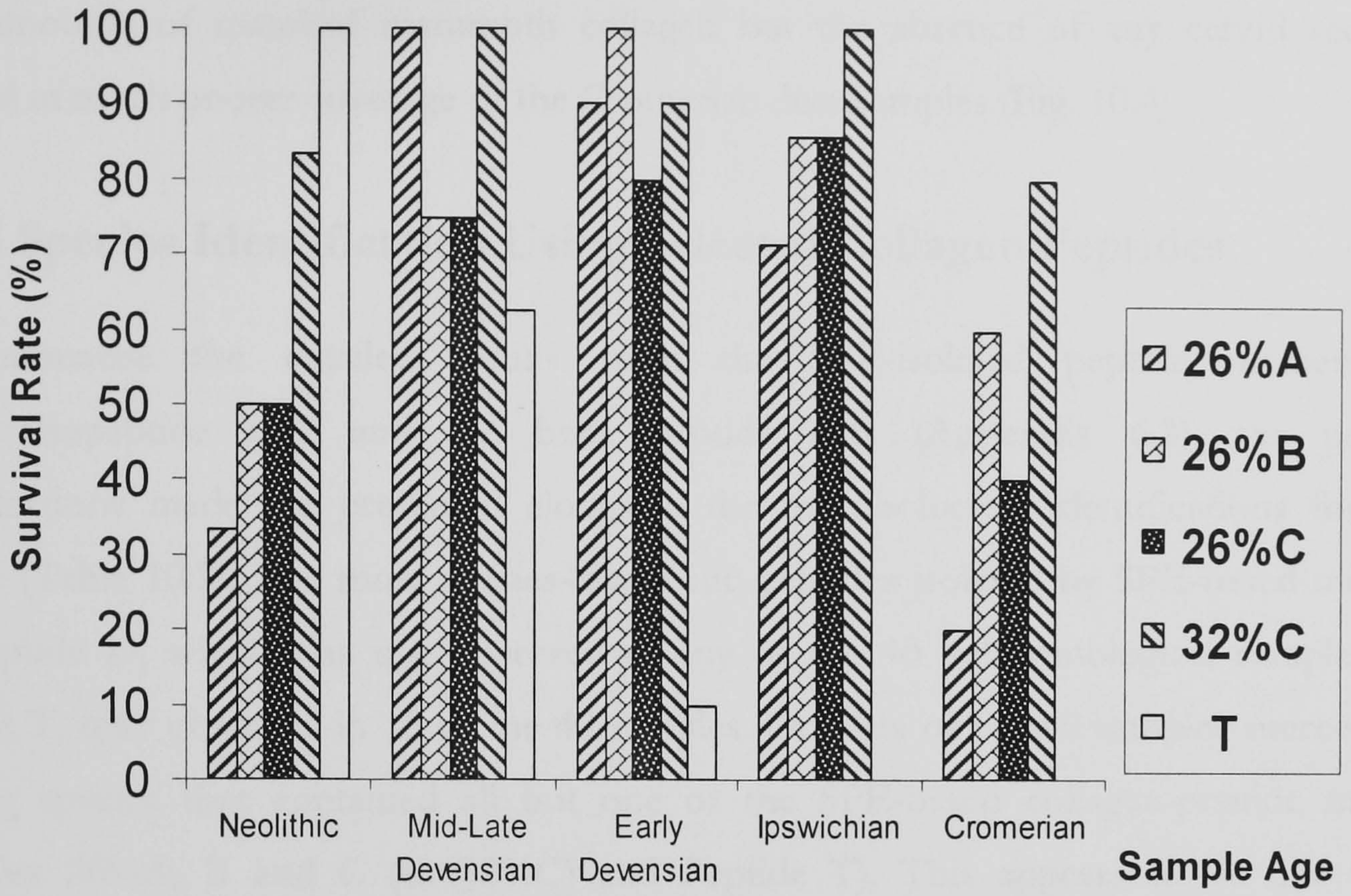


Figure 10.8 – Bar chart of collagen-peptide ('helix' Peptides A-C and carboxytelopeptide Peptide T) marker observation in 40 ancient bone samples. 26% and 32% indicates the concentration of ACN (in 0.1% TFA) of the fraction used to isolate the peptide.

In those samples that were from the Holocene and did show sufficient amounts of carboxytelopeptide for mass spectrometric analysis, signals for Peptide A were surprisingly weak or absent, possibly influenced by larger amounts of other peptides in the sample. In older material such as most of the Devensian and Ipswichian material as well as one Cromerian sample, these helix markers (Peptides 26%A, B and C) were observed. The most reliable SPE-isolated marker presented here, in terms of consistency with survival, appeared to be Peptide 26%B and Peptide 32%C. Because Peptide B has very limited sequence variation between species, the only informative peptide to be extracted from these very ancient samples using this method was Peptide 32%C. In the case of these very ancient samples for which the species is indeterminate, perhaps the most appropriate method of analysing the remaining collagen for species identification is by LC-MS peptide analysis. As Figure 10.4 highlights, large amounts (~50-80%) of the collagen molecule can be matched (and sequenced) even in Cromerian samples using either LC-MALDI or LC-ESI, yet LC-ESI appeared to generate the greater number of peptide matches in the database searches. The primary limiting factor in the LC-MS analyses appeared to be the lack of sequence information in the database, for example, the presence of an elephantid sequence resulted in

large amounts of matched mammoth collagen but the absence of any cervid sequence resulted in much poorer coverage of the Cromerian deer samples (Fig. 10.4).

10.5.3 Species Identification Using Selected Collagen Peptides

To summarise the detailed results from the SPE-isolated peptide markers, the carboxytelopeptide data and the helix peptide data (Appendix 6.3), the possible identifications made are presented alongside the morphological identifications for each sample (Table 10.2). The most species-diagnostic peptides isolated by SPE-based methods are Peptide D, which was not observed in any of the 40 palaeontological samples, and Peptide T, only observed in 15 of the 40 samples. Only six of the 40 samples succeeded in yielding spectra that contained all but one of the SPE-based collagen-peptide markers (Peptides 26%A, B and C (or 32%C) and Peptide T). This appears to be due to the unpredictable observation of Peptides A and B in the spectra of samples that are relatively well-preserved (as determined by amino acid compositions). Both Peptide C and Peptide T were observed together in the spectra of only 13 of the 40 samples. The carboxytelopeptide (either Peptide D or T) was not observed in many of the spectra from the early Devensian samples and not observed in any of the spectra from the older specimens (Fig. 10.8). Without the carboxytelopeptide, species identification using the SPE-isolated peptide markers is much more limited and requires the observation of at least Peptides A and C.

As indicated in Table 10.2, in the Neolithic samples, the collagen-peptide marker m/z values (Appendix 6.4) confirmed the identity of the aurochs specimen as bovine, the pig specimen as porcine, the caprine sample as either sheep or red deer, the horse sample as equine and the roe deer sample as specific to roe deer (due to its Peptide C m/z value being unique to the species investigated throughout this thesis). However, it should be noted that they are not species-specific enough to distinguish between cattle and aurochs, or even cattle and bison. Use of the 'helix' peptides made the identification of the pig, horse and the roe deer samples possible, which could not be distinguished from other taxa using the carboxytelopeptide marker alone. The roe deer carboxytelopeptide m/z values are identical to that of the red deer, sheep and water buffalo, and the horse carboxytelopeptide was consistent with that described in section 5.3.2. However, although the SPE-isolated 'helix'

peptide markers were capable of distinguishing between sheep and goat, they could not distinguish between sheep and red deer. The Etches Cave samples all belonged to species that had not been analysed previously; bear (*Ursus sp.*) and felid (small cat and large cat). The peptide *m/z* values for the carboxytelopeptide fragments of felid (ET5) and bear (ET3 & ET4) are consistent with the partial sequence of cat collagen (I) from the UCSC alignments (Appendix 5.1), which is identical to that of pig and dog collagen. A unique Peptide B *m/z* of 2163.1 was observed in one of the bear specimens (ET1), but was absent in the other two Etches Cave bear specimens (ET3 and ET4) and inconsistent with the only other known bear specimen (UP3) with *m/z* 2147.0 (Appendix 6.3). In order to identify whether or not the peptide of *m/z* 2163.1 was Peptide B, *de novo* sequencing of this peptide was carried out (Fig. 10.9) and indicated similarities with the cattle Peptide B (GLPGVAGSVGEPGPLGIAGPPGAR, hydroxylated residues are underlined) but where the alanine at position 18 is replaced with a serine residue in the sequence GLPGVAGSVGEPGPLGISGPPGAR (hydroxylated residues are underlined).

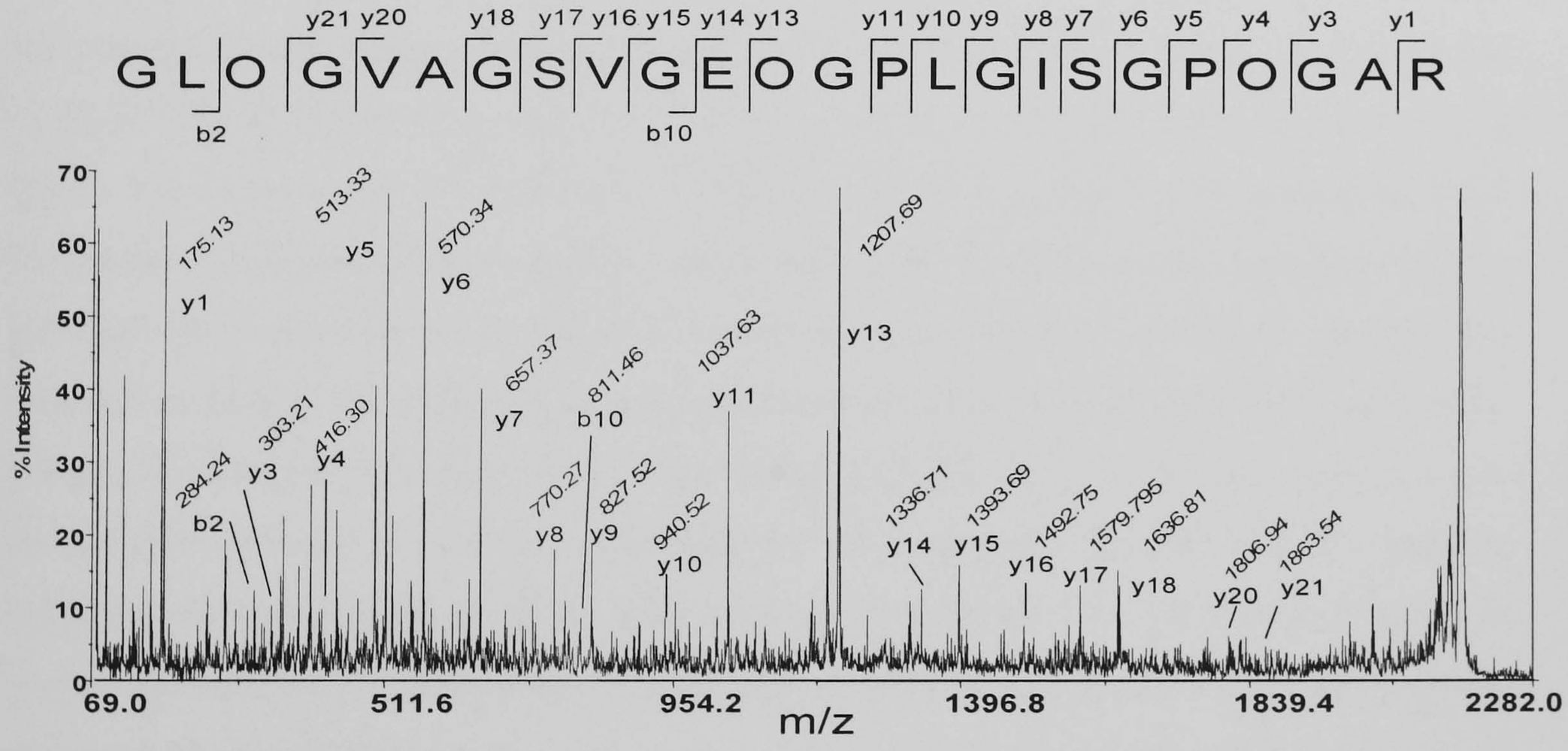


Figure 10.9 - *Manual interpretation for de novo sequencing of product ion spectra showing the identification of the Peptide B sequence for bear marker at m/z 2163.1.*

Thus the peak at *m/z* 2163.1 appears to be Peptide B from a morphologically-determined bear sample (ET3); the peak of *m/z* 2147.0 observed in the other bear sample (ST36) is

possibly indicative of the same peptide with one fewer hydroxylation (-16 Da) but could not be confirmed due to the poor quality of the product ion mass spectrum obtained. Thus, the use of either peak *m/z* as a peptide marker to identify ST36 as bear is unsupported without further research using modern bear specimens. Regarding the y-ions observed in this example of a peptide from the proline-rich helical region of collagen, the effect where fragmentation carboxyterminal to proline (and hydroxyproline) residues is less commonly observed in CID spectra than fragmentation aminoterminal to the proline residue (Breci *et al.* 2003; Vaisar and Urban 1996), can be observed in the greater abundance of the y11 ion than the y10 ion, and the dominant y13 ion (y12 is absent); although note the y3 ion is of a similar intensity to the y4 ion (both of very weak intensity). However, this does not appear problematic for identification of the sequence.

Of the 13 Devensian samples, carboxytelopeptide peaks were only observed in the spectrum of a single specimen (ST33) allowing the identification of that sample as cattle/bison. The 'helix' peptide markers were observed in the spectra of most of these Devensian samples, allowing the identifications of canid specimens (ST11 and ST12), and several samples limited to 'bovid' (ST18, ST23, ST33, and TP1). As noted above, the possible bear Peptide B at *m/z* 2163.1 was observed in the spectrum of specimen ST36 (Appendix 6.3). Several of these Devensian specimens were incorrectly identified by SPE-isolated peptide markers as goat although they originate from reindeer (based on morphological criteria), a species not analysed previously but sharing the same peptide sequences. This highlights the fact that the Peptide C, used for differentiation between sheep and goat, is not only less appropriate in the possible presence of red deer (matching the sheep peptide markers), but also reindeer (matching the goat peptide markers). Carboxytelopeptide peaks were not observed in any of the spectra from the 12 older samples (Ipswichian-Cromerian), and the 'helix' peptide peaks that were observed were uninformative (most identifications were to the taxonomic level of family, such as 'bovid'). The only specimen showing different peptide marker *m/z* values was from the sample speculated to be rhinoceros, with Peptide A being represented by a peak of *m/z* 1437.7 and Peptide B being represented by a peak of *m/z* 2145.1 (Appendix 6.3) (the product ion mass spectra of both ions were too poor in quality to determine the peptide sequences and there is no database sequence for a rhinoceros species).

There were obvious limitations of identification using SPE-based and LC-MS-based collagen-peptide markers, most notably the difficulty in distinguishing between cattle and bison, a problem with two similar species that occupy different niches but whose post-cranial bones are very similar. Another major limitation using the SPE-isolated peptide markers was the differentiation between red deer and sheep (although this is less likely to be as problematic as the cattle and bison distinction using morphological criteria because of the more obvious size differences between sheep and red deer). In the older samples where the carboxytelopeptides were not detected, using only Peptide C (26%C and/or 32%C) it was not possible to distinguish between pig, cattle and red deer. The real potential of the selected SPE-isolated peptide markers appears to be for targeting particular species distinctions, such as the sheep/goat paradigm (Chapter 8), or perhaps distinguishing roe deer from sheep and goat, or water buffalo from cattle/bison, and reindeer from red deer. However, when the LC-MS-based peptide markers are also investigated for some of these species, the species-specificity is greatly improved. The most obvious improvement is the discrimination between cervids (deer) and bovids (cattle, sheep, goat, etc.), which was useful in the identifications for samples WR8, WR9 and DR1. These could only be identified as cattle/bison/sheep/red deer, goat/reindeer and sheep/red deer respectively using the SPE-isolated peptide markers, whereas with the additional LC-MS peptide markers they could be identified as red deer, reindeer and red deer respectively. Ideally, further investigations into increasing the number and reproducibility of collagen peptides observed in the spectra from SPE-isolated fractions should be carried out, with the particular aim to isolate and analyse some of the LC-MS peptide markers capable of discriminating between cervids and bovids, and a wider range of modern samples that take into consideration all possible species present in British archaeological sites should also be investigated (e.g., modern bear, cat, rhino specimens). It is quite clear that at this stage in the development of collagen-peptide marker approaches for archaeological species identification, archaeozoological knowledge, whether experience with morphological analysis of various relevant species or simply background knowledge regarding the faunal history of the site/assemblage of study is still beneficial to simplify the possible species of the specimens.

10.5.4 Amino Acid Analyses as a Screening Tool for Ancient Samples

As described in section 1.4.3, a nitrogen content of about 4% for modern bone corresponds to around 2500 nmol/mg of bone and as long as the (total) amino acid content of bone remains above about 250 nmol/mg (~0.4% nitrogen content), the relative amino acid pattern resembles that of collagen, with large amounts of glycine and substantial amounts of proline and hydroxyproline; below about 50 nmol/mg of fossil bone, the relative amino acid pattern was suggested to be unlike collagen (Hare 1980). Because all samples, even those of no yield (using <1 g) showed collagen-like profiles (Fig. 10.5) it is plausible that if more material was sampled, collagen (I) could have been retrieved from even the lowest protein-content samples. Interestingly, despite predictions that only the amino acids in the collagen telopeptides are able to racemise in intact collagen, as in helical collagen they are structurally constrained (van Duin & Collins, 1998), the extent of racemisation in the majority of Pleistocene samples analysed was substantially higher than could derive from racemisation of amino acids in the telopeptides alone. Thus the large number of samples with high Asx D/L values (>0.2) could be interpreted to indicate that at least the majority of collagen remaining in the sample has been structurally altered. Perhaps the helical structure of the majority of these collagen molecules has partially denatured, allowing Asx racemisation to occur, yet a network of structurally altered collagen remains insoluble either due to extensive cross-linking with unaltered collagen fibrils, or large amounts of altered protein-humic substances.

Although collagen was only confirmed in samples of glycine concentration ~200 nmol/mg (approximate 'total' amino acid concentration ~600 nmol/mg where glycine makes up approximately 1/3 of all amino acids in the collagen molecule), the observation of collagen-like profiles in the samples with glycine concentrations as low as ~3 nmol/mg (approximate 'total' amino acid concentrations ~9 nmol/mg) reduce the amino acid concentration cut-off for collagen survival as suggested by Hare (1980) from 50-250 nmol/mg to ~10 nmol/mg. Although the amino acid compositions suggested collagen survival in samples with Asx D/L as high as 0.605, it could be postulated based on these results that the majority of samples with an Asx D/L >0.25 are unlikely to be worthwhile sampling for collagen-peptide marker analysis. However, this is more likely to be due to the low levels of amino acids (and hence protein) present rather than the Asx D/L values, as supported by the two outliers in Figure

10.7 that did yield collagen peptides, TI1 and KC6. However, these two outliers in the Asx D/L vs Asx concentration plot are both within the range for Holocene samples in the principal component analysis shown in Figure 10.6, possibly indicating that amino acid composition plots as shown in Figure 10.6 are more suitable for screening suitable samples than the commonly-used concentration vs Asx D/L plots (such as Fig. 10.7). However, Asx concentrations plotted against Asx D/L of the 51 samples (Fig. 10.6) indicates that amino acid concentration is a more suitable predictor of collagen survival than Asx D/L, despite the use of Asx D/L for this aim (Poinar *et al.* 1996).

10.6 Conclusions

With the between-sample variability in the SPE-based collagen-peptide marker approaches described, the LC-MS proteomics-based approach may be the most appropriate approach for important unidentified specimens, such as unidentified butchered bones from Boxgrove for example. However, this can be costly and thus not appropriate for analyses of lots of samples from most archaeological sites, particularly with the limited number of species's collagen sequences that are present in the currently used databases. The SPE-isolated collagen-peptide marker approaches are ideal for bulk analyses of samples where some information regarding potential species can be derived, such as discriminating between sheep/goat bones, or between several deer species, etc. But this approach may not be suitable if the samples are completely indeterminate unless the methods are further developed to include more peptide markers. However, for the routine analysis of large numbers of archaeological samples, simple methods where only a select number of peptides are isolated as species-specific markers would be more ideal than the more expensive LC-MS approaches and the described methods should be modified if necessary (by trial and error) in order to apply them to other species of interest.

11 Comment on "Protein Sequences from Mastodon and Tyrannosaurus rex Revealed by Mass Spectrometry"

Reprinted from *Science, Volume 319, 4 January 2008, Pages 33c*, Copyright 2008, with permission from the American Association for the Advancement of Science.

This is in response to Asara, J., Schweitzer, M. H., Freimark, L. M., Phillips, M., 2007. Protein Sequences from Mastodon and Tyrannosaurus rex Revealed by Mass Spectrometry. Science 31, 280

Matthew Collins proposed the reply to Asara et al. (2007) and was responsible for the write-up. I, along with the many co-authors, contributed to the data interpretation that forms a major part of this chapter, as well as to the final write-up. I also carried out the amino acid analysis of many of the samples referred to by this chapter.

Comment on "Protein Sequences from Mastodon and *Tyrannosaurus rex* Revealed by Mass Spectrometry"

Mike Buckley,[1] Angela Walker,[2] Simon Y. W. Ho,[3] Yue Yang,[1] Colin Smith,[4] Peter Ashton,[1] Jane Thomas Oates,[1] Enrico Cappellini,[1] Hannah Koon,[1] Kirsty Penkman,[1] Ben Elsworth,[1] Dave Ashford,[1] Caroline Solazzo,[1] Phillip Andrews,[2] John Strahler,[2] Beth Shapiro,[6] Peggy Ostrom,[5] Hasand Gandhi,[5] Webb Miller,[6] Brian Raney,[7] Maria Ines Zylber,[8] M. Thomas P. Gilbert,[9] Richard V. Prigodich,[10] Michael Ryan,[11] Kenneth F. Rijsdijk,[12] Anwar Janoo,[13] Matthew J. Collins[1]*

We used authentication tests developed for ancient DNA to evaluate claims by Asara *et al.* (Reports, 13 April 2007, p. 280) of collagen peptide sequences recovered from mastodon and *Tyrannosaurus rex* fossils. Although the mastodon samples pass these tests, absence of amino acid composition data, lack of evidence for peptide deamidation, and association of α1(I) collagen sequences with amphibians rather than birds suggest that *T. rex* does not.

