# Comparing the Use of Magnetic Beads with Ultrafiltration for Ancient Dental Calculus Proteomics

**Document Version**
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

# Comparing the use of magnetic beads with ultrafiltration for ancient dental calculus proteomics

*Karren S Palmer[1], Cheryl A.Makarewicz[2], Alexey A. Tishkin[3], Svetlana S. Tur[3], Amartuvshin Chunag[4], Erdenebaatar Diimajav[5], Bayarsaikhan Jamsranjav[6], Michael Buckley[1*]*

[1]School of Natural Sciences, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, U.K.
[2]Kiel University, Institute for Prehistoric and Protohistoric Archaeology, Johanna-Mestorf Strasse 2-6, D-24118, Kiel, Germany
[3]Altai State University, Department of Archeology, Ethnography, and Museology, Lenin Prospect 61, 656049, Barnaul, Russia.
[4]Mongolian Institute of Archaeology, Ulaanbaatar, Mongolia
[5]Department of Archaeology and History, Ulaanbataar State University, 13343, Luvsantseveen Street, 5th khoroo, 15th khorgoolol, Bayanzurk District, Ulaanbaatar, Mongolia
[6]National Museum of Mongolia, Ulaanbaatar Mongolia, Juulchin Street-1, Ulaanbaatar, Mongolia

**Corresponding author: M.Buckley@manchester.ac.uk**

## Abstract

Over the past two decades, proteomic analysis has greatly developed in application to the field of biomolecular archaeology, coinciding with advancements in LC-MS/MS instrumentation sensitivity and improvements in sample preparation methods. Recently, human dental calculus has received much attention for its well-preserved proteomes locked in mineralized dental plaque which stores information on human diets and the oral microbiome otherwise invisible to other biomolecular approaches. Maximizing proteome recovery in ancient dental calculus, available only in minute quantities and irreplaceable after destructive analysis, is of paramount importance. Here, we compare the more traditional ultrafiltration-based and acetone precipitation approaches with the newer paramagnetic bead approach in order to test the influence of demineralization acid on recovered proteome complexity obtained from specimens as well as the sequence coverages matched for significant proteins. We found that a protocol utilising EDTA combined with paramagnetic beads increased proteome complexity, in some cases doubling the number of unique peptides and number of proteins matched, compared to protocols involving the use of HCl and either acetone precipitation or ultrafiltration. Although the increase in the number of proteins was almost exclusively of bacterial origin, a development that has implications for the study of diseases within these ancient populations, an increase in the peptide number for the dairy proteins β-lactoglobulin and casein was also observed reflecting an increase in sequence coverage for these dietary proteins of interest. We also consider structural explanations for the discrepancies observed between these two key dietary proteins preserved in archaeological dental calculus.

## Keywords:

β-lactoglobulin, casein, ultrafiltration, SP3, acetone precipitation, archaeological dairy protein, dental calculus proteome.

**Significance** Proteomic analysis of ancient tissues and organic residues have gained rapid prominence in archaeological research as a means to trace dietary intake, but methods used for extracting valuable proteomes from these finite and often minute resources are still under development. In an effort to amplify recovery of ancient proteomes from human dental calculus, we compare traditional ultrafiltration and acetone precipitation extraction method with a newer paramagnetic bead approach SP3 with no nomimal molecular weight cut-off for low abundance and poorly preserved proteins. We also compare different demineralization agents (ethylenediaminetetraacetic acid (EDTA) and hydrochloric acid (HCl) in combination with the chaotrope guanidine hydrochloride (GuHCl), and observed the demineralising agent EDTA in combination with GuHCl and SP3 compared to HCl or the more traditional proteome extraction methodologies exhibited a greater peptide recovery number and proteome complexity. We identified a marked increase in the amount of ancient proteins recovered using a combination of EDTA and SP3 paramagnetic beads, and explore how protein structure, burial environment, and oral enzyme activity may influence ancient proteome preservation.

## 1. Introduction

Dental calculus, the mineralized form of dental plaque, has been recognized as a high-resolution record of ancient dietary and environmental information since Armitage[1] identified preserved phytoliths entrapped within it (or cementum growth layers) in the teeth of cattle, sheep and horses using scanning electron microscopy. More recently, dramatic improvements in contamination control involving the use of designated cleanroom environments and rapid advancements within the field of paleogenomics, in particular high-throughput Next Generation Sequencing facilitating the isolation and analysis of preserved host mitochondrial DNA[2] and other biomolecules[3,4] has established ancient dental calculus as a suitable long-term reservoir of ancient biomolecules (e.g., DNA, protein and lipids) and micro-remains (e.g. plant microfossils and environmental debris). Over 600 different oral bacterial, viral and fungal taxa have so far been isolated from modern and ancient dental calculus specimens including disease-causing microorganisms associated with the respiratory tract[5-8]. Recently, proteins, lipids and/or mineral particles stored in dental calculus have provided insights into ancient human diets and migration[9,-11], the paleomicrobiome influenced by health status and disease[7], living conditions including inhaled environmental pollutants[12] and occupational activities[13]. However, while proteomic analyses hold much promise for revealing ancient human life histories at both the individual and population level, the combination of minute quantities of dental calculus visibly preserved on teeth, finite supply of calculus provided by a single sampled ancient individual, and diagenetic alteration of proteins in the burial environment, demands maximum recovery of ancient proteins while retaining *in vivo* taxonomic diversity.

### 1.1 Dental Calculus Formation and Composition

Dental calculus is composed primarily of inorganic calcium and phosphorus, along with numerous minor constituents (including carbonate, sodium, magnesium, silicon, iron, fluoride, brushite, whitlockite, octacalcium phosphate and hydroxyapatite[14] and is classified as either supra- or subgingival, relating to its location on the crown or root of the tooth surface respectively[14-17]. It is the mineralized form of dental plaque, a multi-layered oral biofilm that captures the microbial community in a matrix of salivary and bacterials polymers laid down on the tooth crown surface. Dental plaque is formed through interactions between early oral bacterial colonisers including *Streptococcus gordonii and Actinomyces naeslundii,* as well as proteins, glycoproteins, lipids and glycolipids derived from saliva and additional extracellular molecules from bacteria, food residues and other biomolecules of host origin e.g., lysozyme, statherin and

immunoglobulins[18]. Without intervention, dental calculus begins to form within 14 days of plaque formation on the surface of the tooth through the deposition of calcium phosphates, underneath a non-mineralized layer of dental plaque[16]. It is formed largely on the alkali lingual surfaces of the anterior teeth located adjacent to the salivary ducts and is continuously generated throughout the lifetime of an individual, thus providing an accumulative record of dietary intake and oral health[16].

The organic composition of dental plaque and, ultimately, calculus varies considerably between individuals and populations based on diet, age, systematic health, oral microbiome, genetics and oral self-care regime[14,17,19-20]. The relative contribution of each of these factors on rate of dental plaque and calculus formation is still unclear. Similarly, the effects of oral cavity pH and salivary flow on formation rate is unknown, although relatively high alkaline conditions caused by a high-protein diet appear to deter calculus formation while the ingestion of high carbohydrate foods promotes it[17,21]. Differences in the protein intake may contribute to the wide variation in the amount of dental calculus preserved on archaeological tooth specimens on both the individual and population level. Contemporary hunter-gatherers consuming traditional diets rarely yield dental calculus in sharp contrast to the ubiquity of dental calculus in individuals consuming carbohydrate-rich western diets[22].

## 1.2 Biomolecules in Ancient Dental Calculus

Both ancient and modern biomolecules variably preserve in archaeological and clandestine burial environments depending on local sediment pH, temperature, water-content, presence or absence of oxygen, ionic radiation, burial textiles and microorganisms, including the soil[23-31]. Short fragments of DNA and/or proteins from ancient specimens theoretically can survive up to at least one million years under favorable conditions, e.g. within permafrost[32-34]. Yet, the degree of molecule preservation can differ across similar burial environments so that predicting the survival of these ancient biomolecules is difficult[23,26,35-37]. Further compounding this challenge, the structural composition of the matrix surrounding the entrapped organic molecule also influences preservation. DNA, proteins and other biomolecules encased in a mineralized non-porous material (such as bone and teeth) are more likely to survive in greater abundance than those encased in a porous mineralized material (such as dental calculus) or a soft tissue (such as the muscle) as a result of limited chemical and microbial actions from the surrounding burial environment[38].

Over the last decade, isolating and identifying the dietary origins of biomolecules entrapped within dental calculus has become of increasing importance as a tool for studying past societies in archaeology[4,39]. Shotgun sequencing of ancient DNA recovered from hominin dental calculus revealed Neanderthal dietary intake was heavily shaped by environmentally-defined resource availability with meat-based diets characteristic for vegetation-poor steppe environments but plant-based forest diets including mushrooms, pine nuts, and moss for Neanderthal groups inhabiting more temperate climates[39]. Proteomic analysis of dental calculus recovered from ancient pastoralist identified Bovidae and Equidae dairy proteins β-lactoglobulin and casein, confirming dairy consumption by established pastoralist communities inhabiting the Mongolian steppe[10], with no evidence of dietary supplementation from foraging local flora within these communities. The absence of the isolation of local flora from these samples is not that surprising as these proteins would be prone to additional decay upon ingestion by salivary amylase and lingual lipase compared to their dairy counterparts and less likely to be encased as an intact protein and therefore theoretically more susceptible to further decay during burial. Recent research has also explored disease through analysis of the ancient proteome, for example, isolating fimbrial proteins and gingipains from *Porphyromonas gingivalis*

related to periodontal disease, and also immune response proteins specific to neutrophils that are the first line response to bacterial infection[7].

## 1.3 Methodological Approaches to Studying Ancient Dental Calculus

At present, there is no consensus technique for sample preparation in terms of the acids, buffers or protocols implemented for the isolation of ancient proteins from dental calculus for proteomic research. Potential contamination from the burial and laboratory environment adds to the challenges of working with this rare and finite archaeological resource, especially as instrument sensitivity is pushed to their limits, constantly redefining data handling approaches to account for poor protein preservation and/ or laboratory and burial contaminants. The negative implications of which are two fold in terms of the potential data lost from comparative analysis across archaeological studies performed by different research groups regarding preservation status and the application of a 'gold standard' for interpreting data and contaminant removal. Previously published dental calculus sample preparation methodologies for proteomic analysis can be broadly classified as either 'filter-assisted' e.g., Filter Aided Sample Preparation (FASP) or 'gel-assisted' e.g., Gel Aided Sample Preparation (GASP), depending on how the proteins are digested and purified prior to analysis by mass spectrometry. In the present study, we consider Single Pot Solid Phase Enhanced Sample Preparation (SP3) as a viable alternative for the extraction of these ancient proteomes to the filter-assisted methodologies[40].

SP3 was first pioneered by Hughes *et al.*[41] (for eukaryotic cells/ drosophila embryos) and refined by Sielaff *et al.*[42] (for HELA cells), Modridge *et al.*[43] (for HEK-293 cells) and Cleland[44] (for bone) for bottom up sample preparation. It is a paramagnetic bead approach with carboxylate-modified surfaces known to work with a small amount of proteins and a variety of buffers[44]. Unlike filtration-based approaches (FASP and ultrafiltration) that typically yield a recovery rate of ~50% relative to the initial material after digestion, as previously observed with modern samples with FASP[42], SP3 is not limited by the loss of small peptides and quantitative reproducibility below 10 $\mu$g (for eukaryotic cells) imposed by a nominal molecular weight cut-off or overloading of the filtering device[41-45]. The advantages offered by SP3 for ancient dental calculus, which likely contains small protein fragments caused by degradation as well as partial digestion by amylase within saliva, include no nominal molecular weight cut-off, a simple protocol that can be reproduced efficiently with multiple acid/ buffer variants and the absence of potential polymer contaminants from the filter devices used to isolate the peptides which could mask low abundance protein signal. Furthermore, unlike GASP, SP3 is a less labor-intensive methodology in terms of isolating the proteins/peptides directly via amine binding as opposed to retrieving the proteins indirectly from a gel matrix, a process which requires extensive cleanup prior to digestion and analysis.

To determine the feasibility of SP3 as a valid alternative to filter/ultrafiltration based sample preparation approaches for the recovery of ancient proteomes, we performed a comparison of a refined version of the SP3 methodology implemented by Cleland[44] and an ultrafiltration approach[40] on several historical dental calculus samples dating from the late 4th to the late 1st millennium BC. Secondary aims were also to compare instrument platforms, with our analyses being conducted on either an Orbitrap Elite (Thermo Fisher Scientific, Waltham, MA) and/or QE HF (Thermo Fisher Scientific, Waltham, MA) mass spectrometer.

## 2. Experimental Section

### 2.1 Materials

With appropriate permissions, ancient dental calculus was recovered from the tooth remains of human individuals interred in monumental mortuary complexes, located in Russia and Mongolia (housed in Institute for Ethnography and Archaeology at Altai State University, and the Mongolian Institute of Archaeology, the Department of Archaeology at Ulaanbataar State University, and the Mongolian National Museum respectively), dating from the late fourth to late 1st millennia BC. Initial method development was carried out on three specimens (equating to 32 sample preparations including experimental blanks), but overall a total of 153 human individuals were sampled including 44 ancient individuals from Mongolia and 109 from Russia. All 153 specimens were processed and analysed for the presence of dietary and/or oral disease and immune response associated proteins. Initial sample preparation was performed in a dedicated clean room facility at the University of Manchester used exclusively to study ancient biomolecules except for the three initial method development samples (FIR-14, FIR-15 and STP-29) which were prepared and extracted in a standard laboratory environment.

All chemicals were purchased from the following manufacturers: Fisher Scientific (hydrochloric acid (HCl), acetonitrile (ACN) and ethylenediaminetetraacetic acid (EDTA, pH 8.0, Corning 46-034-C1)), Fluka (ammonium bicarbonate (ABC)), Sigma-Aldrich (dithiothreitol (DTT), iodoacetamide (IAM), trifluoroacetic acid (TFA), guanidine hydrochloride (GuHCl), Vivaspin (3 kDa molecular weight cut off (MWCO) ultrafiltration units), Promega (sequencing grade trypsin, Agilent Technologies (OMIX C18 reversed phase ZipTips) and GE Healthcare (Sera Mag SpeedBeads).

### 2.2 Methods

#### 2.2.1 Demineralisation acid and method comparison

Initial methodological comparison focused on the demineralisation acid used to isolate any surviving ancient protein within the dental calculus. Choosing the most appropriate demineralisation acid required in order to remove calcium salts is fundamental to the efficiency of the extraction in terms of proteome complexity and chemically induced post translational modifications. Previous studies for proteomic analysis of dental calculus[46] have utilized either EDTA or acetic acid (AA) as the demineralization agent. Here, we compare EDTA with a more aggressive demineralization agent - HCl, which is known to extract a more complex proteome from bone in terms of proteome variety and protein numbers than EDTA[40], but at a cost of increased chemically-induced deamidation[40]. Three dental calculus specimens of Russian origin were chosen: FIR014, FIR015 and STP029 from the 153 archaeological specimens identified previously as sufficient in terms of material for analysis i.e., >2 mg (Figure 1A). These samples were chosen based on the amount of material for analysis (as all three samples exceeded 30 mg in total) and their age (probability of ancient proteins surviving). These specimens were pre-washed with 10 µl of 1% (v/v) bleach and rinsed several times with milliQ water prior to analysis. The pH of the dental calculus surface was also checked prior to processing to ensure all residual bleach was removed by placing pH indicator strips on the damp surface of the dental calculus specimens.

Dental calculus was first pre-washed using either 0.5 ml of 0.6 M HCl or 1 ml of 0.5M EDTA was added to ~15-30 mg of each of the dental calculus samples in order to remove surface contamination such as soil particles. The specimens were then treated for 66 h at 4 °C (HCl) and 37 °C (EDTA), reflecting time

and temperature intervals utilized by Procopio and Buckley[40] and Cleland[44] for comparison. On completion of the acid treatment, the samples were centrifuged at 8,000 rpm for 2 min to pellet the insoluble debris prior to the removal of the supernatant. The supernatants were aliquoted and the insoluble pellets were placed in 1ml of a buffer containing 6 M GuHCl and 100 mM Tris at pH 7.4 and incubated at 4 °C for 18 h. The aliquoted supernatants (the protein containing acid fractions) were stored temporarily at -20 °C for 24 h to enable the soluble and insoluble proteins to be extracted in parallel. On completion of the GuHCl treatment, the insoluble supernatant was removed and any remaining dental debris left to air-dry in the fume-hood for 48 h prior to long-term storage at -20 °C. The soluble (HCl or EDTA) and insoluble (GuHCl) supernatants were then processed as follows: all the fractions for each sample were processed separately and the buffers exchanged from EDTA, HCl and GuHCl to 50 mM ABC for the fractions to be processed by either 3kDa MWCO ultrafiltration columns or acetone (acetone precipitation). The fractions for extraction by SP3 were left in either EDTA or GuHCl following demineralisation and extraction of the soluble and insoluble protein fractions respectively. Acetone precipitation was used as an alternative to SP3 for the HCl soluble protein fractions only, as covalent coupling between the proteins and the surface of the particles can be negatively impaired by the pH of the buffer used during the extraction procedure
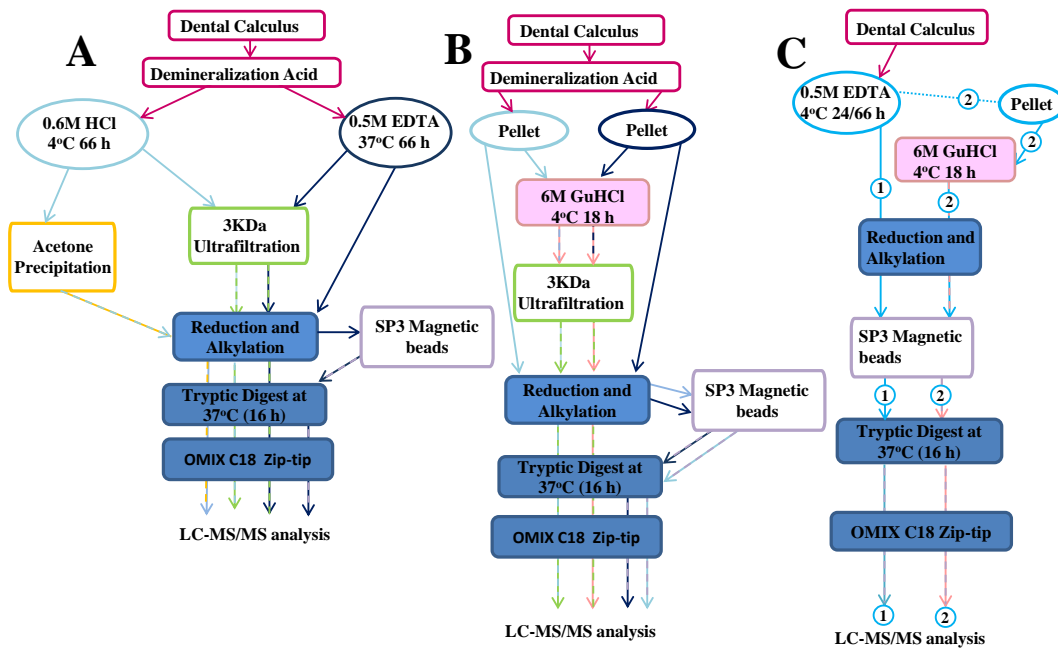


**Figure 1 -** *Flow charts of the overall experimental designs implemented in this comparison study:*
*Initial tests of different acids for demineralisation, ultrafiltration versus SP3 magnetic beads versus acetone precipitation. (a).Dental calculus samples, FIR-14, FIR-15 and STP29 (batch 1) ground, halved and demineralised in either 0.5 M EDTA for 66 h at 37 °C or 0.6 M HCl for 66 h at 4 °C (soluble protein fractions). Acid fractions removed and stored temporarily at -20 °C. (b). Remaining 'dental debris' was treated for 18h at 4°C with GuHCl (to isolate the insoluble proteins). After GuHCl treatment, the samples isolated using the SP3 methodology were reduced and alkylated prior to protein isolation on the magnetic beads. In contrast to the samples isolated by SP3, the samples isolated by ultrafiltration and acetone precipitation, were isolated and then reduced and alkylated. Soluble and insoluble protein fractions underwent trypsin digestion at 37 °C prior to OMIX 18 zip-tipping and resuspension in 80 % (v/v) ACN with 0.1 % TFA. (c). SP3 magnetic bead methodology for 0.5 M EDTA only. Dental calculus samples were prepared and processed as stated for the samples in figures 1(a) and 1(b) with the following modification. The samples were treated with EDTA for either 24 or 66 h at 4 °C. Stages of sample preparation/ processing depicted by solid and dashed arrows. Colour coding key: sky and dark blue represent the samples demineralized in HCl or EDTA respectively, pink GuHCl treatment, yellow acetone precipitation, green ultrafiltration and purple SP3.*

The retentate for the proteins isolated by the 3 kDa columns from 400 µl of protein containing acid fractions were collected using a 100 µl volume of 50 mM ABC, and the proteins eluted, reduced and alkylated. For reduction, 4.2 µl of 100 mM DTT was added to each sample and incubated at RT for 45 min in the dark. After the samples had been reduced, alkylation of these samples was achieved by adding 8.4 µl of 100 mM IAM and incubating the samples at RT for 45 min in the dark. The alkylation step was then quenched by the addition of another 4.2 µl of 100 mM DTT and the samples digested using 0.4 µg of Promega sequencing-grade modified trypsin at 37 $^{o}$C for 18 h. A 100 µl volume of each acid retentate, for acetone precipitation and SP3 separation, were temporarily stored at 4 $^{o}$C. The 100 µl of HCl for each sample were precipitated with Acetone overnight at -20 $^{o}$C at an acetone volume of 4 to 1, and resuspended the following day in 50 mM ABC prior to reduction, alkylation, quenching and tryptic digestion. The 100 µl volumes of EDTA for each SP3 extracted sample were reduced and alkylated alongside the filter-aided and acetone fractions and then processed according to a modified version of the Cleland protocol. A 1:1 mixture of the hydrophobic and hydrophilic Sera-Mag Speedbeads (GE Healthcare), were washed three times in DNAse and RNase free water (Life Sciences Technologies). 10 µl of the prepared beads (equivalent to 1/10 of the volume of sample protein solution) were added to the reduced and alkylated dental calculus protein solutions and a volume of ethanol equivalent to the total volume (sample + bead + alkylation/reduction volumes), and incubated at room temperature for 10 min with rotation. On completion of the incubation phase, the beads were placed on a magnetic rack to collect the beads, the supernatant removed and then stored at -20 $^{o}$C. The magnetic beads were then washed three times in 80% (v/v) ethanol (with rotation and collection on a magnetic rack) and the resulting supernatant stored at -20 $^{o}$C. The beads for each sample were resuspended in 98 µl of 50 mM ABC and digested in 0.4 µg of trypsin (2 µl, sequencing-grade modified; Promega, UK) for 18 h at 37 $^{o}$C alongside the filter-based and acetone precipitated fractions (Figure 1A).