Early reports of DNA preservation in multimillion-year-old bones (i.e., from dinosaurs) have been largely dismissed (*1*, *2*) (table S1), but reports of protein recovery are persistent [see (*3*) for review]. Most of these studies used secondary methods of detection, but Asara *et al.* (*2*) recently reported the direct identification of protein sequences, arguably the gold standard for molecular palaeontology, from fossil bones of an extinct mastodon and *Tyrannosaurus rex*. After initial optimism generated by reports of dinosaur DNA, there has been increasing awareness of the problems and pitfalls that bedevil analysis of ancient samples (*1*), leading to a series of recommendations for future analysis (*1*, *4*). As yet, there are no equivalent standards for fossil protein, so here we apply the recommended tests for DNA (*4*) to the authentication of the reported mastodon and *T. rex* protein sequences (*2*) (Table 1).

First, the likelihood of collagen survival needs to be considered. The extremely hierarchical structure of collagen results in unusual, catastrophic degradation (*5*) as a consequence of fibril collapse. The rate of collagen degradation in bone is slow because the mineral "locks" the components of the matrix together, preventing helical expansion, which is a prerequisite of fibril collapse (*6*). The packing that stabilizes collagen fibrils (*6*) also increases the temperature sensitivity of degradation (E_a 173 kJ mol^{-1}) (Fig. 1). Collagen decomposition would be much faster in the *T. rex* buried in the then-megathermal (>20°C) (*7*) environment of the Hell Creek formation [collagen half-life ($T_{½}$) = ~ 2 thousand years (ky] than it would have been in the mastodon lying within the Doeden Gravel Beds (present-day mean temperature, 7.5°C; collagen $T_{½}$ = 130 ky) (Fig. 1).

This risk of contamination also needs to be evaluated. Collagen is an ideal molecular target for this assessment because the protein has a highly characteristic motif that is also sufficiently variable to enable meaningful comparison between distant taxa if enough sequence is obtained (Fig. 2). Compared with ancient DNA amplification, contamination by collagen is inherently less likely. Furthermore, because the bones sampled in (*2*) were excavated by the

Table 1. Key questions to ask about ancient biomolecular investigations [adapted from (*4*)].

Test	Sample	Pass	Observation
Do the age, environmental history, and preservation of the sample suggest collagen survival?	Mastodon, 300 to 600 ky old	Yes	Collagen $T_{½}$ at 7.5°C = 130 ky
	T. rex, 65 million years old	No	Collagen $T_{½}$ at 20°C = 2 ky
Do the biomolecular and/or macromolecular preservation of the sample, the molecular target, the innate nature of the sample, and its handling history suggest that contamination is a risk?	Biomolecular preservation	?	Range of evidence presented (*8*) but no amino acid compositional data
	Macromolecular preservation	Yes	Macromolecular preservation is not the equivalent of biomolecular preservation (*9*)
	Molecular target	Yes	
	Handling history	Yes?	Large (2.5 g) samples increase risk of contamination?
Do the data suggest that the sequence is authentic, rather than the result of damage and contamination?	Mastodon and *T. rex*	No	Errors in interpretation of spectra [see table S1 and (*13*)]? Damage-induced errors in sequence
Do the results make sense, and are there enough data to make the study useful and/or to support the conclusions?	Mastodon	Yes	Weak affinity to mammals
	T. rex	No	Affinity of α1(I) peptides to amphibians, not birds or reptiles

[1]BioArch, Departments of Biology, Archaeology, Chemistry and Technology Facility, University of York, Post Office Box 373, York YO10 5YW, UK. [2]Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI 48109–0404, USA. [3]Evolutionary Biology Group, Department of Zoology, University of Oxford, OX1 3PS, UK. [4]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103, Leipzig, Germany. [5]Department of Zoology, Michigan State University, East Lansing, MI 48824, USA. [6]Department of Biology, Pennsylvania State University, University Park, PA 16802, USA. [7]Center for Biomolecular Science and Engineering, University of California–Santa Cruz, CA 95064, USA. [8]Department of Parasitology, Kuvin Center, Hebrew University of Jerusalem, Israel. [9]Biological Institute, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark. [10]Chemistry Department, Trinity College, 300 Summit Street, Hartford, CT 06106, USA. [11]Cleveland Museum of Natural History, 1 Wade Oval Drive, University Circle, Cleveland, OH 44106, USA. [12]National Museum of Natural History "Naturalis," P.O. Box 9517, 2300 RA Leiden, Netherlands. [13]National Heritage Trust Fund Mauritius, Mauritius Institute, La Chaussée Street Port Louis, Mauritius.

*To whom correspondence should be addressed. E-mail: mc80@york.ac.uk

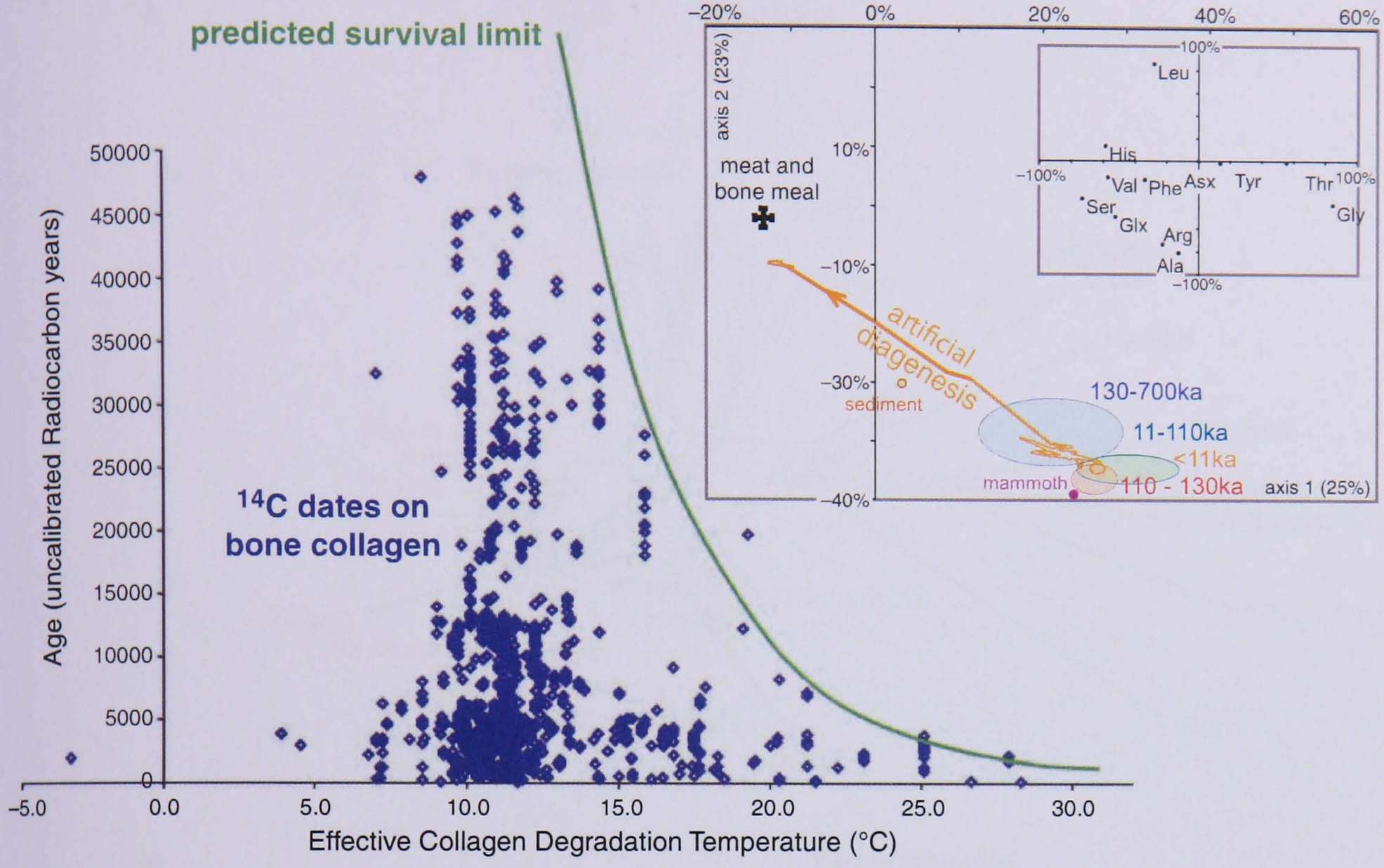


Fig. 1. Plot of radiocarbon age versus estimated effective collagen degradation temperature for radiocarbon-dated bones from laboratory databases (principally Oxford and Groningen). The line represents the expected calendar age at which 1% of the original collagen remains following a zero-order reaction; almost no bone collagen survives beyond this predicted limit. (**Inset**) The 99% confidence intervals of amino acid compositions by first two principal component analyses (48% of total variance) for bones from NW Europe aged <11 ky (n = 324), 11 to 110 ky (n = 210), 110 to 130 ky (n = 26), and 130 to 700 ky (n = 31). Pliocene samples are not plotted, as their composition (n = 8) is highly variable and yields of amino acids are low. The orange line indicates a compositional trend observed when compact bone is heated for 32 days at 95°C, which reduces collagen to 1% of the initial concentration [each inflection represents a separate analysis; n = 32)]. The composition becomes more similar to mixed tissue samples (meat and bone meal; n = 32), principally due to the depletion of Gly. An amino acid profile for mammoth is consistent with collagen, unlike the associated sediment sample [data from (*11*)].

authors, obvious contamination sources such as animal glue (used in conservation) can be excluded. However, concentrating protein from the large amounts of bone used (2.5 g) may have heightened the risk of extraneous proteins entering the sample during extraction, although there have been no systematic studies of this phenomenon. Independent extraction and analyses would have strengthened claims for the authenticity of the origin of the peptides (and potentially ameliorated the original problems of data interpretation) (*4*).

The remarkable soft-tissue preservation of the investigated *T. rex* specimen (MOR 1125) has been documented (*8*). However, microscopic preservation does not equate with molecular preservation (*9*). Immunohistochemistry provides support for collagen preservation, but Asara *et al.* (*2*) presented no data regarding inhibition assays with collagen from different species or cross-reactivity with likely contaminants [e.g., fungi (*10*)]. Curiously, no amino acid compositional analysis was conducted [see (*11*)], although immonium ions were identified by time-of-flight secondary ion mass spectrometry. In our experience, collagen-like amino acid profiles have been obtained in all bones from which we could obtain collagen sequence (Fig. 1, inset).

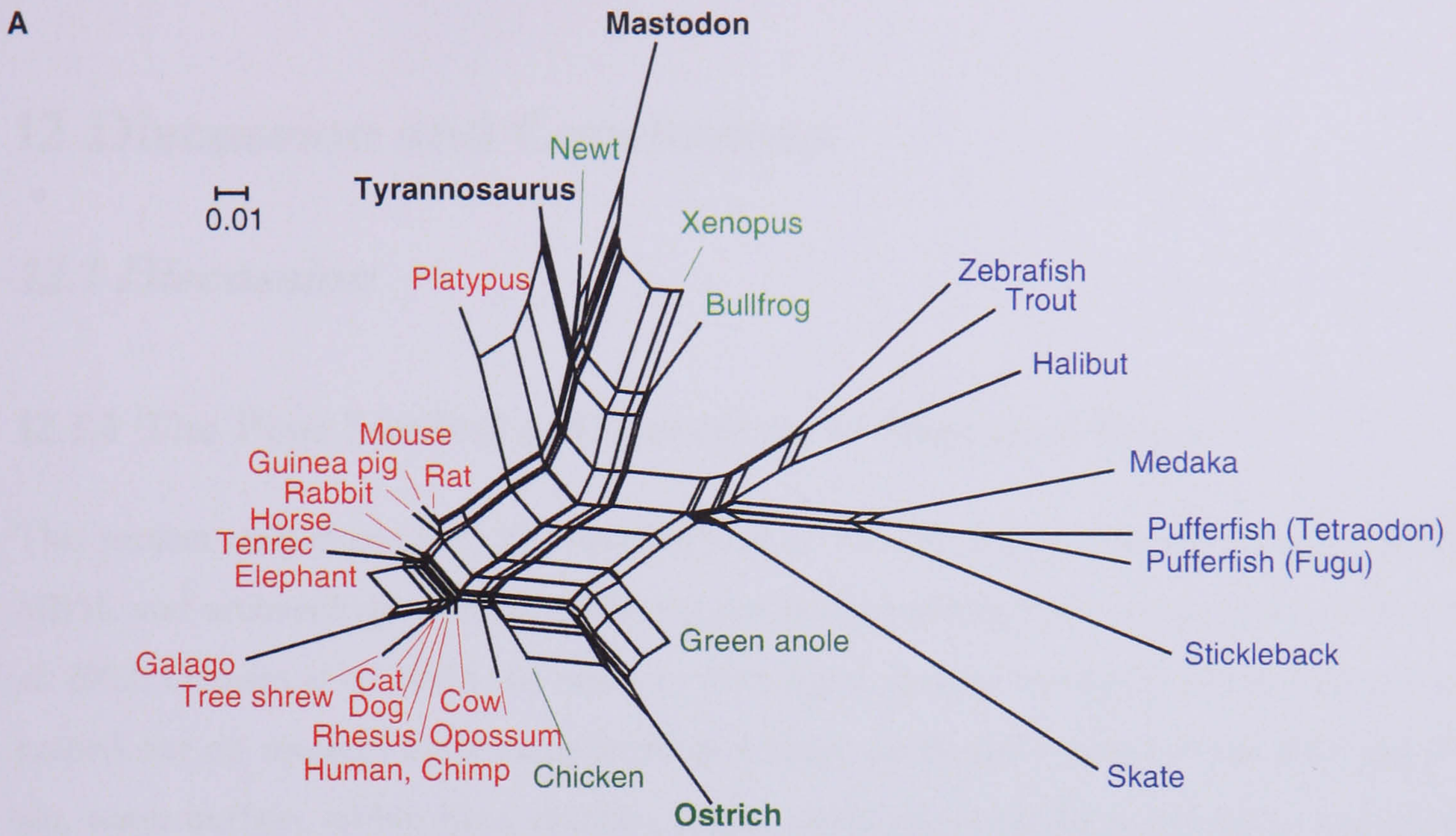
Regarding the proof of sequence authenticity, the spectra reported by Asara *et al.* (*12*) are inconsistent with some of the sequence assignments (*13*) (table S1). A common diagenetic modification, deamidation, not considered in (*2*), may shed light on authenticity. The facile succinimide-mediated deamidation (*14*) of asparagine occurred at $N_{229}G$ and $N_{1156}G$ in ostrich peptides (Ost 4 and Ost5) (see table S1 for nomenclature), presumably during sample preparation. Direct hydrolytic deamidation is slower (*14*), and an expectation of elevated levels of such products is reasonable for old samples. We agree with the most recent interpretation (*13*) of the spectrum illustrated in Fig. 2B as α1(I) $G_{362}SEGPEGVR_{370}$, the deamidated ($Q \rightarrow E_{367}$) form of the sequence found in most mammals (*12*). By way of contrast, none of the three glutamine residues in the reported *T. rex* peptides are deamidated (table S1). Only time will tell if Q→E is a useful marker for authentically old collagen, but from the evidence presented, the mastodon sequence looks more diagenetically altered than *T. rex*.

The unusual, fragmented nature of the reported *T. rex* sequence does not make it amenable to standard, model-based phylogenetic analysis. Instead, we examined the phylogenetic signal of the α1(I) fragments of mastodon and *T. rex* using Neighbor-Net analysis and uncorrected genetic distances. Using the sequences reported in (*13*), both the *T. rex* and mastodon signal display an affinity with amphibians (Fig. 2A). Our reinterpretation of the spectra (*12*) changes the affinity of mastodon but not of *T. rex* (Fig. 2B). In addition to the α1(I) peptides used in the Neighbor-Net analysis, Asara *et al.* reported two other peptides from *T. rex* (*13*); we question the interpretation of the α1(II) spectra (identical to frog) but not the α2(I) spectra (identical to chicken).

We require more data to be convinced of the authenticity of the *T. rex* collagen sequences reported by Asara *et al.* Nevertheless, the handful of spectra reported for the temperate Pleistocene mastodon fail neither phylogenetic nor diagenetic tests, thus highlighting the potential of protein mass spectrometry to bridge the present gulf in our understanding between the fate of archaeological and fossil proteins. To avoid past mistakes of ancient DNA research (*1*), we recommend that future fossil protein claims be considered in light of tests for authenticity such as those presented here.

Reference and Notes

1. E. Willerslev, A. Cooper, *Proc. R. Soc. London. B. Biol. Sci.* **272**, 3 (2005).
2. J. M. Asara, M. H. Schweitzer, L. M. Freimark, M. Phillips, L. C. Cantley, *Science* **316**, 280 (2007).
3. M. H. Schweitzer, *Palaeontologia Electronica* **5**, editorial 2 (2003); http://palaeo-electronica.org/2002_2/r_and_p.pdf
4. M. T. P. Gilbert, H.-J. Bandelt, M. Hofreiter, I. Barnes, *Trends Ecol. Evol.* **20**, 541 (2005).
5. M. J. Collins, M. S. Riley, A. M. Child, G. Turner-Walker, *J. Archaeol. Sci.* **22**, 175 (1995).
6. C. A. Miles, M. Ghelashvili, *Biophys. J.* **76**, 3243 (1999).
7. K. R. Johnson, D. J. Nichols, J. H. Hartman, *The Hell Creek Formation and the Cretaceous-Tertiary Boundary in the Northern Great Plains: Geological Society of America Special Paper* **361**, 503–510 (2002).
8. M. H. Schweitzer *et al.*, *Science* **316**, 277 (2007).
9. N. S. Gupta, D. E. G. Briggs, R. D. Pancost, *J. Geol. Soc. London* **163**, 897 (2006).
10. P. Sepulveda *et al.*, *Infect. Immun.* **63**, 2173 (1995).



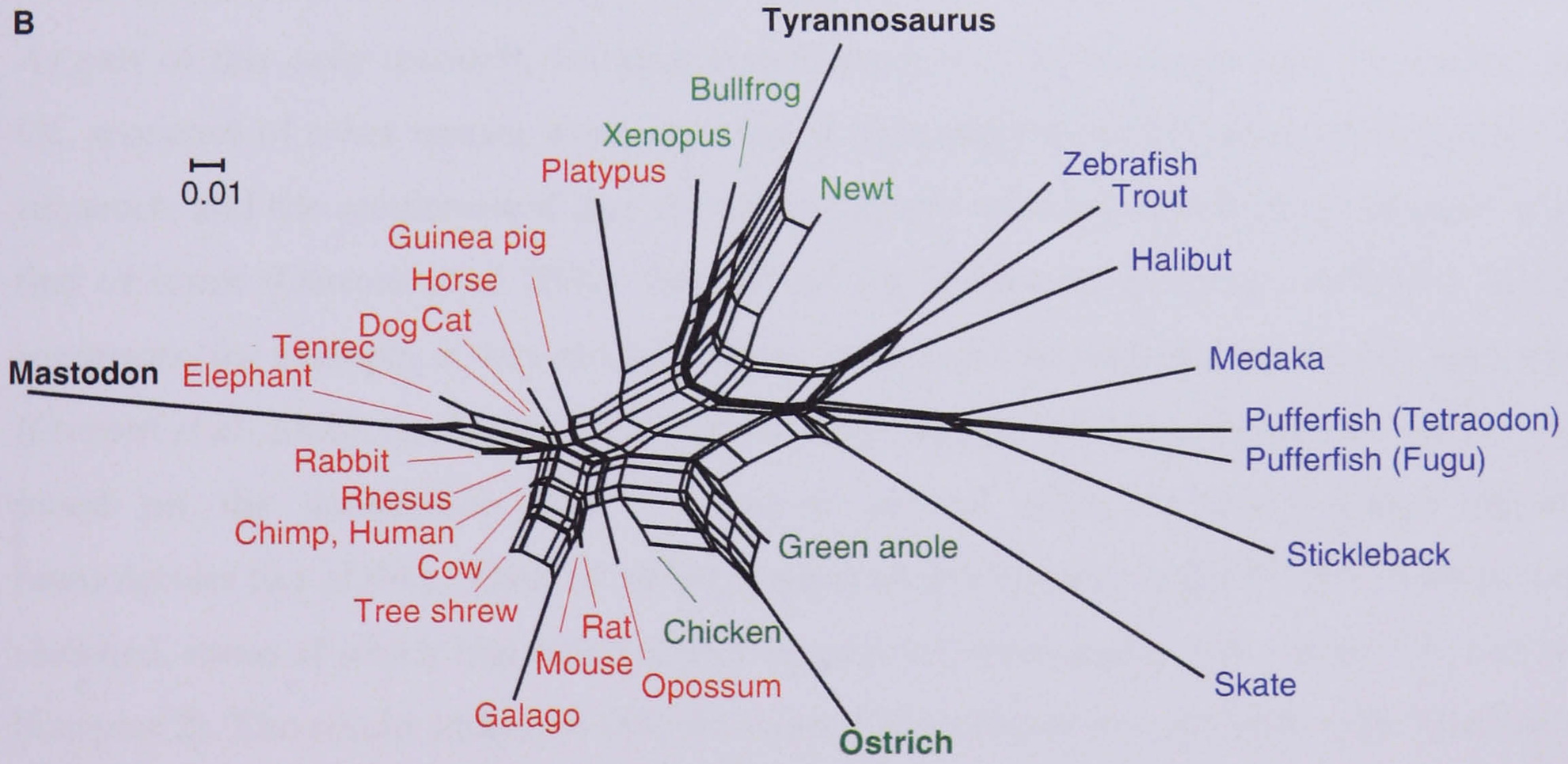


Fig. 2. Phylogenetic networks of α1(I) sequences using Neighbor-Net analysis (**A**) with the most recent Asara *et al.* assignments (*13*) and (**B**) after our reinterpretation of the mass spectrometric data (*12*). *T. rex* does not group with bird/reptile using either set of sequence alignments. More sequence is required for a full, model-based phylogenetic analysis.

11. M. Schweitzer, C. L. Hill, J. M. Asara, W. S. Lane, S. H. Pincus, *J. Mol. Evol.* **55**, 696 (2002).
12. See supporting material on *Science* Online.
13. J. M. Asara *et al.*, *Science* **317**, 1324 (2007).
14. N. E. Robinson *et al.*, *J. Pept. Res.* **63**, 426 (2004).

15. This work was supported by NSF (EAR-0309467), National Environment Research Council (NE511148, GR9/01656), National Center for Research Resources (P41-18627), European Commission (MEST-CT-2004-007909, MEST-CT-2005-020601), Wellcome Trust Bioarchaeology Fellowships (K.P. and H.K.) Analytical Chemistry Trust Fund, the Royal Society of Chemistry Analytical Division, Engineering and Physical Sciences Research Council, and the Michigan Proteome Consortium.

Supporting Online Material
www.sciencemag.org/cgi/content/full/319/5859/33c/DC1
SOM Text
Table S1
References

26 June 2007; accepted 20 November 2007
10.1126/science.1147046

12 Discussion and Conclusions

12.1 Discussion

12.1.1 The Poor Survival of Osteocalcin in Degraded Bone

This project started out with the investigation of OC for species identification purposes in MBM, and archaeological bone, following methods published previously (Nielsen-Marsh *et al.* 2002; Ostrom *et al.* 2000). In order to investigate species variability, initial analyses were carried out on modern bone samples from a range of species: cattle, sheep, goat, red deer, pig, water buffalo, rabbit, hare, chicken, turkey, duck, pheasant, hippopotamus, and the m/z of the protonated OC molecules of each species identified (M. Buckley unpublished data). As part of this early research, working with Ostrom *et al.* at Michigan State University, the OC sequence of other species were investigated including the re-evaluation of the horse OC sequence, and the confirmation that the OC sequences of zebra and donkey are identical to that of horse (Ostrom *et al.* 2006). Because of the limitations in species-variability of OC sequences, for example, it was not possible to distinguish between horse, donkey and zebra (Ostrom *et al.* 2006), the major benefit of investigating OC as a species-specific marker was based on the assumption that it survived beyond other more-informative ancient biomolecules like aDNA. Thus 34 archaeological samples from a larger aDNA analysis were obtained, some of which had failed aDNA extractions, were analysed for intact OC survival (Chapter 2). The results indicated that the intact OC molecule was not detectable in as many samples as those yielding aDNA. Intact OC was not observed in any of the thirteen samples from the Greek open sites (T_{eff} 21°C), yet six of these thirteen samples succeeded the aDNA analyses. In contrast, the majority of samples from the German open sites (T_{eff} 12.5°C) and the Spanish cave site (T_{eff} 9.5°C) succeeded in yielding both OC and DNA. Hence, although limited by sample size, a preliminary analysis of the factors that affect burial diagenesis, such as temperature and site type, appeared to indicate a greater absence of OC in the sites of warmer climate. Interestingly, the success of aDNA retrieval appeared to be less affected by burial temperature than OC retrieval.

The Asx D/L data from these samples showed some correlation between OC and aDNA survival, where no sample of Asx D/L >0.140 was successful for either molecule (Chapter 2). Where the survival of aDNA appeared to bear some relationship to amino acid concentration, in which the samples that failed aDNA amplification appeared to have lower concentrations than those samples that were successful (although not quite significant), the survival of OC did not correlate with amino acid concentration (several samples with amino acid concentrations as high as those of modern samples failed OC screening). It was also observed that those samples that failed both OC and aDNA amplification had amino acid compositions indicative of the persistence of bone collagen beyond both biomolecules.

Attempts at OC extraction from chicken MBM steam-heated (3 bar pressure) at 133°C, 137°C, 141°C and 145°C for 20 minutes resulted with the observations of an intact OC $M+H^+$ peak in the spectrum from the lowest temperature (133°C) rendered sample, and no signal for the $M+H^+$ in any of the higher temperature rendered samples, despite several purification steps (Chapter 3). The results of OC analyses following the removal of gelatine from the MBM by ultrafiltration indicates that the apparent lack of OC was not due to LC interferences from gelatine, although the presence of other biomolecules in the MBM that were not adequately removed, such as lipids and starches, may have been causing some signal suppression. The OC molecule appears to be hydrolysed at temperatures greater than 133°C under these specific MBM rendering conditions (i.e., 3 bar pressure, 20 min). This is consistent with the match to the hydrophobic carboxy-terminus fragment peptide of OC in the LC-MALDI search results of the three higher temperature rendered samples. With the failure to observe the OC $M+H^+$ in the sedimented bone chips from the lowest temperature MBM sample, which also had high crystallinity values indicative of mineral alteration (Wess *et al.* unpublished data), it is likely that the OC protein was not remaining bound to the bone mineral as originally hypothesised. It is plausible that the alteration of the bone apatite due to the presence of high pressure water leads to partial destruction of the bone mineral/protein composite and liberation of the OC into solution where it is decomposed.

12.1.2 The Investigation into the Use of Other Non-Collagenous Proteins

As it became evident in the early stages of this research project that OC was not the ideal species-specific marker as it was initially proposed, investigations into other potential protein markers became necessary. The main difficulties in analysing potential species-specific proteins in bone was considered to be the removal of NCPs from the largely dominant collagen (I) protein, which makes up 90-95% of the proteins in bone (by weight). Simple gel-electrophoretic methods are notoriously inapplicable to the separation of proteins due to the various molecular weight fragments of hydrolysed collagen in degraded (archaeological and heated) bone (Tuross and Stathoplos 1993). Thus methods of removing collagen from bone samples were investigated. The most common methods for the removal of particular NCPs was via forms of anionic exchange with size exclusion liquid chromatography, particularly for the isolation of small NCPs like decorin, biglycan, osteopontin, SPARC, etc. In experimental work undertaken by the author, but beyond the remit of this thesis, the anionic exchange chromatography methods of Goldberg and Sodek (1994), were used to remove the more basic collagen (although somewhat dependent upon extraction methods) from the more acidic NCPs that interact more with the anion exchange column. Following the isolation of the bound NCP fractions, the proteins were digested and analysed by MALDI-MS for protein identification. However, analyses of the NCP fractions for modern and unheated cattle and chicken bone samples showed that collagenous peptides remained present in the NCP fraction. Samples of chicken MBM rendered at four increasing temperatures 133-145°C, were also analysed to investigate potentially species-informative peptides. Although the peptides observed were not present in great enough S/N to obtain adequate MS/MS spectra, a decrease in the bound NCP fraction with increasing MBM samples was apparent, which suggested that these NCPs were also being degraded in a similar way to OC.

Early considerations were also given to the serum protein albumin, due to the amount of literature pertaining to its long-term survival (Cattaneo *et al.* 1992; Cattaneo *et al.* 1995; Montgelard 1992; Tuross 1989). Sequence analyses showed that serum albumin also exhibits similar amounts of variation between species as OC (Fig. 12.1).

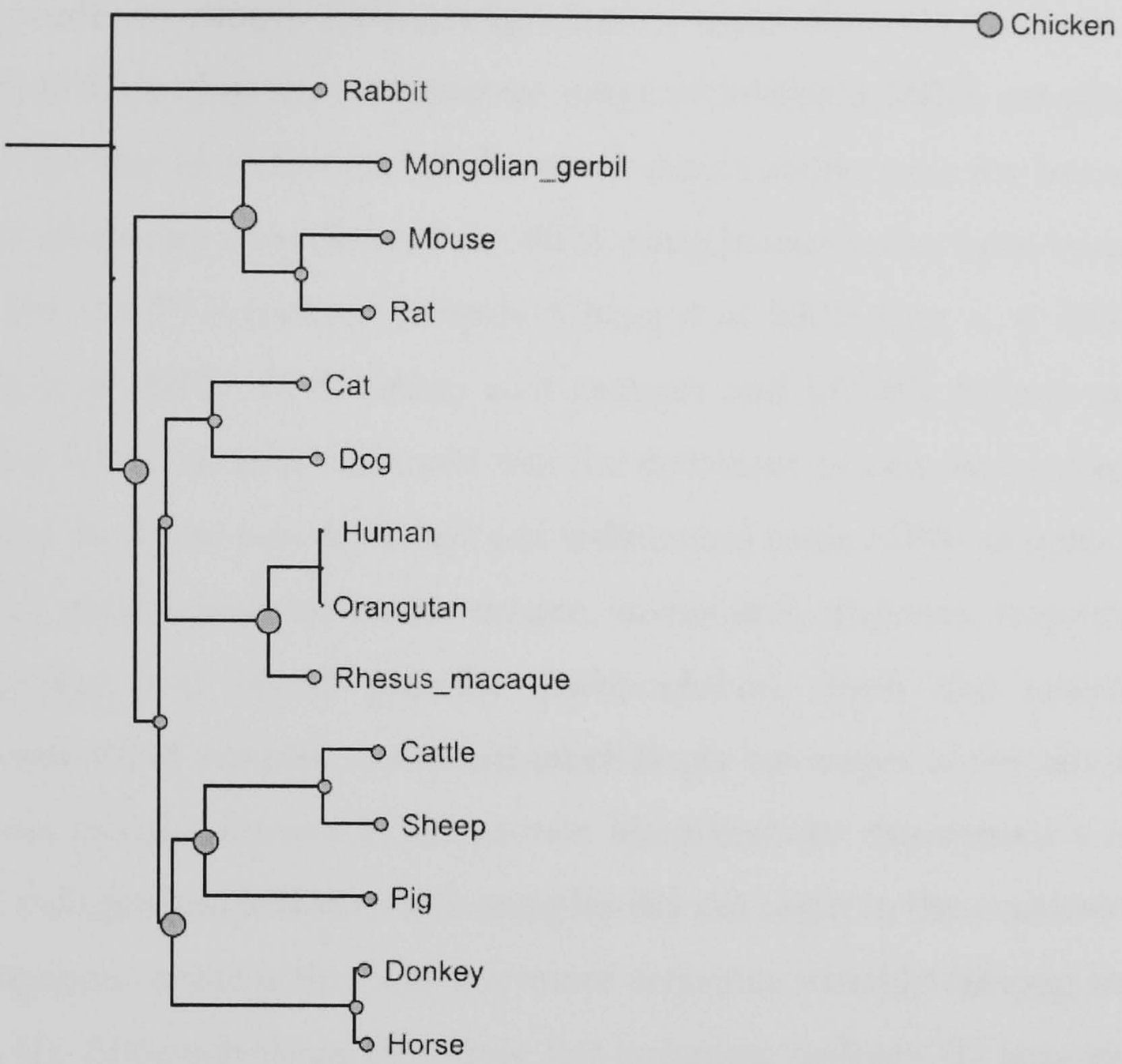


Figure 12.1 - Phylogenetic tree of albumin sequences taken from UniProt protein database and created using Geneious 3.5.6.

However, conventional methods of isolating the protein were time-consuming and costly, due to the minute amounts of the protein in modern bone (and thus even less in ancient bone). Commercially available Albumin/IgG protein extraction cartridges (Calbiochem), designed to remove the protein from serum in order to simplify mixtures, were purchased and methods to extract albumin from bone using these cartridges were investigated. However, the efficiency of these affinity columns varied greatly between species (86% of human and rabbit albumin was found to bind to the column, yet only 37% of cattle albumin was found to bind (Calbiochem 2004) and all attempts at obtaining albumin from modern cattle were unsuccessful. This highlighted the conundrum of attempting to develop affinity-based approaches to isolate a protein or peptide that has a high degree of variability between species, and so only LC (and SPE) based isolation methods using reverse phase monolith or C18 columns were considered.

12.1.3 Identifying the Most Thermostable Protein

Although a simple method of isolating particular proteinaceous markers was desired, it was first necessary to characterise and understand the protein content of heat-treated MBM

(Chapter 4) in order to choose the most appropriate target. Samples rendered at the highest temperature (145°C) within the temperature range of available MBM samples (133-145°C) were analysed for the identification of the most thermostable proteins because numerous other methods of species identification in MBM using biomolecules have been shown to be successful in the ~133°C-rendered samples (Cheng *et al.* 2003; Kim *et al.* 2005; Lahiff *et al.* 2001; Toyoda *et al.* 2004). Both amino acid analyses and LC-MS peptide analyses of the 145°C-rendered MBM samples indicated that the dominant protein was collagen (I), as it is in bone samples. Both the unsedimented and sedimented cattle MBM samples also indicated the presence of muscle proteins (actin, myosin, myoglobin, troponin, tropomyosin), serum proteins (albumin) and blood proteins (haemoglobin). Both the unsedimented and sedimented avian MBM samples, contained much larger coverages of myosin than any other non-collagenous protein. However, the protein identifications throughout a range of these modern bone collagen and MBM residue samples did not result in the consistent observation of a non-collagenous protein that was any more sequence variable (among known species) than collagen (I). Although there were only five complete collagen (I) sequences present in the public protein databases, there appeared to be sufficient species variability to distinguish between the five known mammals (human, cattle, dog, rat and mouse) and hence potential to distinguish between the three mammalian species of economic interest to the agricultural science industry (cattle, pig and sheep). This was subsequently confirmed by sequence compilation of 32 collagen sequences into a local database and subsequent searches using the LC-MS data (Chapter 9). The long-term survival of collagen as the dominant protein in palaeontological bone was also confirmed with amino acid analyses of Cromerian (~600,000 years old) samples and subsequent LC-MS analyses (Chapter 10).

Asx racemisation analyses indicated that extensive denaturation of the proteins in these MBM samples had occurred, even in the acid-insoluble residue thought to consist of only unaltered collagen, likely caused by the extreme conditions of the rendering processes. The Asx racemisation results indicated that its applications to the measure of heat-induced degradation could potentially be used to establish food processing history in MBM. However, the Asx racemisation results (Chapter 3) highlighted the problems with using this particular set of MBM standards, where the pig and sheep MBM 133°C-rendered samples

appeared to be contaminated with a higher temperature rendered sample, consistent with the results from mineral analysis of the MBM sample set (Wess *et al.* unpublished data).

The lowest D/L values of the MBM (the 133°C-rendered cattle and chicken MBM samples and 137°C-rendered pig and sheep MBM samples) were much higher than the proposed cut-offs (Asx D/L 0.10-0.15) for screening out samples unlikely to yield aDNA (Poinar *et al.* 1996; Poinar and Stankiewicz 1999). However, the ability to extract DNA from 133°C-rendered MBM samples (Baeten *et al.* 2004) suggests that the degradation processes of MBM do not appear to mirror the degradation processes in ancient bone, as initially considered. According to the AAR data, the cattle MBM protein appeared significantly more degraded than the chicken MBM protein. This result was unexpected as chicken collagen (I) is considered more soluble than cattle collagen (I) due to a lower concentration in cross-links (Gerstenfeld *et al.* 1994). However, the extent of degradation could have been influenced by the relative ages of the livestock being slaughtered, e.g., young poultry vs. mature cattle (or the precise conditions of heating), the mixtures of such individuals within each MBM sample. Miller *et al.* (1983) reported increased percentages of soluble collagen with the feeding of high-energy diets to mature cattle. This is because during the resulting periods of rapid growth, the rate of protein synthesis is elevated, which results in an increased proportion of newly synthesised collagen (Boleman *et al.* 1996). Newly synthesised collagen contains fewer inter-molecular cross-links, resulting in less stable collagen fibers with higher solubility (McClain and Wiley 1971).

12.1.4 The Analysis of Collagen for Species Identification

At the current state of collagen (I) sequences in the public databases, the ability to distinguish between species using LC-MS peptide analyses was limited to the five known species. A single LC-MS analysis results in large amounts of data to be studied, including peaks analysed by MS/MS, quality of the matched peptide spectra, the ion scores for each peptide match, the total ion scores for the matching proteins, the ranking of protein matches above and below a specific threshold, lists of unmatched peptides, etc (for example see Appendix 3.2, 5.3, & 6.2 listing all information for every LC-MS run). One problem with analysing the data presented as a Mascot result file is that of the top 10 best matches, most

of them are collagen (I) sequences from the α1 and α2 chains of the several species' sequences present in the protein databases (whether UniProt or 'Collagens'). The information required from the search result files is usually only a list of the identified peaks with their respective sequences, as well as a list of the unidentified peaks analysed by MS/MS. Unique peaks that have been identified as belonging to a collagen sequence which is not the highest rank match should also be investigated along with the often large number of unidentified peaks, which is a very important source of information regarding the possibility of species identification as many of the species of interest to the food industry and studies of ancient bone involve species with unknown collagen sequences. The reason behind why these peaks were not matched against any particular sequence should ideally also be further investigated, e.g., whether it is due to poor product ion spectrum quality, incorrect sequence information in the databases, or unknown and complicated PTMs (such as cross-links of unknown additional mass), but the current software is not able to tackle these issues yet. The aims of using LC-MS herein were to investigate the most thermostable proteins that could survive in MBM, and hence long-term survival in ancient bone. Once these initial aims were met, with the identification that collagen (I) is the most appropriate protein to study, alternative approaches that were more reproducible and less costly were investigated.

In order to reduce costs and simplify the data analysis, attempts were made to design methods of collagen (I) peptide isolation, similar to that of OC, whereby cheap SPE cartridges (£2 each sample) could be used in place of using the nanoLC (currently ~£45 each sample) and the analysis time reduced from ~ 3-5 hours/£200-300 per sample to only ~10 seconds/10p per sample. With the knowledge that the α2 (I) chain was much more variable than the α1 (I) chain, a method was developed based on the possibility that the α2 (I) chain would be less affected by the collagen helix-targetting bacterial collagenase, than the α1 (I) chain, and hence perhaps survive digestion more intact. Although the number of observed peptides surviving digestion (in large enough fragments to be analysed adequately by MALDI-MS) was not as great as initially anticipated, this resulted in the ability to isolate a single peptide of the α2 (I) chain, the carboxytelopeptide (Chapter 5).

In order to investigate the same problem of survival as investigated for OC (Chapter 3) for the carboxytelopeptide, analyses were carried out on 19 samples that had previously failed

OC screening (including the 12 samples that failed aDNA analyses). Additional samples of extinct dodo and giant tortoise from Mauritius, also having previously failed aDNA analyses were included in this analysis. The expected carboxytelopeptide peaks were observed in all 21 samples for a 100% success-rate in 'recent' archaeological samples (approx. 4 ka) (Chapter 6). However, although variable enough to distinguish between most taxonomic families of interest, particularly for the MBM industry (Fig. 12.2), the carboxytelopeptide isolated by this method is only 18 amino acid residues long, and marginally more limited in sequence information content than OC (the 49 amino acid OC protein could distinguish between sheep and goat, whereas the 18 amino acid carboxytelopeptide could not).

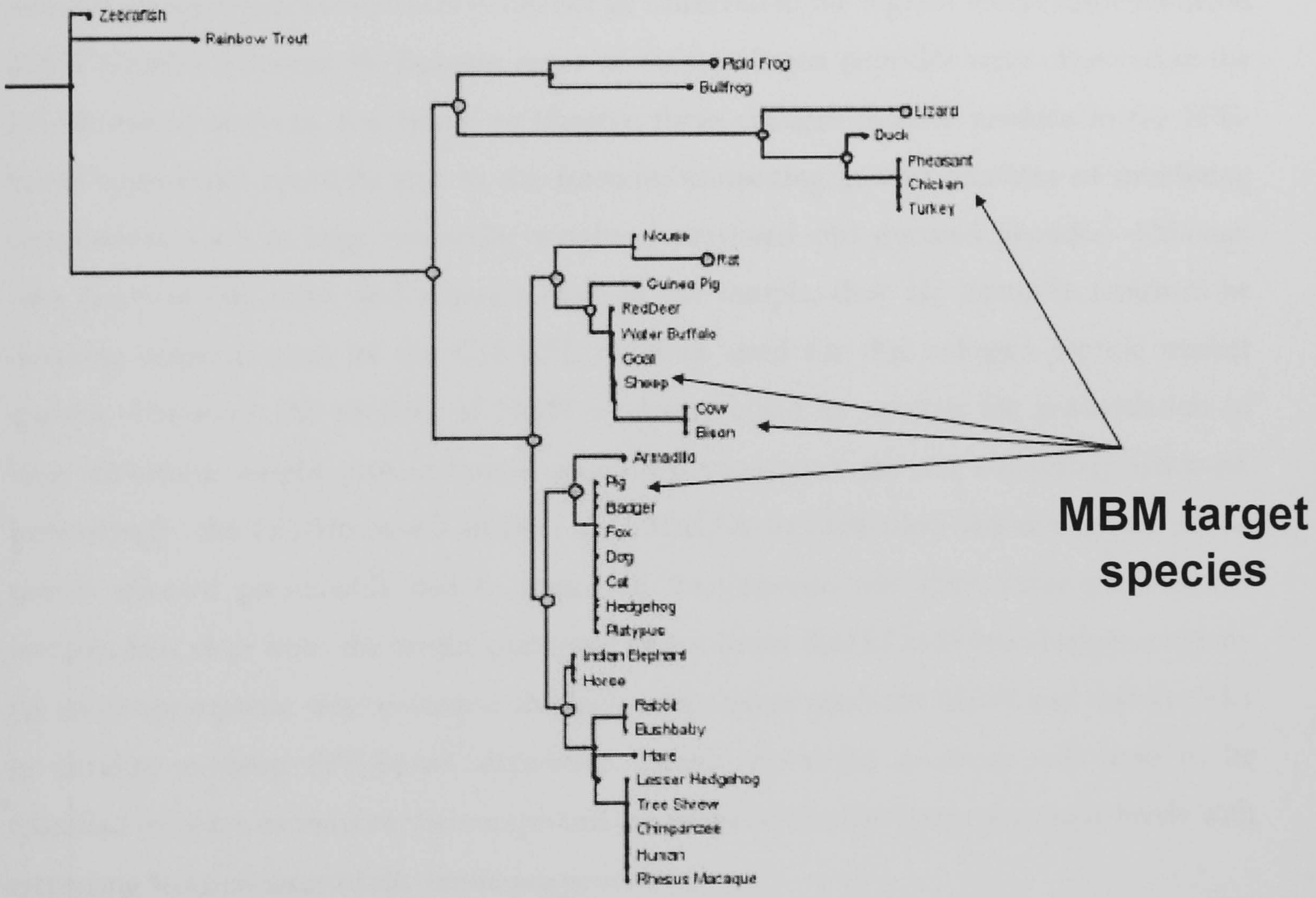


Figure 12.2 - Phylogram of collagen a2 (I) chain carboxy-telopeptide sequences from 36 species (Chapter 5) showing amount of amino acid variation. The four species of interest to the MBM industry are highlighted.