### 2.2.2 Further analysis considerations

Demineralisation in HCl and protein isolation by acetone precipitation were excluded from further experimental consideration as EDTA and the ultrafiltration/SP3 methodologies appeared better suited for the overall objects of the study (Figure 1B and 2). Based on peptide numbers and protein identification (Figure 2B and C), all remaining samples (Table 1) were processed with a demineralization period of 24 or 66 hr at 4 $^{o}$C and isolated using the SP3 methodology for indirect comparison. Demineralisation time was reduced to 24 h as the specimens demineralized for 24 h at a temperature of 4 $^{o}$C should generate a more complex proteome caused by a reduction in chemically induced post-translational modification compared to a 66 h incubation period as observed in previous studies with modern bone[40].

*Table 1.* *Methodology used per dental calculus exhumed from skeletal remains in this study. These range from key (e)Neolithic, Bronze and Iron age mortuary monuments located in Mongolia and Russia dating from the late 4$^{th}$ to the late 1$^{st}$ millennium BC.*

| Sample ID | EDTA or HCl? | Demineralization interval (h)? | Methodology |
|---|---|---|---|
| FIR-15 | Both | 66 | SP3 Beads, Filters & Acetone |
| FIR-14 | Both | 66 | SP3 Beads, Filters & Acetone |

| | | | |
|---|---|---|---|
| STP-29 | Both | 66 | SP3 Beads, Filters & Acetone |
| TUZ-06, TUZ-07, TUZ-12, SAL-10, SAL-12 | EDTA | 66 | SP3 Beads & Filters |
| CHO-10, SAL22, SAL-24, TUZ-20, CHO-01, FIR-4, TUZ-03, TEL-4, TEL-7, TEL-14, TEL-17, TEL-22, MYS-2, MYS-5, KAL-1, MYS-04, TEL-27, BOT-1, TUZ-13, TUZ-15, STP-26, STP-27, STP-28, STP-32, #6365,#6255, #6888, #6366, #6851, #6406, #6308, #6821, #6417, #6855, #6837, #6847, #6848, #6572,#6873, #6507, #6510, #6516 | EDTA | 24 | SP3 Beads only |
| TEL-13, TEL-19, TEL-26, TEL-12, MYS-04,BOT-2, BOT-3, INS-1, INS-2, CHB-1, CHB-2, CHB-3, BER-1, AIR-1, TYT-1, SAL-01, SAL-02, SAL-04, SAL-05,SAL-06, SAL-07, SAL-08, SAL-09, SAL-11, SAL-13, PMR-04, BER-02, BOT-4, CHB-4, STA-8, STP-22, STP-23, STP-24, STP-35, STP-04, FIR-13, FIR-14a, FIR-17, FIR-19, FIR-16FIR-20, FIR-21, FIR-22, FIR-24, FIR-25, FIR-26, FIR-28, FIR-29, FIR-32a, FIR-33, FIR-34, FIR-35, FIR-36, FIR-37, FIR-39, FIR-41, FIR-43, FIR-44, FIR-45, TUZ-10, FIR-30,STP-1, STP-3, STP-7, STP-12, STP-13, STP-16, STP-17, STP-20, #6178, #6370, #6371, #6384, #6440a, #6440b, #6442, #6646, #6650-P, #6660, #6662, #6673, #6678,#6685, #6382, #6832, #6835, #6856, #6857, #6841, #6308, #6310, #6313, #6828, #6838, #NO ASIL | EDTA | 66 | SP3 Beads only |

### 2.2.3 LC-MS/MS Analysis

All LC-MS/MS analysis was performed by the Biological Mass Spectrometry Facility at the University of Manchester. 1 % (v/v) TFA was added to the peptide fractions generated for each sample by tryptic digest for 18 h at 37 °C, to achieve a final concentration of 0.1 % TFA for inhibition of further tryptic digestion. This was applied for all stages of the parameters/ methods comparison described. These peptide containing soluble and insoluble fractions were processed independently and prepped for LC-MS/MS analysis using an OMIX C18 reversed phase Zip-tip to desalt, purify and concentrate the samples as described in Procopio and Buckley[40] with two exceptions. Firstly, the peptides were eluted in 80 % (v/v) ACN + 0.1 % (v/v) TFA instead of 50 % (v/v) ACN + 0.1 % (v/v) TFA and secondly they were dried for 4 h at 37 °C using a centrifugal evaporator prior to resuspension in 20 µl of 0.1 % TFA/5 % ACN for subsequent LC-MS/MS analysis. The resulting peptides were analysed by LC-MS/MS using an UltiMate® 3000 Rapid Separation LC (RSLC, Dionex Corporation, Sunnyvale, CA) coupled to an Orbitrap Elite (Thermo Fisher Scientific, Waltham, MA) mass spectrometer. Extraction blanks were processed and analysed alongside the experimental samples and two samples of Pierce™ 6 Protein Digest, equimolar, LC-MS grade (88342) were ran prior to and after each sample batch to prevent cross-contamination between sample batches. Pierce™ 6 Protein Digest, equimolar, LC-MS grade (88342) is used as an alternative to the standard injection blanks as it is more efficient at removing lingering peptides from the LC-MS/MS columns and enables the user to check the status of the instrument between batches.

Mass spectrometry data, for the Orbitrap Elite, was acquired in a data directed manner for 60 minutes in positive mode. Peptides were selected for fragmentation automatically by data-dependent analysis on a basis of the top 12 peptides with m/z between 300 to 1750Th and a charge state of 2, 3 or 4 with a dynamic exclusion set at 15 sec. The MS Resolution was set at 120,000 with an AGC target of 3e6 and a maximum fill time set at 20 ms. The MS2 Resolution was set to 30,000, with an AGC target of 2e5, a maximum fill time of 45 ms, isolation window of 1.3Th and collision energy of 28. Mobile phase A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in acetonitrile and the column used

was a 250 mm x 75 μm i.d. 1.7 mM nanoE MZ PST CSH130 C18, analytical column (Waters). A 3ul aliquot of the sample was transferred to a 5ul loop and loaded on to the column at a flow of 300nl/min for 13 minutes at 5% B. The loop was then taken out of line and the flow was reduced from 300nl/min to 200 nl/min in 0.5 min.  Peptides were separated using a gradient that went from 5% to 18% B in 34.5 minutes, then from 18% to 27% B in 8 minutes and finally from 27% B to 60% B in 1 min. The column is washed at 60% B for 3 min before re-equilibration to 5% B in 1 minute. At 55 minutes the flow is increased to 300 nl/min until the end of the run at 60 min. Samples identified as containing at least one unique peptide for either β-lactoglobulin or a member of the casein protein family were re-run on an UltiMate® 3000 Rapid Separation LC (RSLC, Dionex Corporation, Sunnyvale, CA) coupled to a QE HF (Thermo Fisher Scientific, Waltham, MA) mass spectrometer for independent validation by a second instrument were possible.

Mass spectrometry data, for the QE HF, was acquired in a data directed manner for 60 minutes in positive mode. Peptides were selected for fragmentation automatically by data dependent analysis on a basis of the top 12 peptides with *m/z* between 300 to 1750Th and a charge state of 2, 3 or 4 with a dynamic exclusion set at 15 sec. The MS Resolution was set at 120,000 with an AGC target of 3e6 and a maximum fill time set at 20 ms. The MS2 Resolution was set to 30,000, with an AGC target of 2e5, a maximum fill time of 45 ms, isolation window of 1.3 Th and collision energy of 28. Mobile phase A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in acetonitrile and the column used was a 250 mm x 75 μm internal diameter 1.7 250 mm x 75 μm i.d. 1.7 μM nanoE MZ PST CSH130 C18, analytical column (Waters). A 3 μl aliquot of the sample was transferred to a 5ul loop and loaded on to the column at a flow of 300 nl/min for 13 min at 5% B. The loop was then taken out of line and the flow was reduced from 300 nl/min to 200 nl/min in 0.5 min.  Peptides were separated using a gradient that went from 5% to 18% B in 34.5 min, then from 18% to 27% B in 8 minutes and finally from 27% B to 60% B in 1 minute. The column is washed at 60% B for 3 min before re-equilibration to 5% B in 1 min. At 55 minutes the flow was increased to 300 nl/min until the end of the run at 60 min.

### 2.2.4 Spectrum Analysis and Quality Control

Raw spectra were converted to Mascot generic files (created assuming 0 #[13]Cs with peak list generation using ExtractMSN) and initially searched against: Swiss-Prot (556,568 sequences; 199,530,821 residues) and HOMD (4,026,163 sequences; 1,246,432,637 residues) using the Mascot Search engine (version 2.4.1; Matrix science, London, UK) against a decoy database. The searches engines TrEMBL (177,754, 527) and cRAP (version 2012.01.01) were excluded from analysis, on the basis that the peptides from the proteins of interest (diet and disease) and possible laboratory contaminants were included in the databases Swiss-Prot and HOMD respectively. Each search included the fixed carbamidomethyl modification of cysteine (+57.02 Da), the variable modifications of deamidation (+0.98 Da) of asparagine/glutamine, and oxidation (+15.99 Da) of proline and methionine residues to account for chemical induced and natural post translational modifications associated with sample processing, diagenesis and hydroxylation (e.g., Procopio and Buckley[40] and Jeong et al.,[47]). Enzyme specificity was set to trypsin with up to 2 missed cleavages allowed; mass tolerance were set to 10 ppm for the precursor ions and 0.5 Da for the fragment ions, with all spectra considered as having 2+ or 3+ precursors for the data generated by the LC-Orbitrap Elite system. Archaeological dental calculus specimens, containing 1 or more, unique peptides identified as potential matches to one of the dietary proteins: β-lactoglobulin and/or a member of the casein protein family for the tryptic search were re-run on the a QE HF as a diagnostic marker for preservation and for independent validation by a second instrument. All parameters were the

same with the following exception; the fragment ion specificity was reduced to 0.02 Da. Scaffold software version 4.11.0 (Proteome Software Inc., Portland, OR) was used to analyse the Mascot data generated. The false-discovery rate (FDR) for the peptide identity threshold and peptide identity or homology threshold generated by Mascot's reverse peptide decoy databases ranged from 1.9 % to 9.74 % and 4.55 % to 9.71 % respectively; the majority of samples FDR fell between 4-6% for the QE HF and Orbi-trap LC-MS/MS. X!Tandem was selected within Scaffold to validate the peptides and proteins identified previously by Mascot. Peptide identifications were accepted if they exceeded the 95% probability threshold for the Peptide prophet algorithm with Scaffold delta mass correction. Protein identities were accepted if contained 2 unique peptides and established at greater than 99% probability (Supplementary Tables S1-3). The samples were considered individually and as distinct groups for each experimental condition.

## 3. Results and Discussion

Proteins were recovered with varying degrees of success from the 153 dental calculus specimens analysed to determine the optimal current methodology and experimental parameters for extraction. The 'success' of the protein recovery was evaluated on the following criteria: 1) the number of proteins assigned identity within the specimen, 2) the diversity of the proteins identified in terms of proteome complexity without chemically induced post-translational modifications and 3) sequence coverage of the peptides assigned to dietary, oral and disease related proteins previously identified in modern and archaeological dental calculus specimens and used in the course of this study as diagnostic markers of specimen preservation.