In order to investigate the isolation of collagen-peptide markers further, an alternate enzyme trypsin (used for conventional LC-MS analyses), was used in the speculation that the more variable α2 (I) chain peptides are more hydrophobic than the less variable α1 (I) chain

peptides and that these could also be isolated using an SPE cartridge. Four of these could be isolated reproducibly (Chapter 7), one of which being a 33 amino acid peptide that exhibits two amino acid differences between sheep and goat collagen (I), another of which including much the same α2 (I) chain carboxytelopeptide sequence as observed in the peptide resulting from digestion with bacterial collagenase. The application of this method to distinguishing between sheep and goat in 6000 year old archaeological bones from Domuztepe, South East Turkey, was then investigated, also with a 100% success-rate in the 26 archaeological samples analysed (Chapter 8).

However, despite their survival in archaeological bones beyond OC and aDNA, these selected collagen-peptide markers could not be observed in the highest temperature rendered MBM samples (Chapter 9). Because some of these collagen peptides were observed in the LC-MS-based analyses, the failure to observe these collagen-peptide markers in the SPE-based approaches could be due to the fractions containing greater numbers of interfering compounds, such as large molecular weight glycosylated and glycated peptides. Although salts (such as bile salts) and sugars remain in the sample, they are normally removed by desalting columns such as the C18 SPE columns used for this collagen-peptide marker analysis. However, the process of MBM rendering could be causing the accumulation of large molecular weight protein-bound starch structures that are not adequately removed. Interestingly, the LC-MS-based analyses (LC-MALDI in particular) did not appear to be greatly affected presumably due to improved fractionation and thus fewer peptides per fraction. It is clear from the results presented in this thesis that LC-MS-based approaches are the most appropriate way to analyse the high temperature rendered MBM and that in order to develop a cheap SPE-based alternative, sample extraction methods will need to be modified in order to remove the compounds responsible for the increasing noise levels with increasing temperatures of the rendering processes.

Analyses of the collagen-peptide markers in the palaeontological specimens (Chapter 10) highlighted their limitations, not only in terms of the limited species variability in the five peptides analysed (Peptides A-D & T), but also the poor survival of the carboxytelopeptide (Peptide T) in samples from the early Devensian or older. Although the presence of the collagen-peptide markers from digestion with trypsin (Peptides A-D) in samples dating back

to the Cromerian age supports the long-term survival of ancient collagen (I), the reproducibility of observations of the four peptides was not consistent. The carboxytelopeptide-containing 'helix' peptide maker (Peptide D) was not observed in any of the palaeontological samples (including the Neolithic specimens), but the Peptide C was observed in almost all samples. The other two peptide markers (Peptides A and B), necessary for adequate species-determination in the absence of the carboxytelopeptide (Peptide T), were not observed in the youngest of the palaeontological samples (from Neolithic Carsington Pasture Cave), yet were observed in much older samples. As the carboxytelopeptide fragments were observed in most of these samples, the inconsistency of the 'helix' peptide markers (Peptides A-D) observed was likely due to the laboratory analysis rather than the preservation state of the samples used. Ideally, further method development of the 'helix' peptide markers as described in Chapter 7, would need to be carried out in order to obtain a method suitable for palaeontological samples of different species (of unknown sequence information). This method development may be heavily influenced by the target species of interest, in this case being the ability to distinguish between sheep (*Ovis*) and goat (*Capra*), and could be developed to include peptide markers for distinguishing deer species (as indicated in Chapter 10).

The only alternative approach to collagen peptide analysis to the cheap and relatively simple SPE methods would be the more complex and expensive LC-MS approaches. However, as highlighted in Chapter 9 and Chapter 10, although LC-MS methods yield much more information that will undoubtedly help in species-discrimination, the present approach yields a lot of non-consistent information between samples of the same species (see Appendix 3.2 & 6.2). Comparisons of the LC-MS analysis of three replcate analyses of a single cattle bone specimen (Chapter 4) as well as the comparisons of LC-MS analyses of steppe mammoth (*Mammuthus trogontherii*) and Asian elephant (*Elephas maximus*) (chapter 10) and other elephantid specimens, including African elephant (*Loxodonta africana*), woolly mammoth (*Mammuthus primigenius*) and american mastodon (*Mammut americanum*)(Buckley *et al.* unpublished data) indicated difficulties in identifying novel sequence information. This was primarily due to the poor reproducibility of the peptides that were analysed by MS and identified in the database searches. For example, in well-preserved archaeological samples peptides from non-collagenous proteins are observed along with the collagenous ones, and

discerning them would be much more time-consuming. Additional LC steps could be taken, such as using a strong cationic exchange column prior to the monolith (or C18 SPE) column in order to clean up the sample that could possibly improve the reproducibility, but these will only raise the running time and costs of the methods employed. Perhaps a simpler and more cost-effective approach to removing the variable fractions of the collagenous matrix (including associated NCPs) would be to isolate the gelatine from 'collagen' after several more aggressive extractions. The LC-MS methods used throughout this thesis aimed at retaining acid-insoluble NCPs in order to identify the protein most likely to survive. Following acid demineralisation, instead of washes with distilled and deionised water to raise the pH of the solution prior to gelatinisation (at pH 8.4), solvents such as guanidine hydrochloride with EDTA could be used to remove these NCPs (Jiang *et al.* 2007; Termine *et al.* 1981a).

12.1.5 The Potential Species-Specificity of Protein Sequences

The main advantage of protein sequencing over DNA sequencing in ancient material is that structural proteins, such as the fibrous protein collagen (I), survives much greater lengths of time than does aDNA (Curry 1988) with the disadvantage that proteins have a lower degree of species variability than DNA, not only because three DNA base pairs code for a single amino acid, but because structural constraints which result in stability equally limit the opportunity for non-fatal amino acid substitutions and thus such mutations are less likely to be retained within the genepool. Taking this one step further, the more stable proteins also tend to be the least informative; highest sequence variability is found in structurally unconstrained regions, but as we have discovered for both OC and collagen, it is these regions which are the first to be lost. As a consequence of there only being very few published collagen (I) sequences, investigations into the maximum species variability of proteins could be estimated from analyses of other proteins such as albumin (Fig. 12.1) and myoglobin (Fig. 12.3). An analysis of these protein sequences indicates that the taxonomic resolution of these proteins is similar to that of OC. It is likely that species of distinct genera, such as sheep and goat, red deer and roe deer, etc., exhibit variations in protein sequences, which is consistent with the analyses of collagen (I) hererin. However, different species of

the same genus, such as cattle and yak or horse and zebra, are not likely to exhibit amino acid variations in protein sequence, supported by known OC and myoglobin sequences.

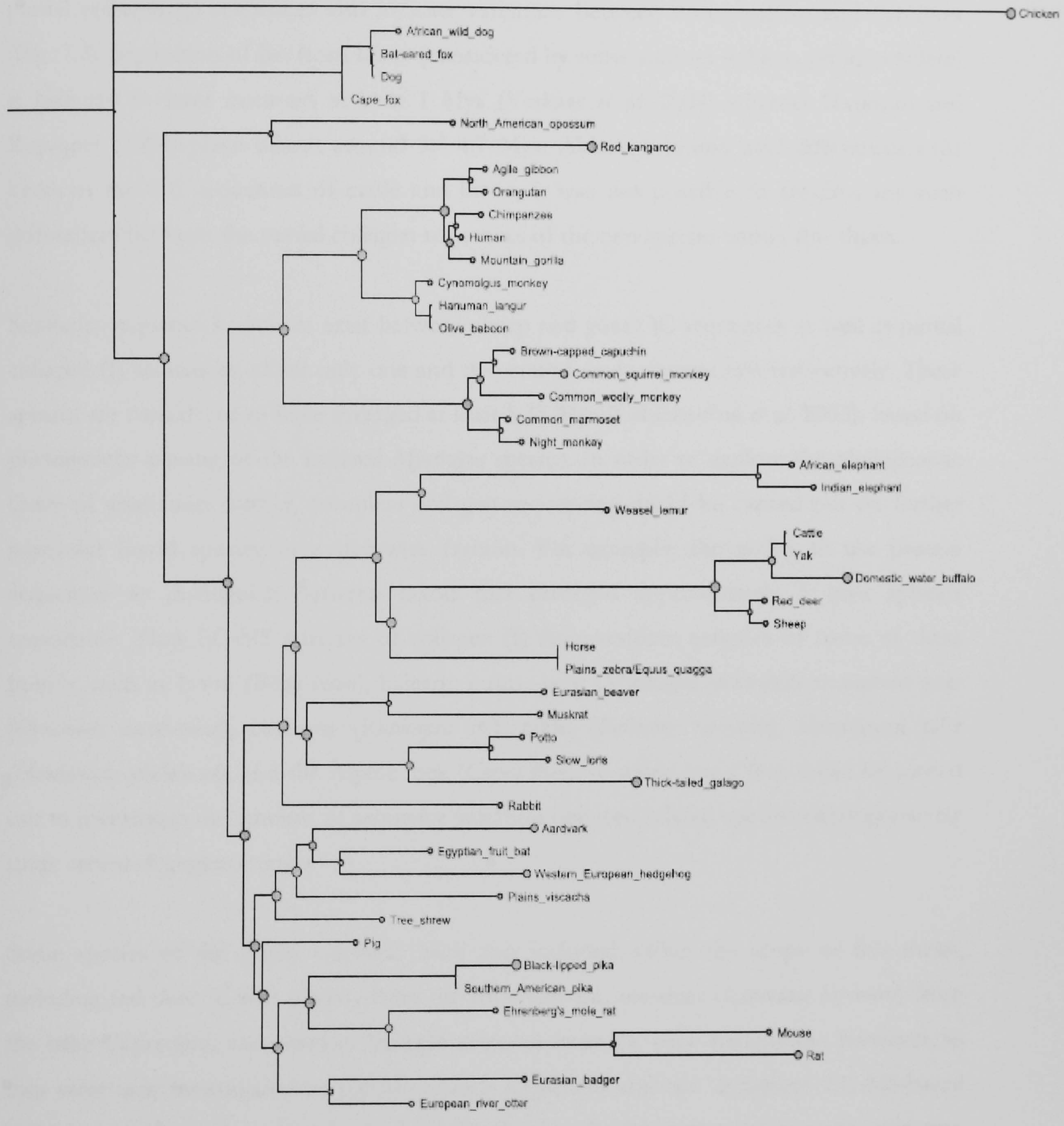


Figure 12.3 - Phylogenetic analysis of selected myoglobin sequences obtained from UniProt created using Geneious 3.5.6

This thesis focused primarily on the distinction between bones of domesticated bovid species such as cattle, sheep and goat using biomolecules. Within the tribe Bovini, the earliest divergence 5 to 10 Mya separated water buffalo (*Bubalus bubalis*), anoa (*Bubalus*

depressicornis), and African buffalo (*Syncerus caffer*) from the *Bos* and *Bison* species (Hassanin and Ropiquet 2004; Janecek *et al.* 1996). As well as the myoglobin sequences (Fig. 12.3), partial collagen (I) sequences also indicate variations between water buffalo and *Bos/Bison* (Fig. 7.4). Separation of *Bos* from *Bison* (considered by some authors to be as yet incomplete) is believed to have occurred at least 1 Mya (Verkaar *et al.* 2004) whereas Hassanin and Ropiquet (2004) place this at around 3.7-4.7 Mya. Although amino acid differences exist between the OC sequences of cattle and bison, it was not possible to confirm any such differences between the partial collagen sequences of the two species within this thesis.

Similarly, sequence variations exist between sheep and goat OC sequences as well as partial collagen (I) sequences, albeit only one and two amino acid substitutions respectively. These species are considered to have diverged at least 5.35 Mya (Lalueza-Fox *et al.* 2005), based on phylogenetic placing of the isolated *Myotragus* species. In order to explore the phylogenetic limits of speciation further, complete collagen sequencing could be carried out on further particular bovid species in a step-wise fashion. For example, the ability to use protein sequences to distinguish between taxon that diverged approximately 5 Mya appears reasonable. Thus LC-MS analyses of collagen (I) from modern samples of some of these bovids, such as bison (*Bison bison*), balearic mouse-goat (*Myotragus balearicus*), mountain goat (*Oreamnos americanus*), chamois (*Rupicapra sp.*), takin (*Budorcas taxicolor*), Himalayan tahr (*Hemitragus jemlahicus*), and the Alpine ibex (*Capra ibex*), to name but a few, could be carried out to investigate the amount of sequence variation between related species of progressively more recent divergence times.

Some species of the family Cervidae were also included within the scope of this thesis, including red deer (*Cervus elaphus*) from the tribe Cervini, roe deer (*Capreolus capreolus*) from the tribe Capreolini, and reindeer (*Rangifer tarandus*) from the tribe Rangiferini. However, as they were only investigated by partial collagen sequences and not 'complete' LC-MS-based collagen peptide analyses, it was not possible to state the limitations of sequence variation within this family. With the partial sequences alone it was possible to demonstrate variations in collagen (I) sequence between the three tribes, where the tribes Capreolini and Rangiferini of the sub-family Capreolidae split approximately 7 Mya, at a similar estimated divergence time as the sub-family Cervini (Pitra *et al.* 2004). As described for the possible further

investigations into proteinaceous species variations amongst the bovids, similar investigations into particular deer species could also be carried out. The identification of deer species from archaeological and palaeontological sites can also be problematic as they can be relatively conservative morphologically (Lister 1996), thus the separation of fragmented material is often difficult. In the range of 'medium sized' deer commonly encountered in European contexts, the most persistent problems arise from the separation of red deer and fallow deer (Lister 1996). As Lister (1996) points out, although the body size of red deer is generally larger than that of the fallow deer, the size of large male fallow deers can overlap with small female red deers and thus using size alone is insufficient to separate remains in many contexts. Both species also show a significant variation in size, both in their geographical range and through time in the Pleistocene and the Holocene (Legge and Rowley-Conwy 1988; Lister 1981). In the British Palaeolithic, fallow deer from late Middle Pleistocene sites such as Swanscombe and Hoxne were larger than today, while the red deer was unusually small (Lister 1986). The two genera *Dama* and *Cervus* appear to have diverged just after the start of the Pliocene around 5.1 Mya (Pitra *et al.* 2004), and thus complete collagen analyses could potentially identify variations elsewhere in the protein. Investigating collagen (I) sequences of deer species may be a more appropriate use of the study of ancient proteins than for bovid species because some species appear to have deeper divergence times, such as European fallow deer (*Dama dama*) and Persian fallow deer (*Dama mesopotamica*), with estimates at around 4-5 Mya (Pitra *et al.* 2004).

The mammalian collagens are highly conserved, as illustrated by sequence analysis and amino acid composition, where placental land mammals show little sequence variation. The whale, on the other hand, shows striking increases in the hydroxyamino acids such as serine and threonine, compared with the land mammals (Eastoe 1955). Even higher levels of these amino acids are found in fish collagens (Neuman 1949; Neuman and Logan 1950), combined with decreased levels of hydroxyproline and proline. Whales, which have long been known as descedents of land mammals, have probably lived in the sea since Eocene times or earlier (Young 1962). The high content of hydroamino acids in whale collagen may therefore be attributed to the long-term effect of a marine diet (Eastoe 1955). Fish collagens have lower thermal stability than mammalian collagens because fish collagens contain lower imino acid contents than mammalian collagens (Kittiphattanabawon *et al.* 2005) and thus

much more variable (Ciarlo *et al.* 1997; Jongjareonrak *et al.* 2005; Yata *et al.* 2001; Zhang *et al.* 2007).

12.1.6 Screening by Amino Acid Analyses

The analysis of many archaeological and palaeontological samples via both SPE-isolated collagen-peptide marker methods, as well as some by LC-MS approaches, was investigated for Chapter 10. The use of amino acid analyses for screening out samples unlikely to yield collagen was investigated. The lowest 'collagen' yield measured in an ancient sample was 0.3%, which had a glycine concentration of ~150 nmol/mg. Using this as an example, assuming 2 mg is necessary for all collagen analyses described within this thesis, approximately 600 mg bone powder would be necessary for the analyses (where samples with ~1% of modern sample glycine concentrations require ~200 mg bone powder). A sample such as WRE2 (*M. trogontherii*) with a glycine concentration of ~50 nmol/mg, would thus need about 1800 mg. Although this appears to be slightly overestimated, it is a similar value to the amounts of sample used. A sample that offered no yield, such as HM2 with glycine concentration ~ 3-5, would require approximately ten-fold this amount, thus about 20 grams of material would be needed to obtain the required 2 mg acid-insoluble 'collagen'. The results indicate that although no yields were obtained using as much as 1 gram of bone, collagen appears to still be present (as determined by amino acid composition analyses) in such small amounts that approximately 20 grams of starting material would be required for some of the samples screened in Chapter 10. This amount of sample destruction was not only considered unrealistic because of the damage induced but also impractical for our current laboratory protocols. Interestingly, using Asx D/L as a screening tool for likely survival of collagen (e.g., Poinar *et al.* 1996) did not appear justified because samples with Asx D/L as high as 0.605 appeared to have amino acid compositions resembling that of collagen. Amino acid concentrations appear to be more useful as a screening tool for the presence/absence of collagen amenable to LC-MS analyses.

If we were to consider much older material, such as the 68 million year old sample analysed by Asara *et al.* (2007), we would anticipate much lower amounts of endogenous protein surviving. Indeed this is the case as Asara and Schweitzer (2008) point out that the remaining

collagen was in the sub-femtomole per milligram range, several orders of magnitude lower than the Pleistocene samples analysed within this thesis. Problems appear to arise at such low levels with regards to contamination as they report that amino acid analyses are no longer fruitful due to the abundance of other contaminant proteins, such as from soil bacteria and human keratin (Asara and Schweitzer 2008). In the case of extra-ordinary examples of collagen (I) survival beyond the expected limits (Table 1.2), such as in Mesozoic fossils (Asara *et al.* 2007), authentication criteria should be put in place for ancient protein analyses as there are for aDNA analyses (Chapter 11). However, Asara and Schweitzer (2008) point out that when collagen (I) is no longer the dominant protein, amino acid analyses are no longer effective to support authenticity and use microscopy and immunological methods instead, the latter of which are questionable due to issues with false positive results (see Montgelard *et al.* 1997). Replication in other laboratories is also more difficult for ancient protein analyses because not only are there very few laboratories worldwide that are solely dedicated to ancient biomolecule analysis, but sample yields of such ancient fossils are very low. The results themselves can be considered a criterion for authentication, namely, if the results are consistent with *a priori* expectations based on the known sample information. With regards to protein sequencing, which can be somewhat more subjective to interpretation than aDNA sequencing (see Chapter 11), this point becomes less clear. For example, Asara *et al.* (2007) identified 'collagen' peptides amongst a mixture of non-collagenous (contaminant) peptides. Not only did they neglect to explain how these collagen peptides were determined to be collagen, but key post-translational modifications were misinterpreted; in many positions, they considered glycine to be hydroxylated (which is not known to occur in any collagen type sequence). In addition to mistakes in interpretation resulting from a reliance on database search parameters, all instances of deamidation of asparagines (N) and glutamines (Q) could be misinterpreted as aspartic acid (D) and glutamic acid (E) respectively. Such misinterpretations of sequence thus have an effect on any subsequent phylogenetic analyses (see Chapter 11). Thus some background knowledge regarding collagen sequences and potential modifications, or likely amino acid substitutions would greatly reduce the number of possible errors in sequence interpretation. Misinterpretations like these will become less likely as the number of collagen sequences increase, a number of additional approximate collagen α1 (I) and α2 (I) sequences are attached in Appendix 5.1.

12.2 Concluding Remarks

The similarities in the protein composition between the unsedimented and sedimented samples, both being primarily composed of collagen (I), indicates that bone is the dominant tissue present in MBM. However, the presence of highly-racemised collagen in the 'acid-insoluble' fractions indicates extensive denaturation and renaturation. Although the potential information gained is vast, the obvious limitations of the technique are the extent of protein sequences in the databases, in the case of collagen (I) there are only five complete protein sequences, only one of which (cattle) has a partial map of its hydroxylation sites. As well as problems inherent in analysing species without a known sequence, which will be common in the analysis of archaeological samples and only two of the four MBM species have protein sequences, there are problems with reproducibility of the current technique. The poor reproducibility from different aliquots of the same sample, let alone different samples of the same species, indicates that with the additional complications of MBM rendering processes (and burial diagenesis in the case of ancient material) modifying peptides, either improved genomic sequences of all possible species needs to be achieved, or the methods for extracting the 'collagen' sample and isolating peptides from it needs to be improved for applications of species identification.

Sequence identification has proven useful in analysing taxonomic identifications in both modern and ancient specimens, most commonly using DNA (Fleischer *et al.* 2000; Greenwood *et al.* 1999; Nielsen-Marsh *et al.* 2002; Thomas *et al.* 1989; Yang *et al.* 2005). However, the older the archaeological/palaeontological material, the more likely that depurination effects of bone diagenesis will have occured throughout the aDNA present in the sample. More recently, the longer-term survival of bone proteins over aDNA are being demonstrated, however, poor results of OC survival (and the relatively minute amounts of other potentially species-informative NCPs) indicate that this long-term survival is probably limited to the major structural protein collagen (I). As indicated by the species-variability of collagen sequences shown within this thesis, there is clearly potential for methods using collagen (I) to be readily applied to answer palaeontological and archaeozoological questions in fragmentary bone specimens.

The question remains is what method is the most appropriate; LC-MS proteomics of the 'collagen' residue, or isolating select few collagen (I) peptides for use as species-specific peptide markers. With the sequence limitations and thus data handling issues of the collagen-peptide marker methods described (Chapter 5-10), the LC-MS proteomics approach is perhaps most appropriate for important unidentified specimens. However, this can be costly and thus not appropriate for the analyses of large numbers of samples, which is likely the ideal approach in the analysis of fragmentary faunal remains from most archaeological sites. The SPE-isolated peptide marker approach is ideal for the analyses of large numbers of samples, but would benefit from some prior knowledge regarding the potential species, such as discriminating between sheep/goat specimens; the SPE-based collagen-peptide marker approach would not be suitable if the samples to be analysed are completely indeterminate using morphological criteria (see Chapter 10). The advantages of carrying out LC-MS-based approaches over the SPE-based approaches are the increased sequence coverage of the protein, as can be seen from the phylogenetic analyses in Chapter 11 compared with those in Chapters 5, 6 and 7. However, for the analysis of many archaeological samples, simple SPE-based identification approaches where only a select few peptides are analysed by MS is more appropriate. The described SPE-based methods could be developed to isolate more peptides of increasing variability using alternate species-distinctions as the target, for example developing the method based on differences between cattle and bison (not yet observed) or differences between red deer and fallow deer (not yet observed).

In the case of archaeolozoology, a combination of both morphological analysis and support from collagen-peptide marker analysis will likely be the most optimum situation. This thesis therefore shows that identifications made using protein sequences (in particular collagen), will either suffer from a) limitations of sequence variability per selected tryptic peptide, or b) expenses of analytical work on 'whole' collagen sequencing. The presented SPE-based approaches were less successful in the analyses of the high temperature rendered (145°C) MBM and thus investigations of this tissue should be carried out using the LC-MS approaches. Where the MBM and food industry may be able to finance more in-depth (LC-MS) analyses, the data analysis involved remains complex and contaminating peptides at the 5% level (where animal tissues of varying species have been purposely mixed) are difficult to distinguish.

12.3 Potential Future Research

Similar to the desire to discriminate between sheep and goat bones, problems also arise with the bones of cetaceans such as porpoise and larger whales. To many past coastal communities in higher latitudes, cetacea were critical resources, both as a source of meat and fat, and as a source of bone and dentine (ivory) for artefact manufacture (Hallén and O'Neill 1994). Relatively undiagnostic bones, such as vertebrae and ribs make up a disproportionate majority of the cetacean skeleton; of 568 cetacean bones recovered from sites in the Western Isles 95% could not be identified to species (Mulville 2002). Different cetacean species have highly characteristic feeding and migration behaviours, making them particularly good indicators of the waters in which they were acquired and the times of year at which hunting or stranding happened. Precise identification of cetacean bone can therefore contribute to understanding the detail of past coastal subsistance strategies. Furthermore, many of the species that were utilised in the past exist today in fragmented and depleted populations. Archaeological records of reliably identified and dated material can make a contribution to mapping the past distribution of species (O'Connor 1996).

A second further application of species identification would be the identification of heavily modified material such as bone tools, or bone fragments remaining in ancient hyaena coprolites. Working bones, antler and ivory into artefacts removes species-diagnostic morphology. Expert examination can discriminate some samples, sometimes (e.g. elephant walrus and hippopotamus ivory) but there would be major benefits from a technique that would allow tentative morphological identifications to be confirmed or refuted by a quite different approach to species identification. In addition, species identification can aid in the interpretation of the artefacts at a site, for instance the use of either reindeer or red deer antler in the manufacture of combs in the Viking North Atlantic (Ambrosiani 1981; Clarke 2002). Bone fragments present in hyaena coprolites are much smaller and much more difficult to speciate by morphological observations alone. If found to preserve well in archaeological cave sites, the identifications of such fragments could be used support current knowledge on what species hyaena predated on in the Devensian and possibly earlier.

Although this thesis does not quite achieve its preferred target in the development of a single method that is both capable of distinguishing between most species found on British archaeological and palaeontological sites, as well as being financially feasible for archaeologists to practise on a large number of specimens, the progress in reaching this target is apparent. Not only has this research identified the most appropriate protein to target for species determination, not only for ancient bone but also for rendered MBM, it has also highlighted the most appropriate form of method development. The ideal method of using collagen (I) peptides for speciation will likely come from a compromise between the LC-MS approach and the SPE-based approach, possibly an improved method of SPE-based peptide isolation that analyses much more than five peptides, either by further elution steps being added. If it is possible to design a financially-affordable method using collagen (I) peptides to speciate between the majority of specimens found on archaeological and palaeontological sites, it would become possible to support or possibly improve our understanding of natural history as well as our understanding of human interaction with animals in the past.

References

Abelson PH. 1956. Paleobiochemistry. *Scientific American 195* (1):83-92.

Agarwal S, and Sohal RS. 1994. Aging and proteolysis of oxidised proteins. *Archives of Biochemistry and Biophysics 309* (1):24-28.

Aiello LC, and Molleson T. 1993. Are more microscopic ageing techniques more accurate than macroscopic ageing techniques. *Journal of Archaeological Science 20*:689-704.

Ajie HO, Hauschka PV, Kaplan IR, and Sobel H. 1991. Comparison of bone collagen and osteocalcin for determination of radiocarbon ages and paleodietary reconstruction. *Earth and Planetary Science Letters 107*(2):380-388.

Ajie HO, and Kaplan IR. 1991. Osteocalcin as the recommended biopolymer for 14C age dating of bone and d13C and d15N palaeodietary reconstruction. *Stable Isotope Geochemistry Geochemical Society, Special Publication (3)*:235-245.

Albarella U. 1999. 'The mystery of husbandry': Medieval animals and the problem of integrating historical and archaeological evidence. *Antiquity 73* (282):867-875.

Ambler R, M. P, and Daniel. 1991. Proteins and molecular palaeontology. *Philosophical Transactions of the Royal Society of London 333*:381-389.

Ambrosiani K. 1981. Viking age combs, comb making and comb makers: in the light of finds from Birka and Ribe: *Stockholm studies in archaeology.*

Anderung C, Bouwman A, Persson P, Carretero JM, Ortega AI, Elburg R, Smith C, Arsuaga JL, Ellegren H, and Gotherstrom A. 2005. Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proceedings of the National Academy of Sciences of the United States of America 102* (24):8431-8435.

Andrews P, Cook J, Currant A, and Stringer C. 1999. *Westbury Cave.* London: Natural History Museum Publications.

Angerbjorn A, Hersteinsson P, Liden K, and Nelson E. 1994. Dietary variation in arctic foxes (Alopex-Lagopus) - an analysis of stable carbon isotopes. *Oecologia 99* (3-4):226-232.

Archibald RM. 1945. Chemical characteristics and physiological roles of glutamine. *Chemical Reviews 37* (2):161-208.

Armstrong WG. 1966. Amino-acid components in fossil calcified tissues. *Nature 210* (5035):481-482.

Armstrong WG, Halstead LB, Reed FB, and Wood L. 1983. Fossil proteins in vertebrate calcified tissues. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences B301* (1106):301-343.

Asara JM, and Schweitzer MH. 2008. Response to comment on "Protein sequences from mastodon and tyrannosaurus rex revealed by mass spectrometry". *Science 319* (5859):33d.

Asara JM, Schweitzer MH, Freimark LM, Phillips M, and Cantley LC. 2007. Protein sequences from mastodon and tyrannosaurus tex revealed by mass spectrometry. *Science 316* (5822):280-285.

Bada J, Wang X, and Hamilton H. 1999. Preservation of key biomolecules in the fossil record: current knowledge and future challenges. *Royal Society of London Philosophical Transactions series B Biological Sciences 354* (1379):77-86.

Bada JL. 1985. Amino acid racemization dating of fossil bone. *Annual Review of Earth and Planetary Science 13*:241-268.

Bada JL, Herrmann B, Payan IL, and Man EH. 1989. Amino acid racemization in bone and the boiling of the German Emporer Lothar I. *Applied Geochemistry 4*:325-327.

Bada JL, Kvenvolden KA, and Peterson E. 1973. Racemization of amino acids in bones. *Nature 245* (5424):308-310.

Bada JL, and Schroeder RA. 1975. Amino acid racemization reactions and their geochemical implications. *Naturwissenschaftern 62*:71-79.

Bada JL, and Shou M-Y. 1980. Kinetics and mechanisms of amino acid racemization in aqueous solution and bones. In: Hare PE, Hoering TC, and King KJ, (Eds.) *Biogeochemistry of amino acids*. New York: Wiley, p 235-255.

Bada JL, Wang XS, Poinar HN, Paabo S, and Poinar GO. 1994. Amino acid racemization in amber-entombed insects: implications for DNA preservation. *Geochimica et Cosmochimica Acta 58* (14):3131-3135.

Baeten V, Reiner AM, Sinnaeve G, and Dardenne P. 2001. Analysis of feedingstuffs by near-infrared microscopy (NIRM): detection and quantification of meat and bone meal (MBM). *Proceedings of the 6th International Symposium on Food Authenticity and Safety.*

Baeten V, von Holst C, Garrido A, Vancutsem J, Renier AM, and Dardenne P. 2005. Detection of banned meat and bone meal in feedstuffs by near-infrared microscopic analysis of the dense sediment fraction. *Analytical and Bioanalytical Chemistry 382* (1):149-157.

Bailey AJ. 2001. Molecular mechanisms of ageing in connective tissues. *Mechanisms of Ageing and Development 122* (7):735-755.

Bailey JF, Richards MB, Macaulay VA, Colson IB, James IT, Bradley DG, Hedges REM, and Sykes BC. 1996. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proceedings of the Royal Society of London Series B-Biological Sciences 263* (1376):1467-1473.

Balasse M, and Ambrose SH. 2005. Distinguishing sheep and goats using dental morphology and stable carbon isotopes in C4 grassland environments. *Journal of Archaeological Science 32* (5):691-702.

Bar-Gal GK, Ducos P, and Horwitz LK. 2003. The application of ancient DNA analysis to identify Neolithic caprinae: a case study from the site of Hatoula, Israel. *International Journal of Osteoarchaeology 13*:120–131.

Bartsiokas A, and Middleton AP. 1992. Characterization and dating of recent and fossil bone by X-ray diffraction. *Journal of Archaeological Science 19*:63-72.

Behnam K, Murray SS, Whiteledge JP, and Brochmann EJ. 2002. Identification of the molecular chaperone alpha B-crystallin in demineralised bone powder and osteoblast-like cells. *Journal of Orthopaedic Research 20*:1190-1196.

Behrensmeyer AK, Kidwell SM, and Gastaldo RA. 2000. Taphonomy and palaeobiology. *Palaeobiology 26*: 103-147.

Behrensmeyer AK. 1978. Taphonomic and ecologic information from bone weathering. *Paleobiology 4* (2):150-162.

Bell LS, Skinner MF, and Jones SJ. 1996. The speed of post mortem change to the human skeleton and its taphonomic significance. *Forensic Science International 82*(2):129-140.

Bella J, Brodsky B, and Berman HM. 1995. Hydration structure of a collagen peptide. *Structure 3* (9):893-906.

Bellagamba F, Moretti VM, Comincini S, and Valfre F. 2001. Identification of species in animal feedstuffs by Polymerase Chain Reaction - restriction fragment length polymorphism analysis of mitochondrial DNA. *Journal of Agricultural Food Chemistry 49* :3775-3781.

Bencsura A, Navale V, Sadeghi M, and Vertes A. 1997. Matrix–Guest Energy Transfer in Matrix-assisted Laser Desorption. *Rapid Communications in Mass Spectrometry 11*(6):679-682.

Bennett RH, Ransom B, Kastner M, Baerwald RJ, Hulbert MH, Sawyer WB, Olsen H, and Lambert MW. 1999. Early diagenesis: impact of organic matter on mass physical properties and processes, California continental margin. *Marine Geology 159* (1-4):7-34.

Berg RA, and Prockop DJ. 1973. The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochemistry and Biophysics Research Communications 52* (1):115-120.

Berna, F., Matthews, A., and Weiner, S. 2004. Solubilities of bone mineral from archaeological sites: the recrystalisation window. *Journal of Archaeological Science 31*: 867-882.

Biemann K. 1988. Contributions of mass spectrometry to peptide and protein structure. *Biomedical and Environmental Mass Spectrometry 16*:99–111.

Bodnar WM, Blackburn RK, Krise JM, and Moseley MA. 2003. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *Journal of the American Society for Mass Spectrometry 14* (9):971-979.

Boessneck J. 1969. Osteological differences between sheep (*Ovis aries* Linné) and goat (*Capra hircus* Linné). In: Brothwell D, and Higgs E, (Eds.). *Science in Archaeology.* New York: Praeger, p 331-358.

Boessneck J, Müller HH, and Teichert M. 1964. Osteologische unterscheidungsmerkmale zwischen schaf (*Ovis aries* Linne) and ziege (*Capra hircus* Linne). *Kuhn-Archives* (78):1–129.

Boleman SJ, Miller RK, Buyck MJ, Cross HR, and Savell JW. 1996. Influence of realimentation of mature cows on maturity, color, collagen solubility, and sensory characteristics. *Journal of Animal Science 74*(9):2187.

Boskey AL. 2007. Mineralization of bones and teeth. *Elements 3*:385-391.

Boyde, A. and Jones, S. 1998. Aspects of anatomy and development of bone: the nm, µm and nm heirarchy. *Advances in organ biology 5A*: 3-44.

Bradley, RS. 1985. *Quaternary palaeoclimatology: methods of palaeoclimtaic reconstruction.* London: Chapman and Hall.

Bradshaw RHW, Hannon GE, and Lister AM. 2003. A long-term perspective on ungulate–vegetation interactions. *Forest Ecology and Management 181*(1-2):267-280.

Brain CK, and Sillen A. 1988. Evidence from the Swartkrans Cave for the earliest use of fire. *Nature 336* (6198):464-466.

Brain, CK. 1981. *The hunters or the hunted?: an introduction to African cave taphonomy.* London: University of Chicago Press.

Brain, CK. 1976. Some principles in the interpretation of bone accumulations associated with man. In Isaac G LI., and McCown E., (Eds). *Human origins: Louis Leakey and the East African evidence.* Menlo Park, California: Staples Press, p 97-116.

Brambilla G, Fumiere O, Aarts HJM, Vaccari G, Von Holst C, Portetelle D, Bouw B, Chiappini B, BDubois M, Georges M and others. 2004. PCR detection of animal by-products in feeds. In Dardenne P, (Ed.) *STRATFEED: Strategies and methods to detect and quantify mammalian tissues in feedingstuffs.* Luxembourg: European Communities.

Breci LA, Tabb DL, Yates JR, and Wysocki VH. 2003. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Analytical Chemistry 75* (9):1963–1971.

Brennan TV, and Clarke S. 1993. Spontaneous degradation of polypeptides at aspartyl and asparaginyl residues - effects of the solvent dielectric. *Protein Science 2* (3):331-338.

Breuker K, Knochenmuss R, and Zenobi R. 1999. Proton transfer reactions of matrix-assisted laser desorption/ionization matrix monomers and dimers. *Journal of the American Society for Mass Spectrometry 10* (11):1111-1123.

Bridgland DR, Bowen DQ, and Wimbledon WA. 1994. *Quaternary of the Thames.* London: Chapman & Hall.

Briggs DEG. 1993. Fossil biomolecules. *The Biochemist 15* (4):8-12.

Brodsky B. 1999. Hydrogen bonding in the triple-helix. *Proceedings of the Indian Academy of Sciences-Chemical Sciences 111* (1):13-18.

Brodsky B, and Ramshaw JAM. 1997. The collagen triple-helix structure. *Matrix Biology 15* (8-9):545-554.

Brown P, and Gajdusek DC. 1991. Survival of Scrapie virus after 3 years' interment. *Lancet (British edition) 337* (8736):269-270.

Brown RS, and Lennon JJ. 1995. Mass resolution improvement by incorporation of Pulsed Ion Extraction in a Matrix-Assisted Laser Desorption/Ionization Linear Time-of-Flight Mass Spectrometer. *Analytical Chemistry 67* (13):1998-2003.

Brown T. 2002. *Genomes.* New York: John Wiley & Sons Inc..

Brown TA, and Brown K. 1992. Ancient DNA and the archaeologist. *Antiquity 66*:10-23.

Brown TA, and Brown KA. 1994. Ancient DNA - using molecular biology to explore the past. *Bioessays 16* (10):719-726.

Brown TA, Nelson DE, Vogel JS, and Southon JR. 1988. Improved collagen extraction by modified Longin method. *Radiocarbon 30*:171-177.

Brunauer LS, and Clarke S. 1986. Age-dependent accumulation of protein residues which can be hydrolysed to d-aspartic acid in human erythrocytes. *Journal of Biological Chemistry 261*:12538-12543.

Buckley M. 2004. *Proteomic studies of osteocalcin bone protein using MALDI mass spectrometry for archaeozoological applications* [MSc Thesis]. Sheffield: University of Sheffield.

Burger J, Rosendahl W, Loreille O, Hemmer H, Eriksson T, Gotherstrom A, Hiller J, Collins MJ, Wess T, and Alt KW. 2004. Molecular phylogeny of the extinct cave lion Panthera leo spelaea. *Molecular Phylogenetics and Evolution 30* (3):841-849.

Campbell S, Carter E, Healey E, Anderson S, Kennedy A, and Whitcher S. 1999. Emerging complexity on the Kahramanmaras plain, Turkey: the Domuztepe project, 1995-1997. *American Journal of Archaeology 103* (3):395-418.

Carbone VA, and Keel BC. 1985. Preservation of plant and animal remains. In: Gilbert Jr RI, and Mielke JH, (Eds.) *The analysis of prehistoric diets.* Orlando, Florida: Academix Press.

Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, and Nesvizhskii A. 2004. The need for guidelines in publication of peptide and protein identification data working group on publication guidelines for peptide and protein identification data. *Molecular & Cellular Proteomics 3*:531-533.

Calbiochem, 2008. User Protocol 122640 Rev. 16-February-04 CML http://search.cosmobio.co.jp/cosmo_search_p/search_gate2/docs/CBC_/122640.20050207.pdf. Accessed 8th December 2007.

Cattaneo C, Gelsthorpe K, Phillips P, and Sokol RJ. 1990. Blood in ancient human bone. *Nature 347*:339.

Cattaneo C, Gelsthorpe K, Phillips P, and Sokol RJ. 1992. Reliable identification of human albumin in ancient bone using ELISA and monoclonal-antibodies. *American Journal of Physical Anthropology 87*:365-372.

Cattaneo C, Gelsthorpe K, Phillips P, and Sokol RJ. 1995. Differential survival of albumin in ancient bone. *Journal of Archaeological Science 22*:271-276.

Cech NB, and Enke CG. 2001. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrometry Reviews 20*(6):362.

Chen FC, Hseih Y-H, and Bridgman RC. 2004. Monoclonal antibody-based sandwich Enzme-Linked Immunosorbent Assay for sensitive detection of prohibited ruminant proteins in feedstuffs. *Journal of Food Protection 67*(3):544-549.

Cheng YH, Wen CM, Ding ST, Kao CC, and Kuo TY. 2003. Detecting meat-and-bone meal in ruminant's feeds by species-specific PCR. *Journal of Animal and Feed Sciences 12* (4):849-858.

Chiappini B, Brambilla G, Agrimi U, Vaccari G, Aarts HJM, Berben G, Frezza D, and Giambra V. 2005. Real-time polymerase chain reaction approach for quantitation of ruminant-specific DNA to indicate a correlation between DNA amount and meat and bone meal heat treatments. *Journal of Aoac International 88* (5):1399-1403.

Child AM, and Pollard AM. 1991. Microbial attack on collagen. In: Pernicka E, and Wagner GA, (Eds.) *Archaeometry '90*,. Basel: BirkhŠuser Verlag. p 617 - 626.

Chu ML, de Wet W, Bernard M, Ding JF, Morabito M, Myers J, Williams C, and Ramirez F. 1984. Human pro alpha 1 (I) collagen gene structure reveals evolutionary conservation of a pattern of introns and exons. *Nature 310*:337-340.

Ciarlo A, Paredi M, and Fraga A. 1997. Isolation of soluble collagen from hake skin (Merluccius hubbsi). *Contribución 191* 65-77.

Clarke D. 2002. Beyond typology: combs, economics, symbolism and regional identity in late norse Scotland. *Norwegian Archaeological Review 35*(2):81-93.

Clutton-Brock J. 1979. The mammalian remains from the Jericho Tell. *Proceedings of the Prehistoric Society 45*:135-157.

Collins MJ, Child AM, van Duin ATC, and Vermeer C. 1998a. Ancient osteocalcin; the most stable bone protein? *Ancient Biomolecules 2*:223-238.

Collins MJ, Gernaey AM, Nielsen-Marsh CM, Vermeer C, and Westbroek P. 2000. Slow rates of degradation of osteocalcin: Green light for fossil bone protein? *Geology 28* (12):1139-1142.

Collins MJ, Nielsen-Marsh CM, Hiller J, Smith CI, Roberts JP, Prigodich RV, Weiss TJ, Csapo J, Millard AR, and Turner-Walker G. 2002. The survival of organic matter in bone: A review. *Archaeometry 44*:383-394.

Collins MJ, and Riley M. 2000. The interpretation of aspartic acid racemization data. In: Goodfriend GA, Collins MJ, Fogel ML, Macko SA, and Wehmiller JF, (Eds.) *Perspectives in Amino Acid and Protein Geochemistry.* New York: Oxford University Press, p 120-144.