### 3.1 Optimal demineralisation acid and methodology

Several considerations were taken into account in the present study, prior to the extraction of the preserved proteins/ peptides entrapped within these ancient dental calculus specimens: (1) laboratory environment (in terms of possible sources of contaminants), (2) demineralisation/decalcification of the sample to achieve a robust and optimal proteome, (3) limitations of the material in terms of preservation/decay and amount, (4) instrumental limitations and (5) cost. Previous publications within the field have used a variety of acids for the decalcification of bone/ dental calculus specimens, including either filter-assisted or gel-assisted extraction methods, but a direct comparison of these methods has not been attempted. In this section we consider the role of the demineralisation acid in combination with the method of extraction in terms of proteome recovery and purpose a weak acid compared to a strong acid will yield a better quality proteome in terms of reduced chemically induced deamidation during processing, while maintaining a diverse proteome for analysis in accordance with the observations of Procopio and Buckley[40] and Tromp et al.[46]. We also propose the SP3 methodology with no molecular weight cut off would improve the number of unique peptides and overall peptide recovery compared to an ultrafiltration based extraction method. The gel-assisted, GASP methodology was considered and excluded for analysis based on the limited amount of material available and potential complications which may arise, for example, if acrylamide is not removed effectively from the samples.

Two acids (EDTA and HCl) used for demineralization and three methodologies (ultrafiltration, SP3 and acetone precipitation) were compared to assess these hypotheses. The soluble and insoluble protein fractions were isolated for each dental calculus specimen, enabling the exploitation of that individual's retrievable dental calculus proteome. For example, the solubility of the globular whey proteins, present in dairy products, increases at pH values farther away from their isoelectric point of 4.5, thus a higher pH (above 4.5) increases the probability of releasing these proteins from their aggregated complexes (formed

at a lower pH) trapped within the dental calculus mineral matrix. Therefore, EDTA and HCl-insoluble fractions (in GuHCl), with a pH of 8.0 and 8.5 respectively, have a higher probability of isolating dairy whey proteins compared to HCl. For the purpose of this study, soluble proteins were defined as proteins that were released from the dental calculus after 24-66 h without the addition of the secondary weak acid/ strong chaotrope – buffered GuHCl solution. Insoluble proteins were defined as those released after treatment with the demineralization agent and chaotropic solution.

Only the HCl-soluble protein fractions were precipitated with acetone, avoiding any potential disruption to the covalent bonding between the carboxylate surface of these paramagnetic beads and the proteins of interest, which could occur as a consequence of low pH and bias the data collection. The pH level for EDTA and the insoluble protein fractions were considered 'less of a risk' at 8.0 and 8.5, as EDTA has been previously utilized to isolate proteins from a variety of materials with SP3. Therefore, it was not considered necessary for comparison, in terms of an acetone precipitated soluble protein fraction for EDTA or acetone precipitated insoluble protein fractions for EDTA and HCl.

Proteome recovery for these samples, in terms of identifying the optimal combination of demineralization agent and methodology (i.e. ultrafiltration, SP3 or acetone precipitation) was evaluated based on the number of unique peptides, the number of unique proteins, and observed deamidation levels in relation to the total amount of peptides for the grouped experimental condition (Figure 2) and as individual specimens (Figure 3) for the soluble and insoluble protein fractions. Deamidation rates (Figure 2D and 3C) were calculated by dividing the number of deamidated unique peptides by the total number of unique peptides per sample[40]. 478 proteins were identified across the three individuals analysed prior to the removal of suspected contaminants, these were subcategorized as either human (5.2 %), bacterial (78.7 %), dietary (0.4 %) or contaminants (15.7 %) and, comparable to the ratios observed by Jersie-Christen *et al.*[7] (Figure 2A). Potential experimental contaminants (e.g. bovine cytochrome c, lysozyme c, alcohol dehydrogenase, bovine serum albumin, apo-transferrin, and beta-galactosidase from the Pierce 6 Protein Digest), laboratory based- and environmental contaminants (e.g. keratins and those stated in the cRAP database excluding the Casein protein family), and contaminants present in the experimental blanks were removed prior to data interpretation. Contaminants in experimental blanks constituted 15.7 % of contaminants observed in Figure 2A. In addition, all human proteins (5.2 %, observed for Figure 2A) were excluded despite the possibility of some of these proteins being from the specimen themselves (e.g. the immune response protein neutrophil defensin-1, a known component of saliva). 101 proteins were considered as processing or environmental contaminants and their peptides removed from the data.
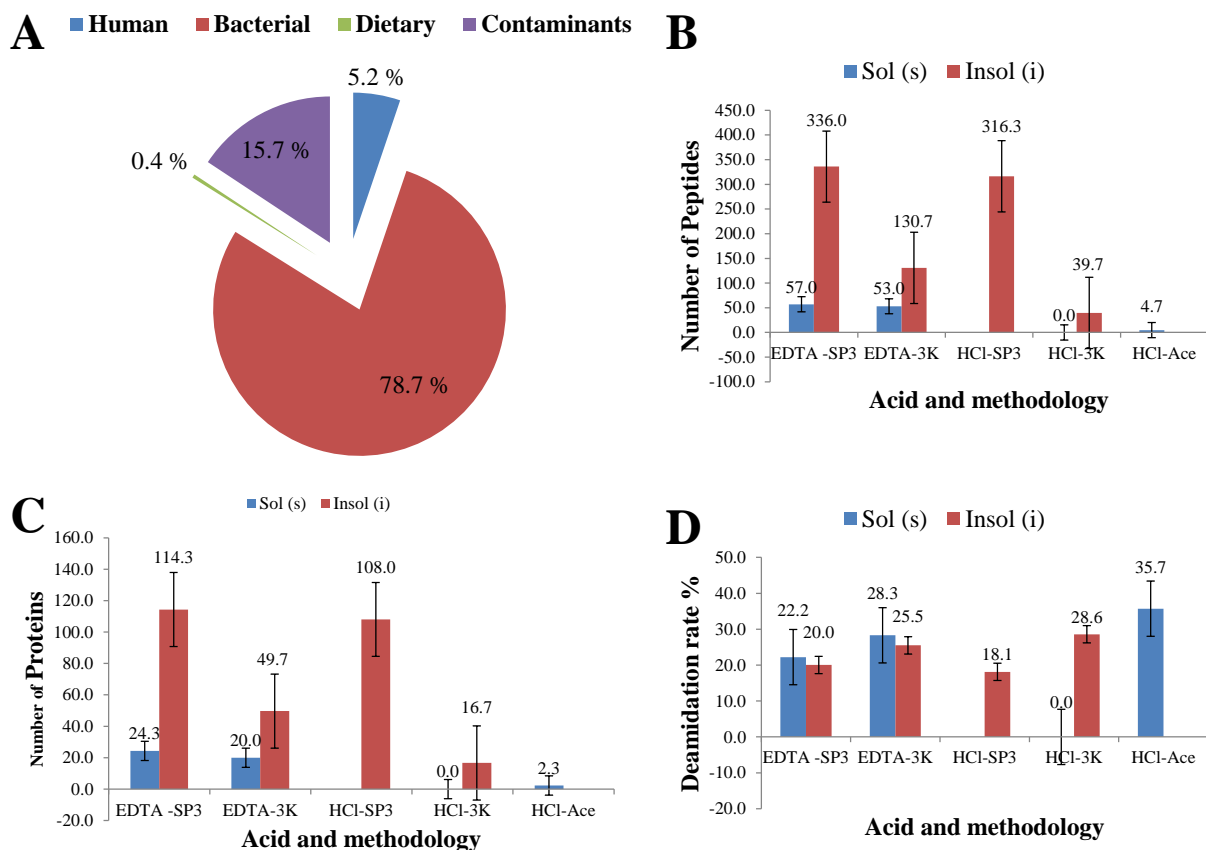
*Figure 2 – Bar charts representing the average and standard errors associated with the grouped experimental conditions to determine the optimal combination of demineralization and protein extraction methodology* (A) Categorization of the 478 proteins identified within the dental calculus specimens: FIR-14, FIR-15 and STP-29 into bacterial, human, dietary and contaminants – experimental and environmental (%), (B) the total number of unique peptides obtained, (C) the total number of proteins and (D) the calculated deamination ratios obtained post-experimental and -environmental contaminant removal.

As shown in figure 2B and 2C, for the grouped experimental conditions, EDTA and SP3 resulted in the recovery of the largest number of unique peptides matched as well as the highest proteome variability for the soluble (57 unique peptides and 24.3 proteins) and insoluble fractions (336 unique peptides and 114.3 proteins). HCl and SP3 resulted in a lower unique peptide score than EDTA with SP3 for the insoluble fraction (316 unique peptides and 108 proteins) and failed to isolate any peptides/proteins of interest for the soluble fraction, contrary to the predicted proteome variety expected, based on the proteome variety observed for modern bone samples demineralized in HCl[40]. Surprisingly, the samples extracted using the 3K ultrafiltration devices seem to have lost a significant amount of proteins in terms of peptide numbers and protein numbers observed than expected. For the 3 kDa ultrafiltration extracted HCl fractions, no peptides/proteins of interest were isolated for the soluble protein fraction as observed with SP3 and only 39.7 unique peptides and 16.7 proteins were observed for the insoluble fraction. Ultrafiltration did not significantly decrease the peptide/ protein numbers observed for the EDTA soluble fraction (53 unique peptides and 20 proteins) but the amount of peptides and proteins was decreased by more than 50 % for the insoluble fraction (130.7 unique peptides and 49.7 proteins). Some loss can be attributed to the nominal

weight cut-off limit for the 3 kDa ultrafiltration devices as predicted, but a loss of greater than a half of the peptide and protein numbers overall across the two acids indicates a secondary factor is likely responsible. In addition, the soluble_protein fraction isolated by acetone precipitation, failed to recover sufficient peptides/proteins of interest for analysis (4.7 unique peptides and 2 proteins) supporting the hypothesis that a secondary factor is contributing to the loss of peptides/protein from the acid.

Deamidation rates for these EDTA and HCl treated fractions (Figure 2D) indicate that a greater amount of deamidated proteins were observed for the soluble protein fraction isolated by acetone (35.7%). However, this calculated rate was discarded as a reference point due to the poor rate of peptides/ proteins of interest recovery rate from the soluble fraction using this methodology, with only 14 peptides recovered for this grouped condition and 5 of these peptides deamidated, leading to an artificially inflated deamidation rate. The second highest deamidation rate was attributed to the insoluble fraction for HCl-3K (28.6%) followed by soluble fraction for EDTA-3 kDa (28.3%), insoluble fraction for EDTA-3K (25.5%), soluble fraction for EDTA-SP3 (22.2%), insoluble fraction for EDTA-SP3 (20.0%), and insoluble fraction for HCl-SP3 (18.1%). Overall, the deamidation rates observed across the acids and methodologies assessed are higher than observed with modern bone (Procopio and Buckley[40]) as expected for archaeological specimens. Unfortunately, from the conflicting data it is difficult to conclude whether a weaker acid is the more efficient demineralization agent in terms of reducing the deamidation observed for the specimens as a grouped experimental condition. For example, while EDTA soluble and insoluble fractions for SP3 and ultrafiltration (using a 3 kDa ultrafiltration device) vary by ~5.5 – 5.9%, the HCl insoluble fractions vary by 10.5%, with the ultrafiltration fraction exhibiting more deamidation than its SP3 counterpart for both EDTA and HCl. In contrast, the HCl treated fraction for SP3, is lower than the observed deamidation rate with EDTA for the same methodology, differing by ~2%.