Collins MJ, Riley MS, Child AM, and Turner-Walker G. 1995. A basic mathematical simulation of the chemical degradation of ancient collagen. *Journal of Archaeological Science 22*:175-183.

Collins MJ, Waite ER, and van Duin ACT. 1999. Predicting protein decomposition: the case of aspartic-acid racemization kinetics. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 354* (1379):51-64.

Collins MJ, Walton D, and King A. 1998. The geochemical fate of proteins. *Nitrogen-Containing Macromolecules in the Bio- and Geosphere 707*:74-87.

Colombo G, Fanti P, Yao CH, and Malluche HH. 1993. Isolation and complete amino acid sequence of osteocalcin from canine bone. *Journal of Bone and Mineral Research 8* (6):733-743.

Colson IB, Bailey JF, Vercauteren M, and Sykes BC. 1997. The preservation of ancient DNA and bone diagenesis. *Ancient Biomolecules 1*:109-117.

Coope GR, Gibbard PL, Hall AR, Preece RC, Robinson JE, and Sutcliffe AJ. 1997. Climatic and environmental reconstructions based on fossil assemblages from Middle Devensian (Weichselian) deposits of the River Thames at South Kensington, Central London, UK. *Quaternary Science Reviews 16* (10):1163-1195.

Croft P. 2003. The animal bones. In: Peltenburg E, (Ed.) *The colonisation and settlement of Cyprus: investigations at Kissonerga-Mylouthkia.* Savedalen: Paul Astroms Forlag, p 49-56.

Csapo J, Csapo-Kiss Zs, Csapo J, and Jr. 1998. Use of amino acids and their racemisation for age determination in archaeometry. *Trends in Analytical Chemistry 17*:140-148.

Currant A, and Jacobi R. 2001. A formal mammalian biostratigraphy for the Late Pleistocene of Britain. *Quaternary Science Reviews 20* (16):1707-1716.

Currey JD. 1984. *The Mechanical Adaptations of Bones.* Princeton: Princeton University Press.

Currey JD. 2006. *Bones: structure and mechanics.* Princeton: Princeton University Press.

Curry GB. 1988. Amino acids and proteins from fossils. In: Broadhead, T.W., (Ed.) *Molecular evolution and the fossil record:* Paleontological Society, p 20 - 33.

Dardenne P. 2000. *Strategies and methods to detect and quantify mammalian tissues in feedingstuffs.* Luxembourg: European Communities.

Davis SJM. 1985. A preliminary report of the fauna from Hatoula: a Natufian Khiamian (PPNA) site near Latroun, Israel. In: Lechevallier M, and Ronen A, (Eds.) *Le site Natoufien-Khiamien de Hatoula press de Latroun, Israel:* Centre de Recherchesfrancais de Jerusalem, p 71–98.

Dawson PH. 1995. *Quadrupole mass spectrometry and its applications.* London: Springer-Verlag.

de Hoffmann E, and Stroobant V. 2001. *Mass Spectrometry: Principles and Applications.* Chichester: Wiley.

Delmas PD, Tracy RP, Riggs BL, Mann K, and . 1984. Identification of the non collagenous proteins of bovine bone by two-dimensional gel electrophoresis. *Calcified Tissue International 36*:308-316.

DeNiro MJ, and Weiner S. 1988. Organic matter within crystalline aggregates of hydroxyapatite: a new substrate for stable isotopic and possibly other biochemical analyses of bone. *Geochimica et Cosmochimica Acta 52*:2415-2423.

Dole M, Mack LL, Hines RL, Mobley RC, Ferguson LD, and Alice MB. 1968. Molecular beams of macroions. *Journal of Chemical Physics 49*:2240-2247.

Dongre AR, Somogyi A, and Wysocki VH. 1996. Surface-induced dissociation: an effective tool to probe structure, energetics and fragmentation mechanisms of protonated peptides. *Journal of Mass Spectrometry 31*(4):339-350.

Dowd TL, Rosen JF, Li L, and Gundberg CM. 2003. The three-dimensional structure of bovine calcium ion-bound osteocalcin using HNMR spectroscopy. *Biochemistry 42*(25):7769-7779.

Dreisewerd K. 2003. The desorption process in MALDI. *Chemical Reviews 103* (2):395-425.

Ducy P, Desbois C, Boyce B, Pinero G, Story B, Dunstan C, Smith E, Bonadio J, Goldstein S, Gundberg C and others. 1996. Increased bone formation in osteocalcin-deficient mice. *Nature 382* (6590):448-452.

Dungworth G, Vinken NJ, and Schwartz AW. 1975. Amino acid compositions of Pleistocene collagens. *Comparative Biochemistry Physiology 51B*:331-335.

Eastoe JE. 1955. The amino acid composition of mammalian collagen and gelatin. *Biochemical Journal 61*:589-602.

EC. 2004. Animal feed safety in Food Safety from the Farm to the Fork. European Comission.

Efremov, JA. 1940. Taphonomy: new branch of palaeontology. *Pan-American Geologist 74*: 81-93.

Eglington G, and Logan GA. 1991. Molecular preservation. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences 333*:315-328.

Elster H, Gil-Av E, and Weiner S. 1991. Amino acid racemization of fossil bone. *Journal of Archaeological Science 18*:605-617.

Emiliani C. 1955. Pleistocene Temperatures. *Journal of Geology 63*:538-578.

Evershed RP, Turner-Walker G, Hedges REM, Tuross N, and Leyden A. 1995. Preliminary results for the analysis of lipids in ancient bone. *Journal of Archaeological Science 22*(2):277-290.

Evershed RP, and Tuross N. 1996. Proteinaceous material from potsherds and associated soils. *Journal of Archaeological Science 23* (3):429-436.

Eyre DR, Dickson, I.R. and Van Ness, K. 1988. Collagen cross-linking in human bone and articular cartilage; Age-related changes in the content of mature hydroxypyridinium residues. *Biochemical Journal 252*:495-500.

Fenn JB. 1993. Ion formation from charged droplets: roles of geometry, energy, and time. *Journal of the American Society for Mass Spectrometry 4*(7):524-535.

Fietzek PP, and Kuhn K. 1976. The primary structure of collagen. *International Reviews of Connective Tissue Research 7*:1-60.

Fisher LW, Termine JD, Dejter JSW, Whitson SW, Yanagishita M, Kimura JH, Hascall VC, Kleinman HK, Hassel JR, and Nilsson B. 1983. Proteoglycans of developing bone. *The Journal of Biological Chemistry 258*(10):6588-6594.

Fleischer RC, Olson SL, James HF, and Cooper AC. 2000. Identification of the extinct Hawaiian eagle (Haliaeetus) by mtDNA sequence analysis. *Auk 117*(4):1051-1056.

Foster JD, Hope J, and Fraser H. 1993. Transmission of Bovine Spongiform Encephalopathy to sheep and goats. *Veterinary Record 133* (14):339-341.

Franzen A, and Heinegard D. 1985. Isolation and characterization of two sialoproteins present only in bone calcified matrix. *Biochemical Journal 232* (15):715-724.

Fraser FC, and King JE. 1954. Faunal remains. In J.G.D. Clarke (Ed.) *Excavations at Star Carr.* Cambridge: Cambridge University Press, p. 70–95.

Fratzl P, Groschner M, Vogl G, Plenk H, Jr., Eschberger J, Ftatzl-Zelman N, Koller K, and Klaushofer K. 1992. Mineral crystals in calcified tissues: a comprehensive study by SAXS. *Journal of Bone Mineral Reserach 3*:329-334.

Frazao C, Simes DC, Coelho R, Alves D, Williamson MK, Price PA, Cancela ML, and Carrondo MA. 2005. Structural evidence of a fourth Gla residue in fish osteocalcin: biological implications. *Biochemistry 44* (4):1234-1242.

Garland AN. 1987. A histological study of archaeological bone decomposition. In Boddington A, Garland AN, and Janaway RC, (Eds.) *Death, decay & reconstruction approaches to archaeology and forensic science.* Manchester: Manchester University Press, p 109-126.

Garrido-Varo A, Perez-Marin DC, Guerrero JE, Gomez-Cabrera A, von Holst C, Murray I, Van Raamsdonk L, and Zegers J. 2005. Construction of the Stratfeed sample bank and preparation of sample sets. In Dardenne P, (Ed.) *Stratfeed: strategies and methods to detect and quantify mammalian tissues in feedingstuffs.* Luxembourg: European Communities.

Geiger T, and Clarke S. 1987. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *Journal of Biological Chemistry 262* (2):785-794.

Geigl E-M. 2002. On the circumstances surrounding the preservation and analysis of very old DNA. *Archaeometry 44* (3):337-342.

Gendreau MA, Krishnaswamy S, and Mann KG. 1989. The interaction of bone Gla protein (osteocalcin) with phospholipid vesicles. *Journal of Biological Chemistry 264* (12):6972-6978.

Georgsson G, Sigurdarson S, and Brown P. 2006. Infectious agent of sheep scrapie may persist in the environment for at least 16 years. *Journal of General Virology 87*:3737-3740.

Gerstenberger J, Hummel S, Schultes T, Hack B, and Herrmann B. 1999. Reconstruction of a historical genealogy by means of STR analysis and Y-haplotyping of ancient DNA. *European Journal of Human Genetics 7* (4):469-477.

Gerstenfeld L, Feng M, Gotoh T, and Glimcher M. 1994. Selective extractability of noncollagenous proteins from chicken bone. *Orthopedic Surgical Forum 55* (3):230-235.

Gizzi GE, Von Holst C, Baeten V, Berben G, and Van Raamsdonk L. 2004. Determination of processed animal proteins, including meat and bone meal, in animal feed. *Journal of Aoac International 87*(6):1334-1341.

Glimcher MJ. 1989. Mechanism of calcification: Role of collagen fibrils and collagen-phosphoprotein complexes in vitro and in vivo. *The Anatomical Record 224* (2):139-153.

Glimcher MJ. 2006. Bone: nature of the calcium phosphate crystals and cellular, structural, and physical chemical mechanisms in their formation. *Reviews in Mineralogy and Geochemistry 64* (1):223-282.

Glimcher MJ, Cohen-Solal L, Kossiva D, and de Ricqles A. 1990. Biochemical analyses of fossil enamel and dentin. *Paleobiology 16*:219-232.

Glimcher MJ, and Krane SM. 1968. The organization and structure of bone and the mechanism of calcification. In Ramachandran GN, and Gould BS, (Eds.) *Biology of Collagen.* New York: Academic Press.

Glowacki J, Rey C, Glimcher MJ, Cox KA, and Lian J. 1991. A role for osteocalcin in osteoclast differentiation. *Journal of Cellular Biochemistry 45* (3):292-302.

Goldberg HA, and Sodek J. 1994. Purification of mineralized tissue-associated osteopontin. *Journal of Tissue Culture Methods 16*:211-215.

Goodfriend GA, Flessa KW, and Hare PE. 1997. Variation in amino acid epimerization rates and amino acid composition among shell layers in the bivalve Chione from the Gulf of California. *Geochimica et Cosmochimica Acta 61*(7):1487-1493.

Gordon CC, and Buikstra JE. 1981. Soil, pH, bone preservation, and sampling bias at mortuary sites. *American Antiquity 46* (3):566-571.

Gotherstrom A, Collins MJ, Angerbjorn A, and Liden K. 2002. Bone preservation and DNA amplification. *Archaeometry 44* (3):395-404.

Grant A. 1982. *The use of tooth wear as a guide to the age of domestic ungulates.* Oxford: British Archaeological Reports: British Series. 91-108 p.

Greenwood AD, Capelli C, Possnert G, and Paabo S. 1999. Nuclear DNA sequences from late Pleistocene megafauna. *Molecular Biology and Evolution 16*(11):1466-1473.

Greenwood AD, Castresana J, Feldmaier-Fuchs G, and Paabo S. 2001. A molecular phylogeny of two extinct sloths. Molecular Phylogenetics and Evolution 18(1):94-103.

Griffth JS. 1967. Cell replication and scrapie. *Nature 215*:1043-1044.

Grine FE, Krause DW, Fosse G, and Jungers WL. 1987. Analysis of individual, intraspecific and interspecific variability in quantitative parameters of caprine tooth enamel structure. *Acta Odontologica Scandinavica 45* (1):1-23.

Grupe G. 1995. Preservation of collagen in bone from dry, sandy soil. *Journal of Archaeological Science 22* (2):193-199.

Gurley LR, Valdez JG, Spall WD, Smith BF, and Gillette DD. 1990. Proteins in the fossil bone of the dinosaur seismosaurus. *Journal of Protein Chemistry 10*:75-90.

Hackett CJ. 1981. Microscopical focal destruction (tunnels) - in exhumed human bones. *Medicine, Science and the Law 21*:243-265.

Hagelberg E, Bell LS, Allen T, Boyde A, Jones SJ, and Clegg JB. 1991. Analysis of ancient bone DNA - techniques and applications. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 333* (1268):399-407.

Haile J, Holdaway R, Oliver K, Bunce M, Gilbert MTP, Nielsen R, Munch K, Ho SYW, Shapiro B, and Willerslev E. 2007. Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Molecular Biology and Evolution 24* (4):982-989.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series 41*:95-98.

Hallén Y, and O'Neill M. 1994. The use of bone and antler at Foshigarry and Bac Mhic Connain, two Iron Age sites on North Uist, Western Isles. *Proceedings of the Society of Antiquaries of Scotland 124*:189-231.

Halstead P. 1996. Pastoralism or household herding? Problems of scale and specialization in early Greek animal husbandry. *World Archaeology 28* (1):20-42.

Halstead P, Collins P, and Isaakidou V. 2002. Sorting the sheep from the goats: morphological distinctions between the mandibles and mandibular teeth of adult Ovis and Capra. *Journal of Archaeological Science 29*:543-553.

Hare PE. 1980. Organic geochemistry of bone and its relation to the survival of bone in the natural environment. In: Behrensmeyer AK, and Hill AP, (Eds.) *Fossils in the making.* Chicago: University of Chicago Press, p 208 - 219.

Hare PE, Miller GH, and Tuross NC. 1975. Simulation of natural hydrolysis of proteins in fossils. *Carnegie Institute of Washington Year Book 74*:609-612.

Hassanin A, and Ropiquet A. 2004. Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, Bos sauveli Urbain 1937. *Molecular Phylogenetics and Evolution 33* (3):896-907.

Hauschka PV. 1977. Quantitative determination of γ-carboxyglutamic acid in proteins. *Analytical Biochemistry 80*:212-223.

Hauschka PV. 1980. Osteocalcin: a specific protein of bone with potential for fossil dating. In: Hare PE, Hoering TC, and King KJ, (Eds.) *Biogeochemistry of Amino Acids.* New York: Wiley, p 75 - 82.

Hauschka PV. 1981. Osteocalcin structure: Ca^{2+}-dependence of a-helical domains. In: Veis A, (Ed.) *The Chemistry and Biology of Mineralised Connective Tissues.* North Holland: Elsevier, p 337 - 341.

Hauschka PV. 1986. Osteocalcin: the vitamin K-dependent Ca^{2+}-binding protein of bone matrix. *Haemostasis 16* (3-4):258-272.

Hauschka PV, and Carr SA. 1982. Calcium-dependant α-helical structure in osteocalcin. *Biochemistry 21*:2538-2547.

Hauschka PV, Lian JB, Cole DE, and Gundberg CM. 1989. Osteocalcin and matrix Gla protein: vitamin K-dependent proteins in bone. *Physiological Reviews 69* (3):990-1047.

Hauschka PV, and Wians FH. 1989. Osteocalcin-hydroxyapatite interaction in the extracellular organic matrix of bone. *Anatomical Record 224*:180-188.

Hedges REM, Millard A, R., and Pike AWG. 1995. Measurements and Relationships of Diagenetic Alteration of Bone from Three Archeological Sites. *Journal of Archaeological Science 22*:201-209.

Hedges REM, and Millard AR. 1995. Bones and groundwater - Towards the modeling of diagenetic processes. *Journal of Archaeological Science 22* (2):155-164.

Heinegard D, Hultenby K, Oldberg A, Reinholt F, and Wendel M. 1989. Macromolecules in bone matrix. *Connective Tissue Research 21*(1-4):3-11.

Helgen KM. 2003. Major mammalian clades: a review under consideration of molecular and palaeontological evidence. *Mammalian Biology 68* (1):1-15.

Higham CFW. 1967. Stock rearing as a cultural factor in prehistoric Europe. *Proceedings of the Prehistoric Society 33*:84-106.

Higuchi R, Bowman B, Freiberger M, Ryder OA, and Wilson AC. 1984. DNA Sequences from the quagga, an extinct member of the horse family. *Nature 312*:282-284.

Hirsila M, Koivunen P, Gunzler V, Kivirikko KI, and Myllyharju J. 2003. Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *Journal of Biological Chemistry 278* (33):30772-30780.

Hoang QQ, Sicheri F, Howard AJ, and Yang DSC. 2003. Bone recognition mechanism of porcine osteocalcin from crystal structure. *Nature 425* (6961):977-980.

Hodge AJ. 1989. Molecular models illustrating the possible distributions of 'holes' in simple systematically staggered arrays of type I collagen molecules in native-type fibrils. *Connective Tissue Research 21*(1-4):137-147.

Hofmann HJ. 1973. Stromatolites: characteristics and utility. *Earth Science Reviews 9*:339-373.

Hole F, Flannery KV, and Neely JA. 1969. *Prehistory and human ecology of the Deh Luran plain: an early village sequence from Khuzistan, Iran.* Ann Arbor: University of Michigan.

Horton HR, Moran LA, Ochs RS, Rawn JD, and Scrimgeour KG. 2002. *Principles of Biochemistry.* New Jersey:Prentice Hall.

Chen FC. and Hsieh YHP, 2002. Monoclonal antibodies against troponin I for the detection of rendered muscle tissues in animal feedstuffs. *Meat Science 62*: 405-412.

Hubert JF, Panish PT, Chure DJ, and Prostak KS. 1996. Chemistry, microstructure, petrology, and diagenetic model of Jurassic dinosaur bones, Dinosaur National Monument, Utah. *Journal of Sedimentary Research 66* (3):531-547.

Hulmes DJ, Miller A, Parry DA, Piez KA, and Woodhead-Galloway J. 1973. Analysis of the primary structure of collagen for the origins of molecular packing. *Journal of Molecular Biology 79 (*1):137-148.

Hulmes GM. 1992. The collagen superfamily - diverse structures and assemblies. *Essays in Biochemistry 27*:49-67.

Hummel S, and Herrmann B. 1996. aDNA typing for reconstruction of kinship. *Homo 47*(1-3):215-222.

Huq NL, Teh LC, Christie DL, and Chapman GE. 1984. The amino acid sequences of goat, pig and wallaby osteocalcins. *Biochemistry International 8* (4):521-527.

Huq NL, Tseng A, and Chapman GE. 1990. Partial amino acid sequence of osteocalcin from an extinct species of ratite bird. *Biochemistry International 21*:491-496.

Hyland D, Tersak J, Adovasio J, and Siegel M. 1990. Identification of the species of origin of residual blood on lithic material. *American Antiquity 55*:104-112.

IHGSH. 2004. Finishing the euchromatic sequence of the human genome. *Nature 431*:931-945.

Iribarne JV, and Thomson BA. 1976. On the evaporation of small ions from charged droplets. The *Journal of Chemical Physics 64*:2287.

Jackson DS, and Bentley JP. 1960. On the significance of the extractable collagens. *The Journal of Cell Biology 7*(1):37-42.

Janecek LL, Honeycutt RL, Adkins RM, and Davis SK. 1996. Mitochondrial gene sequences and the molecular systematics of the artiodactyl subfamily Bovinae. *Molecular Phylogenetics and Evolution 6* (1):107-119.

Jiang X, Ye M, Jiang X, Liu G, Feng S, Cui L, and Zou H. 2007. Method development of efficient protein extraction in bone tissue for proteome analysis. *Journal of Proteome Research 6* (6):2287-2294.

Johnson RE. 1994. Models for matrix-assisted desorption by a laser-pulse. *International Journal of Mass Spectrometry and Ion Processes 139*:25-38.

Jones JS. 1985. Mummified human DNA cloned. *Nature 314*:576.

Jongjareonrak A, Benjakul S, Visessanguan W, Nagai T, and Tanaka M. 2005. Isolation and characterisation of acid and pepsin-solubilised collagens from the skin of Brownstripe red snapper (Lutjanus vitta). *Food Chemistry 93* (3):475-484.

Kadler K. 1994. Extracellular matrix 1: fibril-forming collagens. *Protein Profile 1*(5):519-524.

Kadler KE, Holmes DF, Trotter JA, and Chapman J. 1996. Collagen fibril formation. *Biochemistry Journal 316*:1-11.

Kagan HM, and Li W. 2003. Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell. *Journal of Cellular Biochemistry 88*(4):660-672.

Kansa SW. in prep. Domuztepe mammal remains. In: Campbell S, and Carter E, (Eds.) *In prehistoric Domuztepe.* Los Angeles: Monumenta Archaeologica.

Karas M, Bahr U, Ingendoh A, and Hillenkamp F. 1989. Laser desorption/ionization mass spectrometry of proteins of mass 100 000 to 250 000 Dalton. *Angewandte Chemie International Edition in English 28* (6):760-761.

Karas M, Glueckmann M, and Schaefer J. 2000. Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *Journal of Mass Spectrometry 35* (1):1-12.

Karas M, and Hillenkamp F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry 60*(20):2299-2301.

Kaufman DS, and Manley WF. 1998. A new procedure for determining enantiomeric DL amino acid ratios in fossils using reverse phase liquid chromatography. *Quaternary Science Reviews 17*(11):987-1000.

Kaufmann R. 1995. Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: a novel analytical tool in molecular biology and biotechnology. *Journal of Biotechnology 41*(2-3):155-175.

Kelly J, Tanaka S, Hardt T, Eikenberry EF, and Brodsky B. 1988. Fibril-forming collagens in lamprey. *Journal of Biological Chemistry 263* (2):980-987.

Kim SH, Huang TS, Seymour TA, Wei CI, Kempf SC, Bridgman CR, Clemens RA, and An H. 2004. Identification of a biomarker for the detection of prohibited meat and bone meal residues in animal feed. *Journal of Food Science 69* (9):C739-C745.

Kim SH, Huang TS, Seymour TA, Wei CI, Kempf SC, Bridgman CR, Momcilovic D, Clemens RA, and An HJ. 2005. *Development of immunoassay for detection of meat and bone meal in animal feed. Journal of Food Protection 68* (9):1860-1865.

King K, and Bada JL. 1979. Effect of in situ leaching on amino acid racemisation rates in fossil bone. *Nature 281*:135-137.

Kittiphattanabawon P, Benjakul S, Visessanguan W, Nagai T, and Tanaka M. 2005. Characterisation of acid-soluble collagen from skin and bone of bigeye snapper (Priacanthus tayenus). *Food Chemistry 89* (3):363-372.

Kivirikko KI, and Pihlajaniemi T. 1998. Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Advances in Enzymology and Related Areas of Molceular Biology 72*:325-398.

Klarlund JK, Tsiaras W, Holik JJ, Chawla A, and Czech MP. 2000. Distinct Polyphosphoinositide Binding Selectivities for Pleckstrin Homology Domains of GRP1-like Proteins Based on Diglycine Versus Triglycine Motifs. *Journal of Biological Chemistry 275* (42):32816-32821.

Klug WS, and Cummings MR. 1997. *Concepts of genetics.* Upper Saddle River, N.J.: Prentice Hall.

Knochenmuss R, and Zenobi R. 2003. MALDI ionization: the role of in-plume processes. *Chemical Reviews 103*:441–452.

Knott L, and Bailey AJ. 1998. Collagen cross-links in mineralizing tissues: a review of their chemistry, function, and clinical relevance. *Bone 22*(3):181-187.

Kohler-Rollefson I. 1992. A model for the development of nomadic pastoralism on the transjordanian plateau. In: Bar-Yosef O, and khasanov A, (Eds.) *Pastoralism in the Levant: archaeological materials in anthropological perspectives.* Madison: Prehistory Press, p 11-18.

Koon H. 2006. *Detecting cooked bone in the archaeological record: a study of the thermal stability and deterioration of bone collagen.* Unpublished PhD Thesis. York: University of York.

Koon HEC, Nicholson RA, and Collins MJ. 2003. A practical approach to the identification of low temperature heated bone using TEM. *Journal of Archaeological Science 30* (11):1393-1399.

Krause E, Wenschuh H, and Jungblut PR. 1999. The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Analytical Chemistry 71* (19):4160-4165.

Krcmar P, and Rencova E. 2003. Identification of species-specific DNA in feedstuffs. *Journal of Agricricultural Food Chemistry 51*:7655-7658.

Krenk P. 1991. An overview of rendering structure and procedures in the European Community. In Bradley,R. Savey M & Marchant B, (Eds.) *Sub-acute spongiform encephalopathies.* Kluwer, Dordrecht, p 161-167.

Kronick PL, and Cooke P. 1996. Thermal stabilization of collagen fibers by calcification. *Connective Tissue Research 33* (4):275-282.

Krutchinsky AN, Zhang W, and Chait BT. 2000. Rapidly switchable matrix-assisted laser desorption/ionization and electrospray quadrupole-time-of-flight mass spectrometry for protein identification. *Journal of the American Society for Mass Spectrometry 11* (6):493-504.

Kukkola L, Hieta R, Kivirikko KI, and Myllyharju J. 2003. Identification and characterization of a third human, rat, and mouse collagen prolyl 4-hydroxylase isoenzyme. *Journal of Biological Chemistry 278* (48):47685-47693.

Lahiff S, Glennon M, O'Brien L, Lyng J, Smith T, Maher M, and Shilton N. 2001. Species-specific PCR for the identification of ovine, porcine and chicken species in meat and bone meal (MBM). *Molecular and Cellular Probes 15* (1):27-35.

Lalueza-Fox C, Castresana J, Sampietro L, Marquès-Bonet T, Alcover JA, and Bertranpetit J. 2005. Molecular dating of caprines using ancient DNA sequences of Myotragus balearicus, an extinct endemic Balearic mammal. *BMC Evolutionary Biology 5* (1):70.

Lander, ES, and Waterman, MS., 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics 2*: 231-239.

Landis WJ. 1996. Mineral characterization in calcifying tissues: atomic, molecular and macromolecular perspectives. *Connective Tissue Research 35* (1-4):1-8.

Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, Vigne JD, Edwards CJ, Schlumbaum A, and Dinu A. 2007. From the cover: ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences 104* (39):15276.

Lees S. 1979. A model for the distribution of HAP crystallites in bone - an hypothesis. *Calcified Tissue International 27* (1):53-56.

Legge AJ. 1996. The beginning of caprine domestication in Southwest Asia. In Harris, D.R. (Ed.) *The origins and spread of agriculture and pastoralism in Eurasia.* London: University College London Press p 238-262.

Legge AJ, and Rowley-Conwy P. 1988. *Star Carr revisited: a re-analysis of the large mammals.* London: Birkbeck College.

Leonard JA, Shanks O, Hofreiter M, Kreuz E, Hodges L, Ream W, Wayne RK, and Fleischer RC. 2007. Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science 34* (9):1361-1366.

Lewis MS, and Piez KA. 1964. The characterization of collagen from the skin of the dogfish shark (Squalus acanthias). *Journal of Biological Chemistry 239* (10):3336-3340.

Liardon, R, and Hurrel, RF. 1983. Racemization of amino acids in heated and alkali-treated food proteins. *Journal of Agricultural and Food Chemistry 31*:432-437.

Liggett WH, Jr., Lian JB, Greenberger JS, and Glowacki J. 1994. Osteocalcin promotes differentiation of osteoclast progenitors from murine long-term bone marrow cultures. *Journal of Cell Biochemistry 55* (2):190-199.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature 362:*709-715.

Linnaeus C. 1758. *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis.* Stockholm: Tomus I. Holmiae Laurenti Salvi. 1-824 p.

Lister A. 1981. *Evolutionary studies on Pleistocene deer.* Unpublished PhD Thesis: University of Cambridge.

Lister AM. 1986. New results on deer from Swanscombe, and the stratigraphical significance of deer in the Middle and Upper Pleistocene of Europe. *Journal of Archaeological Science 13*:319–338.

Lister AM. 1992. Mammalian fossils and Quaternary biostratigraphy. *Quaternary Science Reviews 11*:329–344.

Lister AM. 1996. The morphological distinction between bones and teeth of fallow deer (*Dama dama*) and red deer (*Cervus elaphus*). *International Journal of Osteoarchaeology 6* (2):119-143.
Logan GA, Smiley CJ, and Eglinton G. 1995. Preservation of fossil leaf waxes in association with their source tissues, Clarkia, Northern Idaho, USA. *Geochimica et Cosmochimica Acta 59* (4):751-763.
Longin R. 1971. New method of collagen extraction for radiocarbon dating. *Nature 230*:241 - 242.
Loreille O, Vigne JD, Hardy C, Callou C, Treinen-Claustre F, Dennebouy N, and Monnerot M. 1997. First distinction of sheep and goat archaeological bones by the means of their fossil mtDNA. *Journal of Archaeological Science 24* (1):33-37.
Lowenstam HA, and Weiner S. 1989. *On Biomineralization.* New York: Oxford University Press.
Lowenstein JM. 1985. Molecular approaches to the identification of species. *American Scientist 73*:541-547.
Lubda D, Cabrera K, Kraas W, Schaefer C, and Cunningham D. 2001. New developments in the application of monolithic HPLC columns. *Liquid Chromatography - Gas Chromatography 19* (12):1186–1191.
Lyman, RL. 1994. *Vertebrate taphonomy.* Cambridge: Cambridge University Press.
Mainland I, and Halstead P. 2005. The economics of sheep and goat husbandry in Norse Greenland. *Arctic Anthropology 42*:103-120.
Malone JD. 1982. Recruitment of osteoclast precursors by purified bone matrix constituents. The *Journal of Cell Biology 92* (1):227-230.
Mamyrin BA. 2001. Time-of-flight mass spectrometry: concepts, achievements, and prospects. *International Journal of Mass Spectrometry 206* (3):251-266.
Mamyrin BA, Karataev VI, Shmikk DV, and Zagulin VA. 1973. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Soviet Physics Journal of Experimental and Theoretical Physics* 37:45.
Masters PM. 1987. Preferential preservation of non-collagenous protein during bone diagenesis: implications for chronometric and stable isotopic measurements. *Geochimica et Cosmochimica Acta 51*:3209-3214.
Mayr E, Linsley EG, and Usinger RL. 1953. *Methods and principles of systematic zoology.* New York: McGraw-Hill.
McBride DJ, Jr., Choe V, Shapiro JR, and Brodsky B. 1997. Altered collagen structure in mouse tail tendon lacking the alpha 2(I) chain. *Journal of Molecular Biology 270* (2):275-284.
McClain PE, and Wiley ER. 1971. Influence of diet on metabolism and crosslinking in rat skin collagen. *Federation Proceedings 30*:402.
McKinley J. 2000. The analysis of cremated bone. In *Human Osteology in Archaeology and Forensic Science.* Cox, M. and Mays, S. (Eds.) London: Greenwich Medical Media, p 403-421.
McNulty T, Calkins A, Ostrom P, Gandhi H, Gotffried M, Martin L, and Gage DA. 2002. Stable isotope values of bone organic matter: artificial diagenesis experiments and plaeoecology of Natural Trap Cave, Wyoming. *Palaios 17*:36-49.
Miles CA, Avery NC, Rodin VV, and Bailey AJ. 2005. The increase in denaturation temperature following cross-linking of collagen is caused by dehydration of the fibres. *Journal of Molecular Biology 346* (2):551-556.
Miles CA, and Bailey AJ. 1999. Thermal denaturation of collagen revisited. *Proceedings of Indian Academy of Sciences - Chemical Sciences 111* (1):71-80.
Miles CA, and Bailey AJ. 2001. Thermally labile domains in the collagen molecule. *Micron 32* (3):325-332.
Miles CA, Burjanadze TV, and Bailey AJ. 1995. The kinetics of the thermal-denaturation of collagen in unrestrained rat tail tendon determined by differential scanning calorimetry. *Journal of Molecular Biology 245*:437-446.
Miles CA, and Ghelashvili M. 1999. Polymer-in-a-box mechanism for the thermal stabilization of collagen molecules in fibers. *Biophysical Journal 76* (6):3243-3252.
Miles CA, Sionkowska A, Hulin SL, Sims TJ, Avery NC, and Bailey AJ. 2000. Identification of an intermediate state in the helix-coil degradation of collagen by ultraviolet light. *Journal of Biological Chemistry 275* (42):33014-33020.

Millard A. 2001. Deterioration of bone. In: Pollard AM, and Brothwell D, (Eds.) *Handbook of archaeological sciences*. Chichester: John Wiley and Sons.
Miller PE, and Denton MB. 1986. The quadrupole mass filter: basic operating concepts. *Journal of Chemical Education 63*:617-622.
Miller RK, Tatum JD, Cross HR, Bowling RA, and Clayton RP. 1983. Effects of carcass maturity on collagen solubility and palatability of beef from grain-finished steers. *Journal of Food Science 48*:484-486.
Monboisse JC, Bellon G, Randoux A, Dufer J, and Borel JP. 1990. Activation of human neutrophils by type I collagen. *Biochemical Journal 270*:459-462.
Montgelard C. 1992. Albumin preserved in fossil bones and systematics of Malpaisolys insularis (Muridae, Rodentia), an extinct rodent of the Canary Islands. *Historical Biology 6*:293-302.
Montgelard C, Buchy MC, Gautret P, and Dauphin Y. 1997. Biogeochmical characterization of ichthyosaur bones from Holzmaden (Germany, Lias). *Buletin de la Société Géologique de France 168*:759-766.
Muldoon MT, Onisk DV, Brown MC, and Stave JW. 2004. Targets and methods for the detection of processed animal proteins in animal feedstuffs. *International Journal of Food Science and Technology 39*(8):851-861.
Mulville J. 2002. The role of cetacea in prehistoric and historic Atlantic Scotland. *International Journal of Osteoarchaeology 12* (1):34-48.
Murphy, WJ, Eizirik, E, Johnson, WE, Zhang, YP, Ryderk, OA, O'Brien, SJ, 2001. Molecular phylogenetics and the origins of the placental mammals. *Nature 409* (6820): 614-618.
Murray I, Aucott LS, and Pike IH. 2001. Use of discriminant analysis on visible and near infrared reflectance spectra to detect adulteration of fishmeal with meat and bone meal. *Journal of Near Infrared Spectroscopy 9* (4):297-311.
Muyzer G, Sandberg P, Knapen MHJ, Vermeer C, Collins M, and Westbroek P. 1992. Preservation of the bone protein osteocalcin in dinosaurs. *Geology 20* (10):871-874.
Nelsestuen GL, Zytkovicz TH, and Howard JB. 1974. The mode of action of vitamin K identification of caboxyglutamic acid as a component of prothrombin. *Journal of Biological Chemistry 249* (19):6347-6350.
Nesvizhskii AI, and Aebersold R. 2005. Interpretation of shotgun proteomic data the protein inference problem. *Molecular & Cellular Proteomics 4* (10):1419-1440.
Neuberger A. 1948. Stereochemistry of amino acids. *Advances in Protein Chemistry 4*:297 - 383.
Neuman RE. 1949. The amino acid composition of gelatins, collagens and elastins from different sources. *Archives of Biochemistry 24* (2):289-298.
Neuman RE, and Logan MA. 1950. The determination of hydroxyproline. *Journal of Biological Chemistry 184* (1):299-306.
Newman ME, Parboosingh JS, Bridge PJ, and Ceri H. 2002. Identification of archaeological animal bone by PCR/DNA analysis. *Journal of Archaeological Science 29*:77-84.
Nicholson RA. 1993. A morphological investigation of burnt animal bone and an evaluation of its utility in archaeology. *Journal of Archaeological Science 20* (4):411-428.
Nicholson RA. 1996. Bone degradation, burial medium and species representation: debunking the myths, an experiment-based approach. *Journal of Archaeological Science 23* (4):513-533.
Nicholson RA. 1998. Bone degradation in a compost heap. *Journal of Archaeological Science 25*:393-403.
Nielsen-Marsh C. 2002. Biomolecules in fossil remains. *The Biochemist* 12-14.
Nielsen-Marsh CM, and Hedges REM. 1999. Bone porosity and the use of mercury intrusion porosimetry in bone diagenesis studies. *Archaeometry 41*:165-174.
Nielsen-Marsh CM, and Hedges REM. 2000a. Patterns of diagenesis in bone I: the effects of site environments. *Journal of Archaeological Science 27* (12):1139-1150.
Nielsen-Marsh CM, and Hedges REM. 2000b. Patterns of diagenesis in bone II: effects of acetic acid treatment and the removal of diagenetic CO_3^{2-}. *Journal of Archaeological Science 27*:1151-1159.

Nielsen-Marsh CM, Ostrom PH, Gandhi H, Shapiro B, Cooper A, Hauschka PV, and Collins MJ. 2002. Exceptional preservation of bison bones > 55 ka as demonstrated by protein and DNA sequences. *Geology 30* (12):1099–1102.

Nielsen-Marsh CM, Richards MP, Hauschka PV, Thomas-Oates JE, Trinkaus E, Pettitt PB, Karavanic I, Poinar H, and Collins MJ. 2005. Osteocalcin protein sequences of Neanderthals and modern primates. *Proceedings of the National Academy of Sciences 102* (12):4409-4413.

Nogami HMD, Oohira A, and Ogasawara NMD. 1987. Levels of creatine kinase activity in cartilage of tubular and nontubular bone in relation to pathogenesis of achondroplasia. *Clinical Orthopaedics & Related Research 219*:308-312.

Norman D. 1994. *Prehistoric life: the rise of the vertebrates.* London: Boxtree Limited.

Notbohm H, Nokelainen M, Myllyharju J, Fietzek PP, Muller PK, and Kivirikko KI. 1999. Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis *in vitro*. *Journal of Biological Chemistry 274* (13):8988-8992.

O'Connor T. 2000. *The archaeology of animal bones.* Stroud: Sutton Publishing Limited.

O'Connor TP. 1996. A critical overview of archaeological animal bone studies. *World Archaeology 28* (1):5-19.

Ocana MF, Neubert H, Przyborowska A, Parker R, Bramley P, Halket J, and Patel R. 2004. BSE control: detection of gelatine-derived peptides in animal feed by mass spectrometry. *Analyst 129* (2):111-115.

Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, and Ahn NG. 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & Cellular Proteomics 4* (10):1487-1502.

Olumee Z, Sadeghi M, Tang X, and Vertes A. 1995. Amino acid composition and wavelength effects in matrix-assisted laser desorption/ionization. *Rapid Communications in Mass Spectrometry 9*:744-752.

Ono M, Shitashige M, Honda K, Isobe T, Kuwabara H, Matsuzuki H, Hirohashi S, and Yamada T. 2006. Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. *Molecular & Cellular Proteomics 5* (7):1338.

Orgel JP, Wess TJ, and Miller A. 2000. The *in situ* conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen. *Structure with Folding & Design 8* (2):137-142.

Orlando L, Leonard JA, Thenot A, Laudet V, Guerin C, and Hanni C. 2003. Ancient DNA analysis reveals woolly rhino evolutionary relationships. *Molecular Phylogenetics and Evolution 28* (3):485-499.

Ostrom PH, Gandhi H, Strahler JR, Walker AK, Andrews PC, Leykam J, Stafford TW, Kelly RL, Walker DN, Buckley M and others. 2006. Unraveling the sequence and structure of the protein osteocalcin from a 42 ka fossil horse. *Geochimica et Cosmochimica Acta 70* (8):2034-2044.

Ostrom PH, Schall M, Gandhi H, Shen TL, Hauschka PV, Strahler JR, and Gage DA. 2000. New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. *Geochimica et Cosmochimica Acta 64* (6):1043-1050.

Paabo S, Higuchi RG, and Wilson AC. 1989. Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *Journal of Biological Chemistry 264* (17):9709-9712.

Palsson PA. 1979. Rida (scrapie) in Iceland and its epidemiology. *Slow Transmissible Diseases of the Nervous System 1*:357–366.

Payne S. 1973. Kill-off patterns in sheep and goats: the mandibles from Asvan Kale. *Anatolian Studies 23*:281-303.

Payne S. 1985. Morphological distinctions between the mandibular teeth of young sheep, *Ovis* and goats, *Capra*. *Journal of Archaeological Science 12*:139-147.

Penkman KE. 2005. *Amino acid geochronology: a closed system approach to test and refine the UK model.* Unpublished PhD Thesis. Newcastle: University of Newcastle-upon-Tyne.

Perez-Marin DC, Garrido-Varo A, Guerrero-Ginel JE, and Gomez-Cabrera A. 2004. Near-

infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: chemical composition and open-declaration. *Animal Feed Science and Technology 116* (3-4):333-349.

Pitra C, Fickel J, Meijaard E, and Groves C. 2004. Evolution and phylogeny of old world deer. *Molecular Phylogenetics and Evolution 33* (3):880-895.

Poinar HN, Hoss M, Bada JL, and Paabo S. 1996. Amino acid racemization and the preservation of ancient DNA. *Science 272* (5263):864-866.

Poinar HN, and Stankiewicz BA. 1999. Protein preservation and DNA retrieval from ancient tissues. *Proceedings of the National Academy of Sciences of the United States of America 96* (15):8426-8431.

Poser JW, and Price PA. 1979. A method for the decarboxylation of γ-carboxyglutamic acid in proteins. *Journal of Biological Chemistry 254* (2):431-436.

Price PA. 1988. Role of vitamin-K-dependent proteins in bone metabolism. *Annual Reviews of Nutrition 8*:565-583.

Price PA, Urist MR, and Otawara Y. 1983. Matrix Gla Protein, A new gamma-carboxyglutamic acid-containing protein which is associated with the organic matrix of bone. Biochem *Biophysical Research Communications 117*:765-771.

Price PA, and Williamson MK. 1985. Primary structure of bovine matrix Gla protein, a new vitamin K-dependent bone protein.. *Journal of Biological Chemistry 260* (28):14971-14975.

Prigodich RV, and Vesely MR. 1997. Characterization of the complex between bovine osteocalcin and type I collagen. *Archives of Biochemistry and Biophysics 345*:339-341.

Prummel W, and Frisch HJ. 1986. A guide for the distinction of species, sex and body side in bones of sheep and goat. *Journal of Archaeological Science 13* (6):567-577.

Prusiner SB. 1982. Novel proteinaceous infectious particles cause scrapie. *Science 216* (4542):136.

Qin C, Baba O, and Butler WT. 2004. Post-translational modifications of SIBLING proteins and their roles in osteogenesis and dentinogenesis. *Critical Reviews in Oral Biology & Medicine 15* (3):126.

Radkiewicz JL, Zipse H, Clarke S, and Houk KN. 1996. Accelerated racemization of aspartic acid and asparagine residues via succinimide intermediates: an ab initio theoretical exploration of mechanism. *Journal of the American Chemical Society 118* (38):9148-9155.

Rappsilber J, and Mann M. 2002. What does it mean to identify a protein in proteomics? *Trends in Biochemical Sciences 27*(2):74-78.

Reitz EJ, and Wing ES. 1999. *Zooarchaeology*. Cambridge: Cambridge University Press.

Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, Old WM, Cheung HT, Russell S, and Wattawa JL. 2004. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Analytical Chemistry 76* (13):3556–3568.

Rich A, and Crick FH. 1961. The molecular structure of collagen. *Journal of Molecular Biology 3*:483-506.

Richards MB, and Sykes BC. 1995. Authenticating DNA extracted from ancient skeletal remains. *Journal of Archaeological Science 22*:291-299.

Ritz-Timme S, Cattaneo C, Collins MJ, Waite ER, Schutz HW, Kaatsch HJ, and Borrman HIM. 2000. Age estimation: the state of the art in relation to the specific demands of forensic practise. *International Journal of Legal Medicine 113* (3):129-136.