Evaluation of the specimens as individuals within the experimental conditions concurs with the previous observation that EDTA and SP3 appear to be the optimal demineralising agent and method for extraction in terms of unique peptide numbers and protein identity (Figure 3A and 3B). Additionally, if we consider the proteins recovered from the specimens in terms of origin (geographical climate and age), it is clear from the data that EDTA is more consistent at isolating the dairy proteins β-lactoglobulin and the α-S1 and κ caseins from either the soluble and/or insoluble protein and that SP3 tends to isolate more peptides per protein (Figure 3A). This is consistent with the data previously published for this demineralization agent using FASP[9-10]. For example, the soluble fraction for SP3 yielded more peptides (3 peptides) than its ultrafiltration counterpart (1 peptide) for FIR-14 treated with EDTA. In addition, in the absence of β-lactoglobulin, the casein proteins α-S1 and K were isolated from the soluble and insoluble EDTA fractions for STP-29 only with the SP3 methodology. Interestingly, although the number of peptides and proteins identified isolated for the HCl fractions was generally lower than for the EDTA extracted fractions, the same phenomenon was observed for the insoluble fraction of FIR-15, multiple peptides were matched for SP3 from the insoluble fraction of FIR-15 treated with HCl (2 peptides) and only a singular peptide from ultrafiltration.

For all three specimens, HCl appears to have isolated the lowest unique peptide numbers (Figure 3A) and protein identities (Figure 3B) compared with EDTA overall except in the case of the HCl-SP3 vs EDTA-ultrafiltration insoluble fractions for FIR-15. One possible explanation for this reduction in proteome complexity for HCl is hydrolysis due to the demineralization length of 66 h. Procopio and Buckley[40] observed a greater number of unique peptides and identified proteins for modern bone with HCl

than with the equivalent EDTA treatment for a demineralization length of 66 h. However, this could be due to the nature of the material analysed; Procopio and Buckley[40] analysed modern bone samples, whereas the dental calculus analysed here was not only ancient but also presumably partially digested from salivary amylase activity in the oral cavity while the subject was alive. It is possible that upon release from the inorganic matrices, these low-level, partially digested/decayed proteins were more sensitive to the acid in terms of hydrolysis compared to modern bone counterparts. This may also explain the absence of plant-based dietary proteins identified within these specimens.

For deamidation rates observed in the three test samples (FIR-14, FIR-15 and STP-29), the data do not readily discern whether EDTA or HCl minimizes the chemically induced post-modification observed (Figure 3C). The HCl-soluble fractions precipitated with acetone, soluble EDTA fraction for FIR-014 and FIR-015 (isolated by SP3) and the HCl-insoluble fraction for FIR-14 treated (isolated using ultrafiltration) exhibit, high deamidation ratios as a result of their reduce peptide numbers; deamidation rates across the remaining data points for these acids and methodologies, range between 19.5-28.3 (EDTA) and 13.6-22.9 (HCl) suggesting that HCl yields peptides with lowest deamidation rates, therefore contradicting the overall trends aforementioned.

These discrepancies, in terms of the observed deamidation rates for the demineralisation agents: HCl and EDTA, as grouped experimental conditions or for an individual specimen can be partly attributed to the small number of samples tested and sampling bias caused by the probable loss of peptides during sample preparation (ultrafiltration, precipitation and SP3) in combination with the potential loss of peptides due to hydrolysis from overexposure to the demineralization agent HCl. Hence, based on the data obtained, we consider EDTA the more appropriate demineralization agent for isolating the entrapped ancient proteome within dental calculus in terms of number of unique peptides, number of identified proteins and the consistent presence of the dairy proteins β-lactoglobulin and caseins (α-S1 and κ) (Figure 2 and 3). Further analysis with a larger sample set is required to determine whether HCl or EDTA is the better agent for demineralisation or are equivalent. Comparative analysis of the unique peptide and protein identities for the methodologies for an individual specimen and/ or grouped experimental conditions, reveal SP3 appears to be the better methodology for sample preparation
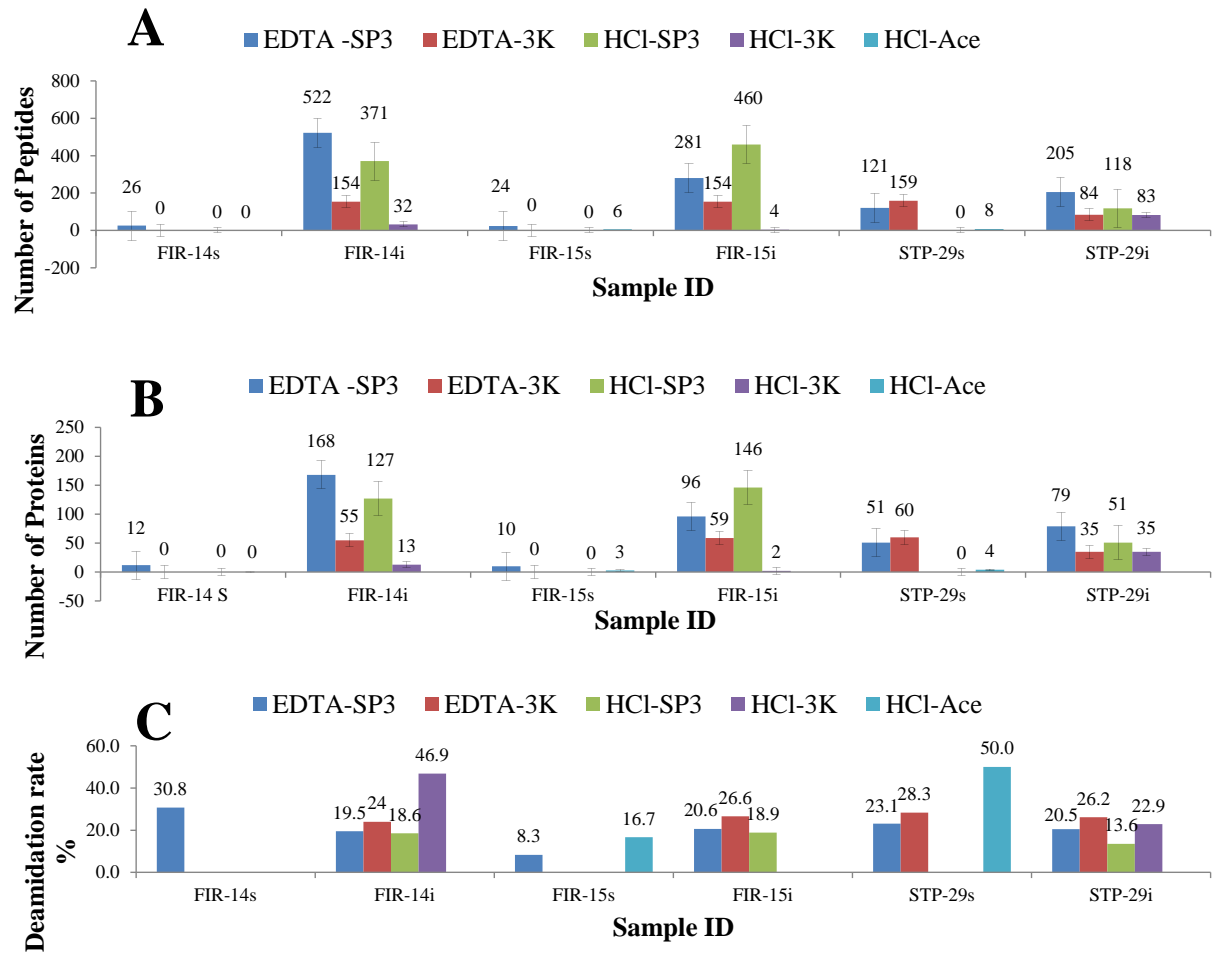
***Figure 3 – Bar charts representing the individual specimens to determine the optimal demineralization and protein extraction methodology*** (A) the total number of unique peptides obtained, (B) the total number of proteins and (C) the calculated deamination ratios obtained. The asterixes and crosses above the bars in (A) & (B) represent the presence of either a unique peptide associated with β-lactoglobulin or a member of the casein protein family respectively. Red symbols indicate 2 or more peptides were identified. Black symbols indicate only 1 peptide was observed. Peptides and proteins from soluble and insoluble fractions represented by (s) and (i) respectively.

## 3.2 Comparing Soluble and Insoluble Fractions

Insoluble fractions were isolated after acid treatment using GuHCl-buffered solution and processed separately from the EDTA and HCl fractions. From the unique peptide numbers (Figure 3A) and number of proteins identified (Figure 3B) it is clear that the majority of proteins in terms of peptides and identification are isolated from the insoluble fractions for two of the specimens: FIR-014 and FIR-015 and the reverse is true of the third specimen: STP-029. From the composition of the individual specimens in terms of protein identity, it is possible for the third specimen STP-029, more soluble proteins were preserved than observed for the insoluble proteins (as indicated by the presence of the casein protein family) and thus explaining this discrepancy. As with the previous data groupings, EDTA and SP3 appear to isolate the higher number of unique peptides and protein identity in terms of number; both the soluble and insoluble

proteins fractions concur except for FIR-015 were the HCl SP3 isolated fraction appears to have greater numbers as previously observed for the grouped condition. One possible explanation for this discrepancy is SP3 paramagnetic bead loss during processing of the EDTA SP3 fraction. Overall greater peptide numbers and protein identifications are observed with the insoluble protein fractions compared to the soluble fractions. However, as the data for FIR-014, FIR-015 and STP-029 indicates the soluble protein fraction in some cases may contain proteins of interest e.g. the dairy proteins β-lactoglobulin and members of the casein family and therefore can't be discounted. Hence, the soluble and insoluble protein fractions for an individual were pooled prior to LC-MS/MS for analysis from this point onwards, to retain the proteome complexity while reducing the cost of analysis.

## 3.3 Comparing LC-Orbitrap with QE

A key element of any proteomics experiment is the sensitivity of the LC-MS/MS instrument used to characterize the recovered proteome. No study to date has compared the sensitivity of Orbi-trap LC-MS/MS to QE LC-MS/MS in measuring proteins stored in ancient dental calculus. Previous ancient dental calculus research has been primarily conducted using QE LC-MS/MS, as the instrument parameters for QE are theoretically superior to the Orbi-trap in terms of instrument sensitivity and, consequently, should yield a higher protein recovery rate in terms of number of unique peptides and proteins. Here we explore the use of the Orbi-trap LC-MS/MS as a valid alternative to QE LC-MS/MS for identifying proteins of interest within an archaeological specimen in the absence of access to a QE LC-MS/MS instrument.

All samples containing a single peptide for the dairy proteins, i.e., β-lactoglobulin or a member of the casein protein family, were analysed on the QE. Dairy peptides identified by Swiss-Prot and HOMD, but below the sample-specific ion score cut-off for homology recommended by Mascot. Several samples which tested negative for the presence of any dairy proteins were re-run to exclude the possibility of data loss due to insufficient instrument sensitivity (for the Orbi-trap). As theorized, an increase in the amount of unique peptides (above the mascot ion-score cut-off and indicative of an identity or extensive homology of ($p<0.05$)) was observed for the dairy proteins, β-lactoglobulin and the caseins with the QE instrument when compared to the data previously generated with the Orbi-trap (Figure 4). Protein identities were assigned for these that were previously 'one-hit wonders' if they exceeded 2 unique peptides and established at greater than 99% probability. The dairy peptide isolated taxonomically belonged to the family Bovidae (Figure 5 and 6). Specimens that tested negative for these dairy proteins such as STP-003, STP-020 and STP-035 previously remained negative (excluded from Figure 4). Notably, samples that previously tested positive for β-lactoglobulin, including dairy peptides below the Mascot ion score cut-off and Scaffold Protein identity and Peptide identity thresholds of 99% and 95%, but had no evidence of containing a member of the Caseins (e.g. TEL-022 and INS-001), demonstrated an increase in the amount of the Casein family proteins in the QE compared with the Orbi-trap.
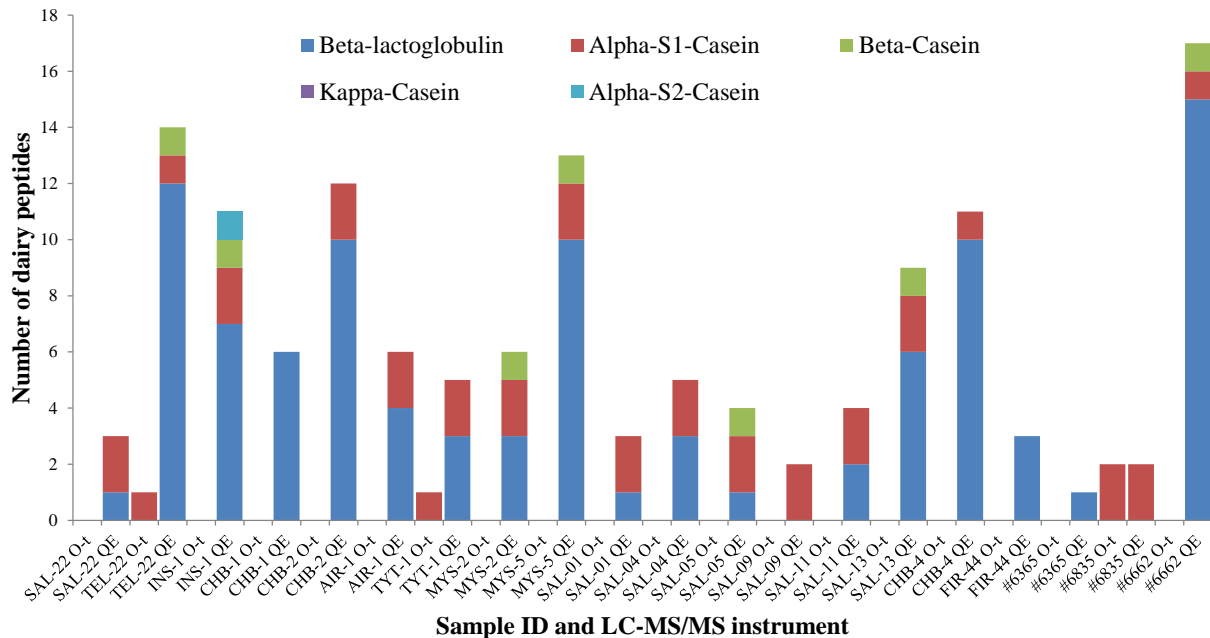
**Figure 4 - Comparative Orbi-trap and QE analysis for the diagnostic dietary proteins: β-lactoglobulin and Casein.**

*The total number of unique peptides observed for β-lactoglobulin and the casein protein family as individual units within the archaeological specimens belong to Bovidae. Only peptides above or equivalent to the Scaffold protein and peptide identity thresholds of 99% and 95% respectively are recorded in this figure accounting for the 'jump' of absence to presence for most of the samples even though these samples were chosen based on the presence of at least one unique dairy peptide within the original Mascot searches.*

Initial comparison of the data generated for the same specimen revealed that the QE instrument appears to identify more unique peptides related to a specific protein than its Orbi-trap counterpart (Figure 4). Such an increase in the number of unique peptides for any given specimen could enable a more specific protein identity assignment in terms of taxonomy from family to genus and/or species (Figure 5).  For example, a comparison of the published sequences available in UniProt (accessions P02758, P02754, P02756, P67975 and P67976 for horse, cow, goat and sheep respectively), identifies numerous positions of amino acid variation between species of interest. Of these, two are located at the N-terminus of the protein at position 13 (M= Horse, T = Cow, A = Goat & Sheep) and 16 (I= Horse & Goat, A= Cow & V= Sheep) and therefore are specific to species.

Assuming the increase in peptides does not reflect the identification of numerous post modified versions of the same peptide generated during the individual's lifetime, diagenetic alteration, or chemically induced post- modification introduced during protein extraction and analysis procedures. Thus, to evaluate whether the increase in peptide number and protein identification observed for the QE instrument reflected an increased 'sensitivity' in terms of instrument performance, directly correlating to the identification of smaller peptides within the dental calculus rather than diagenetic alteration or post-translational modifications (natural and chemically induced) that the Orbi-trap previously missed (due to a combination of reduced sensitivity and/ or peptide calling), we plotted the sequence coverage for these samples (Figure 5).
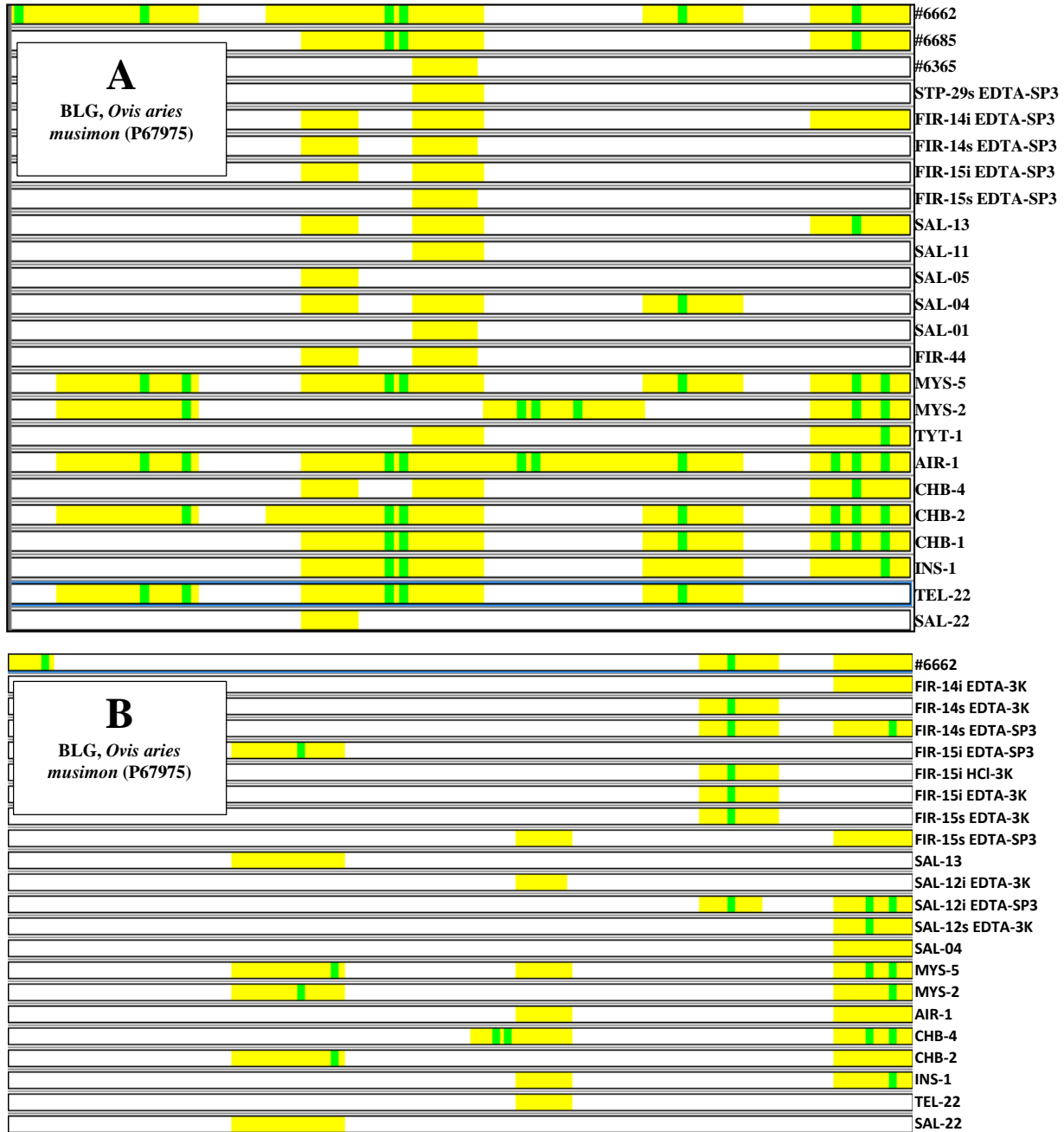
**Figure 5 - B-lactoglobulin sequence coverage for *Ovis orientalis musimon* (Swiss-Prot Accession number: P67975) identified by Orbi-trap and QE LC-MS/MS analysis for the ancient dental calculus specimens identified as containing this dietary protein on the basis of 1 or more peptides matched by Mascot (including those below the ion score cut-off).** (A) Visual representation of the sequence coverage for the specimens analysed by QE LC-MS/MS. (B) Visual representation of the sequence coverage for the specimens analysed by Orbi-trap LC-MS/MS. Peptide coverage and post-modifications are highlighted in yellow and green respectively.

A comparative evaluation of the sequence coverage observed for β-lactoglobulin (Figure 5) confirms the increase in peptide numbers observed in Figure 4 reflects an increase in peptide coverage if we compare the same specimen on the two instruments in most cases [Figure 5A and B, samples #6662

(Mongolia), SAL-013, SAL-004, MYS-005, MYS-002, AIR-001, CHB-004, CHB-002, INS-001 and TEL-022 (Russia)]. The QE instrument also increased the amount of 'dairy positive' samples from 14 (Orbi-trap) to 22 (QE) at the singular peptide level with a Scaffold protein and peptide identity of 99 % and 95 % respectively. This phenomenon is particularly evident for data generated using QE LC-MS/MS versus Orbi-trap LC-MS/MS, when considering the Casein proteins, which appears to be more susceptible to archaeological decay than β-lactoglobulin (Figure 4, sample SAL-22 and Supplementary Figure S1).

The number of unique peptides and the sequence coverage observed for comparative analysis of the Orbi-trap and QE LC-MS/MS instruments concur; for the analysis of ancient dental calculus specimens, the QE appears to be the instrument of choice in terms of identifying low level and smaller peptide fragments generated either by protein decay overtime or a combination of protein decay and tryptic digest for analytical purposes. To further assess the impact of instrument performance on the quality of peptide recovered and validate the presence of these peptides within the sample, it is necessary to examine spectra for each individual peptide e.g., at the fragment ions generated. It is clear from the spectra obtained (e.g., Supplementary Figure S3) that the QE instrument yielded better 'coverage' of the peptides identified from β-lactoglobulin (Supplementary Figure S3B and S3C) and casein (Supplementary Figure S3D) compared to the spectra observed for Orbi-trap (Supplementary Figure S3A), concurring with the notable absence of an α-S1-casein spectra with the Orbi-trap instrument.