Ritz S, Turzynski A, Schutz HW, Hollmann A, and Rochholz G. 1996. Identification of osteocalcin as a permanent aging constituent of the bone matrix: basis for an accurate age at death determination. *Forensic Science International 770*:13-26.

Robbins LL, Muyzer G, and Brew K. 1993. Macromolecules form living and fossil biominerals; implications for the establishment of molecular phyolgenies. In Engle MH, and Macko SA, (Eds.) *Organic Geochemistry*. New York: Plenum, p 799 - 816.

Roberts MB, Parfitt SA, and Heritage E. 1999. *Boxgrove: a middle Pleistocene hominid site at Eartham Quarry, Boxgrove, West Sussex. London*: English Heritage.

Roberts SJ, Smith CI, Millard AR, and Collins MJ. 2002. The taphonomy of cooked bone: characterising boiling and its physico-chemical effects. *Archaeometry 44* (4):485-494.

Robins SP. 1988. Functional properties of collagen and elastin. *Baillieres Clinical Rheumatology 2*:1-35.

Robins SP, and Duncan A. 1987. Pyridinium crosslinks of bone collagen and their location in peptides isolated from rat femur. *Biochimica Biophysica Acta 914* (3):233-239.

Robinson AB, and Rudd CJ. 1974. Deamidation of glutaminyl and asparaginyl residues in peptides and proteins. In: Einsele G, Ricken W, and Seilacher A, (Eds.) *Current topics in cellular recognition.* New York: Academic Press, p 247 - 294.

Roepstorff P, and Fohlman J. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry 11*(11):601.

Royce PM, and Barnes MJ. 1985. Failure of highly purified lysyl hydroxylase to hydroxylate lysyl residues in the non-helical regions of collagen. *Biochemical Journal 230* (2):475-480.

Sakae T, Mishima H, Kowaza Y, and LeGeros RZ. 1995. Thermal stability of mineralized and demineralized dentin: A differential scanning calorimetric study. *Connective Tissue Research 33*:193-196.

Schaaff TG, Cargile BJ, Stephenson Jr JL, and McLuckey SA. 2000. Ion Trap collisional activation of the (M+ 2H) 2-(M+ 17H) 17 ions of human hemoglobin-chain. *Analytical Chemistry 72* (5):899-907.

Schneider T, Butz P, Ludwig H, and Tauscher B. 2003. Pressure-induced formation of pyroglutamic acid from glutamine in neutral and alkaline solutions. *LWT-Food Science and Technology 36* (3):365-367.

Schramm PT, Johnson CJ, Mathews NE, McKenzie D, Aiken JM, and Pedersen JA. 2006. Potential role of soil in the transmission of prion disease. *Reviews in Mineralogy and Geochemistry 64* (1):135-152.

Schreiweis MA, Butler JP, Kulkarni NH, Knierman MD, Higgs RE, Halladay DL, Onyia JE, and Hale JE. 2007. A proteomic analysis of adult rat bone reveals the presence of cartilage/chondrocyte markers. *Journal of Cell Biochemistry 101*:466-476.

Schreve DC. 2001. Differentiation of the British late Middle Pleistocene interglacials: the evidence from mammalian biostratigraphy. *Quaternary Science Reviews 20*(16-17):1693-1705.

Schwalbe RA, Ryan J, Stern DM, Kisiel W, Dahlback B, and Nelsestuen GL. 1989. Protein structural requirements and properties of membrane binding by gamma-carboxyglutamic acid-containing plasma proteins and peptides. *Journal of Biological Chemistry 264* (34):20288-20296.

Schweitzer MH. 2004. Molecular paleontology: some current advances and problems. *Annales de paléontologie 90* (2):81-102.

Shackleton NJ. 1967. Oxygen isotope analyses and Pleistocene temperatures re-assessed. *Nature 215* (5096):15-17.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J and others. 2004. Rise and fall of the Beringian steppe bison. *Science 306* (5701):1561-1565.

Sherrat A, 1981. Plough and pastoralism: aspects of the Secondary Products Revolution. In *Paterrn of the past: studies in honour of David Clarke* (Ed. I Hodder, G. Isaac and N. Hammond). Cambridge: Cambridge University Press. p 261-305.

Sherratt A. 1983. The secondary exploitation of animals in the Old World. *World Archaeology 15* (1):90-104.

Shipman P, Foster G, and Schoeninger M. 1984. Burnt bones and teeth: an experimental study of color, morphology, crystal structure and shrinkage. *Journal of Archaeological Science 11*:307-325.

Shipman, P. 1981. *Life history of a fossil: an introduction taphonomy and paleoecology*.Cambridge, Massachusetts: Harvard University Press.

Shoshani J, Lowenstein JM, Walz DA, and Goodman M. 1985. Proboscidean origins of mastodont and woolly mammoth demonstrated immunologically. *Paleobiology 11*:429-437.

Shukla AK, and Futrell JH. 2000. Tandem mass spectrometry: dissociation of ions by collisional activation. *Journal of Mass Spectrometry 35* (9):1069-1090.

Sillen A. 1989. Diagenesis of the inorganic phase of cortical bone. In: Price DT, (Ed.) *The chemistry of prehistoric human bone.* Cambridge: Cambridge University Press, p 211 - 229.

Smith BJ. 2002. Chemical Cleavage of Proteins at Aspartyl-X Peptide Bonds. In Walker JM (Ed.) *Protein Protocols Handbook*, The Second Edition. New York: Humana Press Inc, p 499-502.

Smith CI, Craig OE, Prigodich RV, Nielsen-Marsh CM, Jans MME, Vermeer C, and Collins MJ. 2005. Diagenesis and survival of osteocalcin in archaeological bone. *Journal of Archaeological Science 32* (1):105-113.

Smith GG, and Evans RC. 1980. The effect of structure and conditions on the rate of racemization of free and bound amino acids. In: Hare PE, Hoering TC, King K, and Jr., (Eds.) *Biogeochemistry of amino acids.* New York: Wiley, p 257 - 282.

Smith GG, Williams KM, and Wonnacott DM. 1978. Factors affecting the rate of racemization of amino acids and their significance to geochronology. *The Journal of Organic Chemistry 43* (1):1-5.

Sobolik KD. 2003. *Archaeobiology.* Walnut Creek: AltaMira Press.

Spadaro JA, Becker RO, and Bachman CH. 1970. The distribution of trace metal ions in bone and tendon. *Calcified Tissue International 6* (1):49-54.

Stahl DA, Flesher B, Mansfield HR, and Montgomery. 1988. Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. *Applied and Environmental Microbiology 56*:11079-11084.

Stanlotter H, Sulzner M, Egelseer E, Norton CF, and Hochstein LI. 1993. Comparison of Membrane ATPases from Extreme Halophiles Isolated from Ancient Salt Deposits. *Origins of Life and Evolution of the Biosphere 23* (1):53-64.

Steen H, and Mann M. 2004. The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology 5* (9):699-711.

Stiner MC, Kuhn SL, Weiner S, and Bar yosef O. 1995. Differential burning, recrystallization, and fragmentation of archaeological bone. *Journal of Archaeological Science 22* (2):223-237.

Stuart AJ. 1975. The vertebrate fauna of the type Cromerian. *Boreas 4*:63-76.

Stuart AJ. 1976. The history of the mammal fauna during the Ipswichian/last Interglacial in England. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences 276* (945):221-250.

Stuart AJ. 1982. *Pleistocene vertebrates in the British Isles.* London: Longman.

Stuart AJ, and Lister AM. 2001. The mammalian faunas of Pakefield/Kessingland and Corton, Suffolk, UK: evidence for a new temperate episode in the British early Middle Pleistocene. *Quaternary Science Reviews 20* (16):1677-1692.

Stuart AJ, Lister AM, and Egginton J. 1993. *Fossil vertebrates. The Lower Paleolithic Site at Hoxne, England.* Chicago: University of Chicago Press.

Stutz AJ, and Tuross N. 1997. Archaeological information preserved in Pleistocene-age bone protein residues. *Journal of Human Evolution 32* (4):A22-A23.

Sutcliffe AJ. 1964. *The mammalian fauna.* Royal Anthropological Institute of Great Britain and Ireland.

Taflin DC, Ward TL, and Davis EJ. 1989. Electrified droplet fission and the Rayleigh limit. *Langmuir 5* (2):376-384.

Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, and Yoshida T. 1988. Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry 2* (8):151-153.

Tartaglia M, Saulle E, Pestalozza S, Morelli L, Antonucci G, and Battaglia PA. 1998. Detection of bovine mitochondrial DNA in ruminant feeds: a molecular approach to test for the presence of bovine-derived materials. *Journal of Food Protection 61* (5):513-518.

Taylor DM. 1998. Inactivation of the BSE agent. *Journal of Food Safety 18* (4):265-274.

Taylor DM, and Woodgate SL. 2003. Rendering practices and inactivation of transmissible spongiform encephalopathy agents. *Revue Scientifique Et Technique De L Office International Des Epizooties 22* (1):297-310.

Taylor RE, Hare PE, and White CD. 1995. Geochemical criteria for thermal alteration of bone. *Journal of Archaeological Science 22*:115-119.

Termine JD, Belcourt AB, Conn KM, and Kleinmann HK. 1981a. Mineral and collagen-binding proteins of fetal calf bone. *Journal of Biological Chemistry 256* (20):10403-10408.
Termine JD, Kleinman HK, Whitson SW, Conn KM, McGarvey ML, and Martin GR. 1981b. Osteonectin, a bone-specific protein linking mineral to collagen. *Cell 26*:99-105.
Thomas RH, Schaffner W, Wilson AC, and Pääbo S. 1989. DNA phylogeny of the extinct marsupial wolf. *Nature 340*:465-467.
Toyoda A, Nakajo M, Kawachi H, Matsui T, and Yano H. 2004. PCR detection of bovine mitochondrial DNA derived from meat and bone meal in feed. *Journal of Food Protection 67* (12):2829-2832.
Triffitt JT, Gebauer U, Ashton BA, Owen ME, and Reynolds JJ. 1976. Origin of plasma alpha2-HS-glycoprotein and its accumulation in bone. *Nature 262*(5565):226-227.
Turekian KK, and Bada J. 1972. *Calibration of hominid evolution: recent advances in isotopic and other dating methods applicable to the origin of man.* Edinburgh: Scottish Academic Press.
Tuross N. 1989. Albumin preservation in the Taima-taima mastodon skeleton. *Applied Geochemistry 4*:255-259.
Tuross N, Behrensmeyer AK, Eans ED, Fisher LW, and Hare PE. 1989. Molecular preservation and crystallographic alterations in a weathering sequence of wildebeest bones. *Applied Geochemistry 4*:261-270.
Tuross N, and Stathoplos L. 1993. Ancient proteins in fossil bones. *Methods in Enzymology 224*:121-129.
Tzaphlidou M. 2005. The role of collagen in bone structure: an image processing approach. *Micron 36* (7-8):593-601.
Ulrich MM, Perizonius WRK, Spoor CF, Sandberg P, and Vermeer C. 1987. Extraction of osteocalcin from fossil bones and teeth. *Biochemical and Biophysical Research Communications 149*:712-719.
Vaisar T, and Urban J. 1996. Probing the proline effect in CID of protonated peptides. *Journal of Mass Spectrometry 31*:1185-1187.
van Duin ACT, and Collins MJ. 1998. The effects of conformational constraints on aspartic acid racemization. *Organic Geochemistry 29* (5-7):1227-1232.
van Kolfschoten T. 2000. The Eemian mammal fauna of central Europe. *Geologie En Mijnbouw-Netherlands Journal of Geosciences 79* (2-3):269-281.
Veis A. 1964. *The Macromolecular Chemistry of Gelatin.* New York: Academic Press.
Verkaar ELC, Nijman IJ, Beeke M, Hanekamp E, and Lenstra JA. 2004. Maternal and paternal lineages in cross-breeding Bovine species. Has Wisent a hybrid origin? *Molecular Biology and Evolution 21* (7):1165-1170.
Vermeulen P, Baeten V, and Dardenne P. 2005. The STRATFEED project: a European initiative to tackle the problematic detection of MBM in compound feed. In: Dardenne P, (Ed.) *STRATFEED: strategies and methods to detect and quantify mammalian tissues in feedingstuffs.* Luxembourg: European Communities.
Vertes A, Gijbels R, and Levine RD. 1990. Homogeneous bottleneck model of matrix-assisted ultraviolet laser desorption of large molecules. *Rapid Communications in Mass Spectrometry 4*:228–233.
Vertes A, Irinyi G, and Gijbels R. 1993. Hydrodynamic model of matrix-assisted laser desorption mass spectrometry. *Analytical Chemistry 65* (17):2389-2393.
Vertes A, and Levine RD. 1990. Sublimation versus fragmentation in matrix-assisted laser desorption. *Chemical Physics Letters 171*:284–290.
von den Driesch A, and Wodtke U. 1997. The fauna of 'Ain Ghazal, a major PPN and early PM settlement in central Jordan. In: Gebel H, Kafafi Z, and Rollefson G, (Eds.) *The prehistory of Jordan II Perspectives from 1997.* Berlin: Ex Orient, p 511-543.
von Endt DW, and Ortner DJ. 1984. Experimental effects of bone size and temperature on bone diagenesis. *Journal of Archaeological Science 11*:247-253.
von Holst C, Honikel KO, Unglaub W, Kramer G, and Anklam E. 2000. Determination of an appropriate heat treatment of animal waste using the ELISA technique: results of a validation study. *Meat Science 54* (1):1-7.

von Holst C, Unglaub W, and Anklam E. 2001. Post process product control of rendering plant sterilization conditions by ELISA. *Journal of Association of Official Analytical Chemists International 84* (6):1793-1799.
Vuorio E, and de Crombrugghe B. 1998. The family of collagen genes. *Annual Reviews of Biochemistry 59*:837-872.
Waite ER, Child AM, Craig OE, Collins MJ, Gelsthorpe K, and Brown TA. 1997. A preliminary investigation of DNA stability in bone during artificial diagenesis. *Bulletin De La Societe Geologique De France 168* (5):547-554.
Walker, M. 2005. *Quaternary dating methods.* Chichester: Wiley.
Wallace JM, Rajachar RM, Chen XD, Shi S, Allen MR, Bloomfield SA, Les CM, Robey PG, Young MF, and Kohn DH. 2006. The mechanical phenotype of biglycan-deficient mice is bone- and gender-specific. *Bone 39* (1):106-116.
Wang RF, Myers MJ, Campbell W, Cao WW, Paine D, and Cerniglia CE. 2000. A rapid method for PCR detection of bovine materials in animal feedstuffs. *Molecular and Cellular Probes 14* (1):1-5.
Wapnish P, and Hesse B. 1988. Urbanization and the organization of animal production at Tell Jemmeh in the middle Bronze Age Levant. *Journal of Near Eastern Studies 47* (2):81-94.
Weber IT, Harrison RW, and Iozzo RV. 1996. Model structure of decorin and implications for collagen fibrillogenesis. *Journal of Biological Chemistry 271*(50):31767-31770.
Weiner S, and Bar-Yosef O. 1990. States of preservation of bones from prehistoric sites in the Near East: A survey. *Journal of Archaeological Science 17*:187-196.
Weiner S, Lowenstam HA, and L. H. 1976. Characterisation of 80-million-year-old mollusk shell proteins. *Proceedings of the National Academy of Sciences USA 73*:2541-2545.
Weiner S, and Price PA. 1986. Disaggregation of bone into crystals. *Calcified Tissue International 39*:365-375.
Weiner S, and Traub W. 1986. Organization of hydroxyapatite crystals within collagen fibrils. *Federation of European Biochemical Societies Letters 206*:262-266.
Weiner S, and Traub W. 1992. Bone structure; from ångstroms to microns. *Journal of the Federation of American Societies for Experimental Biology 6*:879-885.
Weiner S, Traub W, Elster H, and DeNiro MJ. 1989. The molecular structure of bone and its relation to diagenesis. *Applied Geochemistry 4*:231-232.
White CL, Young P, Phillips N, and Rodehutscord M. 2000. The effect of dietary protein source and protected methionine (Lactet) on wool growth and microbial protein synthesis in Merino wethers. *Australian Journal of Agricultural Research 51* (2):173-183.
Wiley WC, and McLaren I. 1955. Time-of-flight mass spectrometer with improved resolution. *Review of Scientific Instruments 26*:12.
Willerslev E, and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences 272* (1558):3-16.
Williams RAD, and Elliot JC. 1989. *Basic and applied dental biochemistry.* Edinburgh: Churchill Livingstone.
Wilson KA, and Walker J. 2005. *Principles and techniques of biochemistry and molecular biology.* Cambridge: Cambridge University Press.
Wood WB, Wilson JH, Benbow RM, and Hood LE. 1981. *Biochemistry - A Problems Approach.* Californica: Benjamin Cummings.
Woodward SR, Weyand NJ, and Bunnell M. 1994. DNA sequence from Cretaceous period bone fragments. *Science 266* (5188):1229-1232.
Wright LE, and Schwarcz HP. 1996. Infrared and isotopic evidence for diagenesis of bone apatite at Dos Pilas, Guetemala: Palaeodietary implications. *Journal of Archaeological Science 23*:933-944.
Wu WW, Wang G, Baek SJ, and Shen RF. 2006. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel-or LC-MALDI TOF/TOF. *Journal of Proteome Research 5* (3):651-658.
Wyckoff RW, and Davidson FD. 1973. Gelatinized Pleistocene collagen. *Comparative Biochemistry and Physiology B 46* (1):143-146.
Wyckoff RWG. 1972. *The biochemistry of animal fossils.* Bristol: Scientechnica.

Yamashita M, and Fenn JB. 1984. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry 88* (20):4451-4459.
Yang DY, Woiderski JR, and Driver JC. 2005. DNA analysis of archaeological rabbit remains from the American Southwest. *Journal of Archaeological Science 32* (4):567-578.
Yata M, Yoshida C, Fujisawa S, Mizuta S, and Yoshinaka R. 2001. Identification and characterization of molecular species of collagen in fish skin. *Journal of Food Science 66* (2):247-251.
Young JZ. 1962. *The Life of Vertebrates*. New York: Oxford University Press.
Zang X, van Heemst JDH, Dria KJ, and Hatcher PG. 2000. Encapsulation of protein in humic acid from a histosol as an explanation for the occurrence of organic nitrogen in soil and sediment. *Organic Geochemistry 31*(7-8):679-695.
Zeder MA. 2006. Reconciling rates of long bone fusion and tooth eruption and wear in sheep (*Ovis*) and goat (*Capra*). In: Ruscillo D, (Ed.) *Ageing and sexing animals from archaeological sites.* Oxford: Oxbow Press, p 87–118.
Zenobi R, and Knochenmuss R. 1998. Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews 17* (5):337-366.
Zohary, D. Tchernov, E. and Horwitz, LK. 1998. The role of unconscious selection in the domestication of sheep and goats. *Journal of Zoology 245*: 129-135.
Zhang Y, Liu W, Li G, Shi B, Miao Y, and Wu X. 2007. Isolation and partial characterization of pepsin-soluble collagen from the skin of grass carp (Ctenopharyngodon idella). *Food Chemistry 103* (3):906-912.
Zhigilei LV, Kodali PBS, and Garrison BJ. 1997. Molecular dynamics model for laser ablation and desorption of organic solids. *Journal of Physical Chemistry B 101* (11):2028-2037.
Zhu W, Robey PG, and Boskey AL. 2007. The regulatory role of matrix proteins in mineralisation of bone. In: Feldman D, Nelson D, and Rosen CJ, (Eds.) *Osteoporosis. Third Edition ed.* New York: Elsevier, p 191-240.

Appendices 1-7

The appendices are attached in electronic form on DVD as various file types.

Appendix 1: Chapter 1 Supplementary Material

1.1 Aligned collagen (I) sequences available from UniProt database

Appendix 2: Chapter 3 Supplementary Material

2.1 Search results of cattle and chicken OC fragment peak masses against theoretical OC fragment masses

2.2 Search results of cattle and chicken MBM OC fraction peak masses against theoretical OC and collagen fragment masses

2.3 Search results of spiked chicken MBM OC fraction peak masses against theoretical OC and collagen fragment masses

2.4 Search results of chicken MBM sediment OC fraction peak masses against theoretical OC and collagen fragment masses

2.5 Search results of chicken MBM OC fraction peak masses following ultrafiltration against theoretical OC and collagen fragment masses

2.6 Search results of chicken MBM OC fraction LC-MS peak masses (following ultrafiltration) against theoretical OC and collagen fragment masses

2.7 LC-MALDI data files for OC fractions of chicken MBM (raw) rendered at 133, 137, 141 and 145°C

2.8 LC-MALDI search results for OC fractions of chicken MBM (raw) rendered at 133, 137, 141 and 145°C using standard scoring UniProt decoy database searches (exported as CSV files)

Appendix 3: Chapter 4 Supplementary Material

3.1 Method development of the standard operating procedures for the LC-MS and SPE-MS methods used throughout this thesis.

3.2 LC-MALDI data for cattle, pig, sheep and chicken bone 'collagen' and cattle and chicken MBM (unsedimented and sedimented) 'collagen' digests

3.3 LC-MALDI search results for cattle, pig, sheep and chicken bone 'collagen' and cattle and chicken MBM (unsedimented and sedimented) 'collagen' digests using both standard scoring and MudPIT scoring decoy UniProt database searching

3.4 Descriptions of proteins identified in Chapter 4

3.5 Amino acid D/L values of 'total', 'soluble' and 'insoluble' fractions of MBM (raw and sedimented) rendered at 133, 137, 141 and 145°C

3.6 Amino acid concentration values of 'total', 'soluble' and 'insoluble' fractions of MBM (raw and sedimented) rendered at 133, 137, 141 and 145°C

Appendix 4: Chapter 5 'Supplementary Online Material' Figures 5-7

Appendix 5: Chapter 9 Supplementary Material

5.1 Sequences used in the 'Collagens' database