### 3.4 β-lactoglobulin vs Casein

Milk is a complex oil in water emulsion secreted by the mammalian gland, composed of fats, proteins, lactose, minerals, enzymes, cells, hormones, immunoglobulins and vitamins[48]. The protein composition of milk can be broadly categorized as either insoluble (casein - αs1, αs2, β, and κ) or soluble (serum/whey - α-lactalbumin, β-lactoglobulin, serum albumin, immunoglobulins, lactoferrin, transferrin, calcium-binding protein, prolactin, folate-binding protein, and proteosepeptone) proteins[49-50]. The exact proportion of these milk proteins varies with time since parturition, stage of lactation, age and health of animal, between species and also with season (the last presumably related to dietary intake and/or habitat)[48]. Casein (αs1, αs2, β, and κ) and β-lactoglobulin constitute the major protein components of milk, for example, within fresh cow's milk β-lactoglobulin accounts for ~ 58% (of the total whey protein fraction) and Caseins ~80 % of these soluble and insoluble protein fractions, respectively. Members of the Casein protein family altogether constitute the most abundant protein fraction present in fresh milk, yet β-lactoglobulin is more routinely observed as a dietary protein isolated from ancient dental calculus specimens[4,7,9]. This study (Figure 4, specimens CHB-001 and FIR-044) concurs with these previously published findings regarding the preservation of β-lactoglobulin over the caseins. αs1, αs2, β, and κ casein collectively constitute between 50 – 80 % of the total protein content of fresh milk with differences in the proportion and overall total protein content differences between species (Table 2). For example, the casein fractions for sheep and cattle milk comprises ~80% of total protein content, compared with 50% observed for horse milk. In addition, despite the α and β caseins being members of a single gene family and conserved across the ruminants (Supplementary Figure 2), the physical and chemical properties of these proteins varies in terms of micelle size, mineralization, hydration and response to environmental factors (including heat and pH). K casein is homologous to y-fibrinogen and also exhibits these differences across species. The low rates of protein identification for αs1, αs2 and κ casein compared to β-casein (except in the case of cattle) can be explained by their lower proportional representation across the ruminants. However, these proteins comprise altogether ~80% of the total protein content of milk compared to the 10-12% observed in β-lactoglobulins. Based on abundance alone at ingestion, the expectation is that caseins would be

observed at a higher rate than β-lactoglobulin universally within ancient dental calculus samples obtained from individuals that consumed milk and possibly other dairy related products.

**Table 2 - Composition of casein proteins in cow, sheep and goat milk (%)[a]**

| Casein | Cow | Goat | Sheep |
|--------|------|------|-------|
| αS1[b] | 39.7 | 5.6 | 6.7 |
| αS2[b] | 10.3 | 19.2 | 22.8 |
| β[b] | 32.7 | 54.8 | 61.6 |
| κ[b] | 11.6 | 20.4 | 8.9 |

[a] Adapted from Balthazar *et al.*, 2017

[b] Percentage of total casein

This discrepancy in protein abundances in consumed milk and those detected in ancient dairy proteome may be explained by differences in the complexity of protein structure and the types of bonds present within these ancient biomolecules. The structural architecture of a protein in the vast majority of cases is determined by its primary amino acid sequence. We propose variation in structure impacts protein survival in the burial environment and may explain some of the differences observed for the proteins recovered here. In bone, the most frequently observed surviving protein from archaeological specimens is the supramolecule forming (e.g. fibrils, beaded filaments, anchoring fibrils, and networks), triple helical glycoprotein collagen followed by the non-collagenous, hydroxyapatite binding proteins A2HSG, serum albumin and biglycan[51-52]. The survival of these proteins within the burial environment can be partly attributed to their self-contained micro-environment within the mineralized bone as is the case with DNA; however this would not explain the discrepancies observed in terms of abundance for these collagenous and non-collagenous proteins. Nor can these discrepancies be explained by the natural abundance of these proteins alone but perhaps more due to their structural properties.

β-lactoglobulin is a structurally more complex and rigid protein at the monomer level than the Caseins. It has a pH dependent quaternary structure and is composed of a conical barrel or calyx, open at only one end, and comprised of eight anti-parallel beta-strands with a 3-turn α helix on the outer surface barrel and has been successfully crystallized. Yet its biological function remains elusive[53]. In contrast, attempts to crystallize the caseins, a family of hydrophobic, heat stable, calcium-binding proteins[54] thus far have proved unsuccessful in terms of elucidating the 'full' secondary and tertiary structures of these proteins as monomers. Leading to the past misnomer of the caseins being perceived as random coil or natively denatured monomers as a side effect of their more flexible nature (compared to the whey proteins counterparts). Farrell proposes these self-associating proteins are intrinsically unstructured and exist in either a natively unfolded state (αs1), pre-molten globular state (αs2) or molten globular state (κ, β), displaying a compact native secondary structure with little tertiary fold, high degrees of hydration and flexible side-chains capable of reacting rapidly to environment changes due to a lack of a fixed tertiary structure as observed with other secretory calcium-binding proteins[55,56,57].

However, at the pH of saliva (6.2 to 7.6), neither β-lactoglobulin nor any member of the caseins would exist as a monomer. Rather, β-lactoglobulin would exist as a dimer (36,700 Da) and the caseins as micelles composed of $10^3$- $10^6$ casein molecules and ranging in diameter from 50 to 700 nm. Casein micelles

are highly hydrated and sponge-like colloidal particles, consisting of multiple entities (αs1, αs2, β, and κ) held together and organized by their non-covalent intermolecular binding interactions and stabilized by the presence of a calcium phosphate[58-59]. These porous micelle structures are theoretically more open to reacting with water and other substances from the environment and within the micelle itself in the case of the former. Water increases the decay rate observed for ancient DNA through nucleophilic attack. Only 15 % of the known water content of a micelle (~4 g of water per g of protein) is bound to the protein itself, the remainder being simply occluded within the particle and leaving these protein structures more susceptible to this nucleophile than β-lactoglobulin[60,61]. Thus, poor survival rates of the casein proteins compared to β-lactoglobulin for archaeological specimens may be attributed in part to their simplified structure and protein arrangement within the porous micelle thus increasing their susceptibility to the usual suspects of pH, temperature, water content and soil microbiome within the burial environment. An argument further supported by the identification of peptides buried within the center of these protein arrangements by LC-MS/MS Orbi-trap and/or QE for both the hydrophobic walled β-lactoglobulin (Figure 6A) and the κ -casein (Figure 6D) protected αs1, αs2, β-caseins (Figure 6B and 6D). K–casein extends outwards on the periphery of the micelle and is responsible for the steric stabilization of the 'interlocked lattice' structure in which the casein molecules surround calcium phosphate nanoclusters which extend as short-chains between the interlocking points and outwards at the particle periphery[62]. The increased presence of κ–casein on the periphery of the micelle, in combination with its low abundance within milk, might also explain its absence within ancient dental calculus more generally as this member of this protein family is most vulnerable to decay caused by salivary enzymes and external factors (such as nucleophile water) within the burial site.

Digestive enzymes within saliva, amylase (salivary) and lingual lipase may also contribute to preservation discrepancies between β-lactoglobulin and the caseins in dental calculus. However, the actions of both lingual lipase and salivary amylase can be discounted as a contributory factor for degradation of these proteins based on the target specificity of these enzymes. Lingual lipase specifically targets lipids breaking them down into triglycerides[63], while salivary amylase specifically targets the 1-4 α glycosidic bonds formed between carbon-1 and carbon-4 of two monosaccharides[64-65], as neither β-lactoglobulin or caseins are lipids or monosaccharides (sugars), both should remain intact when initially encased in dental plaque on the surface of the tooth.
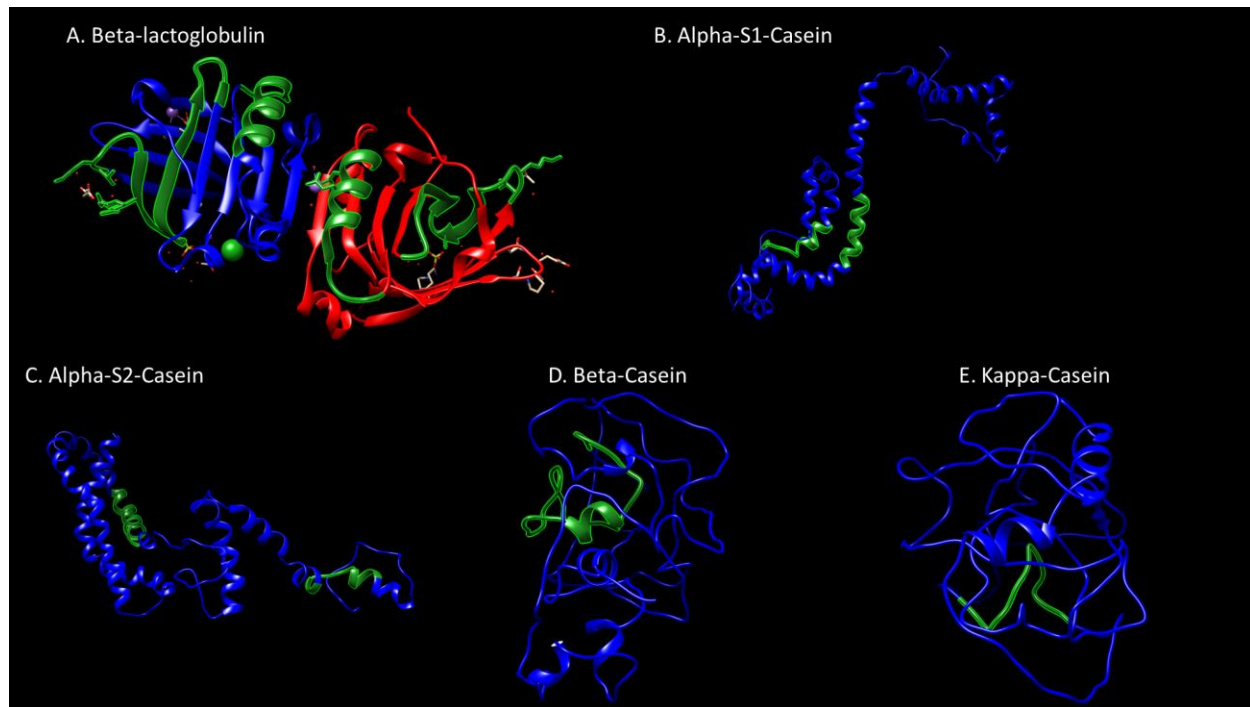
***Figure 6 - Structural representations of *β*-lactoglobulin and the caseins protein family.*** *A. Crystal structure of* Ovis aries *β-lactoglobulin dimer adapted from Loch et al., 2014 (PDB File: 4NLJ). B. Predicted α-S1-casein monomer from* Ovis aries*. C. Predicted α-S2-casein monomer from* Ovis aries*. D. Predicted beta-casein monomer from* Ovis aries*. E. Predicted κ-casein monomer from* Ovis aries*. Casein predicted structures generated by Phyre²(Version 2.0). Monomers (either blue or red) and surviving peptides identified by LC-MS/MS QE and Mascot (white with green border).*

## Conclusions

In this study we compared three methodologies for the extraction and isolation of ancient proteins from dental calculus considering the additional experimental parameters of demineralization acid and instrumentation. We observed a significant increase in the amount of ancient proteins recovered using a combination of EDTA and SP3 paramagnetic beads from these archaeological specimens compared to demineralization with a more aggressive acid: HCl and an ultrafiltration based methodology. We suggest this method can be used as an alternative to filtration based methodologies for the extraction and analysis of ancient proteins from dental calculus specimens, increasing the probability of isolating bacterial (related to disease) and dietary proteins of interest historically. In addition, to our knowledge, our study is the first detailed analysis of proteins isolated from ancient dental calculus using the SP3 methodology to comparatively assess the LC-MS/MS instruments: Orbi-trap Elite (Thermo Fisher Scientific, Waltham, MA) and QE HF (Thermo Fisher Scientific, Waltham, MA) mass spectrometers. Revealing the Orbi-trap Elite is less sensitive (in terms of peptide number and sequence coverage recovery) and therefore less suitable for the study of ancient proteins than its QE HF counterpart. The role of protein structure within the archaeological context of preservation could offer an explanation of the discrepancies observed between the dairy proteins: β-lactoglobulin and casein, but would benefit further investigation.

## Acknowledgements

**SUPPORTING INFORMATION:**

**Table S1. Scaffold peptide report for dental calculus samples FIR14, FIR15 and STP29 for comparative analysis of ultrafiltration vs acetone precipitation vs SP3.**

**Table S2. Scaffold peptide report for dental calculus samples identified as potentially positive for the presence of dairy based on 1 unique dairy peptide identified by LC-MS/MS Orbit-trap Elite and assigned for re-analysis using the QE-HF.**

**Table S3. Scaffold peptide report for dental calculus samples identified by the LC-MS/MS Orbi-trap Elite as containing proteins associated with the ingestion of dairy and re-analysed using the Q E-HF instrumentation.**

**Figure S1. Alpha-S1-casein sequence coverage for *Bos taurus* (Swiss-Prot accession number: P02662) and *Ovis orientalis musimon* (Swiss-Prot Accession number: P4653) identified by Orbi-trap and QE LC-MS/MS analysis for the ancient dental calculus specimens identified as containing this dietary protein on the basis of 1 or more peptides identified by Mascot (including those below the ion score cut-off).**

**Figure S2. CLUSTAL Omega (1.2.4) multiple sequence alignment for the sheep casein proteins.**

**Figure S3. Spectral comparison of the peptides identified as either β-lactoglobulin or αS1-casein assigned to Bovidae for the Early Bronze Age dental calculus specimen TEL-22.**

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023399.

# References

1.      Armitage, P. L., The Extraction and Identification of Opal Phytoliths from the Teeth of Ungulates *J Archaeol Sci* **1975,** 2, (3), 187-197.

2.      De La Fuente, C., Flores, SV. and Moraga, ML., Human bacterial DNA from dental calculus: a new source of genetic material. *Archaeometry* **2012** 55, (4), 766-788.

3.      Preus, H. R.; Marvik, O. J.; Selvig, K. A.; Bennike, P., Ancient bacterial DNA (aDNA) in dental calculus from archaeological human remains. *J Archaeol Sci* **2011,** 38, (8), 1827-1831.

4.      Warinner, C.; Hendy, J.; Speller, C.; Cappellini, E.; Fischer, R.; Trachsel, C.; Arneborg, J.; Lynnerup, N.; Craig, O. E.; Swallow, D. M.; Fotakis, A.; Christensen, R. J.; Olsen, J. V.; Liebert, A.; Montalva, N.; Fiddyment, S.; Charlton, S.; Mackie, M.; Canci, A.; Bouwman, A.; Ruhli, F.; Gilbert, M. T.; Collins, M. J., Direct evidence of milk consumption from ancient human dental calculus. *Sci Rep* **2014,** 4, 7104.

5.      Adler, C. J.; Dobney, K.; Weyrich, L. S.; Kaidonis, J.; Walker, A. W.; Haak, W.; Bradshaw, C. J.; Townsend, G.; Sołtysiak, A.; Alt, K. W., Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat Genet* **2013,** 45, (4), 450.

6.      Dewhirst, F. E.; Chen, T.; Izard, J.; Paster, B. J.; Tanner, A. C.; Yu, W. H.; Lakshmanan, A.; Wade, W. G., The human oral microbiome. *J Bacteriol* **2010,** 192, (19), 5002-17.

7.      Jersie-Christensen, R. R.; Lanigan, L. T.; Lyon, D.; Mackie, M.; Belstrøm, D.; Kelstrup, C. D.; Fotakis, A. K.; Willerslev, E.; Lynnerup, N.; Jensen, L. J., Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. *Nat Commun* **2018,** 9, (1), 1-12.

8.      Liu, B.; Faller, L. L.; Klitgord, N.; Mazumdar, V.; Ghodsi, M.; Sommer, D. D.; Gibbons, T. R.; Treangen, T. J.; Chang, Y.-C.; Li, S., Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PloS One* **2012,** 7, (6), e37919.

9.      Charlton, S.; Ramsøe, A.; Collins, M.; Craig, O. E.; Fischer, R.; Alexander, M.; Speller, C. F., New insights into Neolithic milk consumption through proteomic analysis of dental calculus. *Archaeol Anthropol Sci* **2019,** 11, (11), 6183-6196.

10.     Wilkin, S.; Miller, A. V.; Taylor, W. T.; Miller, B. K.; Hagan, R. W.; Bleasdale, M.; Scott, A.; Gankhuyg, S.; Ramsøe, A.; Uliziibayar, S., Dairy pastoralism sustained eastern Eurasian steppe populations for 5,000 years. *Nat Ecol Evol* **2020,** 4, (3), 346-355.

11.     Yang, Y., Dairying transformed Mongolia. *Nat Ecol Evol* **2020,** 4, (3), 288-289.

12.     Hardy, K.; Buckley, S.; Collins, M. J.; Estalrrich, A.; Brothwell, D.; Copeland, L.; García-Tabernero, A.; García-Vargas, S.; De La Rasilla, M.; Lalueza-Fox, C., Neanderthal medics? Evidence for food, cooking, and medicinal plants entrapped in dental calculus. *Naturwissenschaften* **2012,** 99, (8), 617-626.

13.     Radini, A.; Tromp, M.; Beach, A.; Tong, E.; Speller, C.; McCormick, M.; Dudgeon, J. V.; Collins, M. J.; Rühli, F.; Kröger, R., Medieval women's early involvement in manuscript production suggested by lapis lazuli identification in dental calculus. *Sci Adv* **2019,** 5, (1), eaau7126.

14.     White, D. J., Dental calculus: recent insights into occurrence, formation, prevention, removal and oral health effects of supragingival and subgingival deposits. *Eur J Oral Sci* **1997,** 105, (5), 508-522.

15.     Hayashizaki, J.; Ban, S.; Nakagaki, H.; Okumura, A.; Yoshii, S.; Robinson, C., Site specific mineral composition and microstructure of human supra-gingival dental calculus. *Arch Oral Biol* **2008,** 53, (2), 168-174.

16.     Hillson, S., *Dental Anthropology*. Cambridge University Press: Cambridge, 1996.

17.     Lieverse, A. R., Diet and the aetiology of dental calculus. *Int J Osteoarchaeol* **1999,** 9, (4), 219-232.

18.     Marsh, P. D.; Bradshaw, D. J., Dental plaque as a biofilm. *J Ind Microbiol* **1995,** 15, (3), 169-175.

19.     Hanihara, T.; Ishida, H.; Ohshima, N.; Kondo, O.; Masuda, T., Dental calculus and other dental disease in a human skeleton of the Okhotsk Culture unearthed at Hamanaka‐2 site, Rebun‐Island, Hokkaido, Japan. *Int J Osteoarchaeol* **1994,** 4, (4), 343-351.

20.     Mackie, M.; Hendy, J.; Lowe, A. D.; Sperduti, A.; Holst, M.; Collins, M. J.; Speller, C. F., Preservation of the metaproteome: variability of protein preservation in ancient dental calculus. *STAR* **2017,** 3, (1), 58-70.

21.     Hillson, S., *Teeth*. Cambridge University Press: Cambridge, 2005.

22.     Aufderheide, A. C.; Rodríguez-Martín, C.; Langsjoen, O., *The Cambridge encyclopedia of human paleopathology*. Cambridge University Press Cambridge: 1998; Vol. 478.

23.     Allentoft, M. E.; Collins, M.; Harker, D.; Haile, J.; Oskam, C. L.; Hale, M. L.; Campos, P. F.; Samaniego, J. A.; Gilbert, M. T. P.; Willerslev, E., The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc Royal Soc B* **2012,** 279, (1748), 4724-4733.

24.     Elsner, J.; Schibler, J.; Hofreiter, M.; Schlumbaum, A., Burial condition is the most important factor for mtDNA PCR amplification success in Palaeolithic equid remains from the Alpine foreland. *Archaeol Anthropol Sci* **2015,** 7, (4), 505-515.

25.     Geigl, E. M., On the circumstances surrounding the preservation and analysis of very old DNA. *Archaeometry* **2002,** 44, (3), 337-342.

26.     Smith, C. I.; Chamberlain, A. T.; Riley, M. S.; Stringer, C.; Collins, M. J., The thermal history of human fossils and the likelihood of successful DNA amplification. *J Hum Evol* **2003,** 45, (3), 203-217.

27.     Carter, D. O.; Tibbett, M., Microbial decomposition of skeletal muscle tissue (Ovis aries) in a sandy loam soil at different temperatures. *Soil Biol Biochem* **2006,** 38, (5), 1139-1145.

28.     Carter, D. O.; Yellowlees, D.; Tibbett, M., Moisture can be the dominant environmental parameter governing cadaver decomposition in soil. *Forensic Sci Int* **2010,** 200, (1-3), 60-66.

29.     Forbes, S. L.; Stuart, B. H.; Dent, B. B., The effect of the burial environment on adipocere formation. *Forensic Sci Int* **2005,** 154, (1), 24-34.

30.     Forbes, S. L.; Stuart, B. H.; Dent, B. B., The effect of the method of burial on adipocere formation. *Forensic Sci Int* **2005,** 154, (1), 44-52.

31.     Lowe, A. C.; Beresford, D. V.; Carter, D. O.; Gaspari, F.; O'Brien, R. C.; Stuart, B. H.; Forbes, S. L., The effect of soil texture on the degradation of textiles associated with buried bodies. *Forensic Sci Int* **2013,** 231, (1-3), 331-9.

32.     Cappellini, E.; Welker, F.; Pandolfi, L.; Ramos-Madrigal, J.; Samodova, D.; Rüther, P. L.; Fotakis, A. K.; Lyon, D.; Moreno-Mayar, J. V.; Bukhsianidze, M.; Rakownikow Jersie-Christensen, R.; Mackie, M.; Ginolhac, A.; Ferring, R.; Tappen, M.; Palkopoulou, E.; Dickinson, M. R.; Stafford, T. W.; Chan, Y. L.; Götherström, A.; Nathan, S. K. S. S.; Heintzman, P. D.; Kapp, J. D.; Kirillova, I.; Moodley, Y.; Agusti, J.; Kahlke, R.-D.; Kiladze, G.; Martínez-Navarro, B.; Liu, S.; Sandoval Velasco, M.; Sinding, M.-H. S.; Kelstrup, C. D.; Allentoft, M. E.; Orlando, L.; Penkman, K.; Shapiro, B.; Rook, L.; Dalén, L.; Gilbert, M. T. P.; Olsen, J. V.; Lordkipanidze, D.; Willerslev, E., Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. *Nature* **2019,** 574, (7776), 103-107.

33.     Meyer, M.; Fu, Q.; Aximu-Petri, A.; Glocke, I.; Nickel, B.; Arsuaga, J.-L.; Martínez, I.; Gracia, A.; de Castro, J. M. B.; Carbonell, E., A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **2014,** 505, (7483), 403-406.

34.     Orlando, L.; Ginolhac, A.; Zhang, G.; Froese, D.; Albrechtsen, A.; Stiller, M.; Schubert, M.; Cappellini, E.; Petersen, B.; Moltke, I., Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **2013,** 499, (7456), 74-78.

35.     Campos, P. F.; Kristensen, T.; Orlando, L.; Sher, A.; Kholodova, M. V.; Götherström, A.; Hofreiter, M.; Drucker, D. G.; Kosintsev, P.; Tikhonov, A., Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (Saiga tatarica) since the Pleistocene. *Mol Ecol* **2010,** 19, (22), 4863-4875.

36.     Molak, M.; Ho, S. Y., Evaluating the impact of post-mortem damage in ancient DNA: a theoretical approach. *J Mol Evol* **2011,** 73, (3-4), 244-255.

37.      Olalde, I.; Allentoft, M. E.; Sánchez-Quinto, F.; Santpere, G.; Chiang, C. W. K.; DeGiorgio, M.; Prado-Martinez, J.; Rodríguez, J. A.; Rasmussen, S.; Quilez, J.; Ramírez, O.; Marigorta, U. M.; Fernández-Callejo, M.; Prada, M. E.; Encinas, J. M. V.; Nielsen, R.; Netea, M. G.; Novembre, J.; Sturm, R. A.; Sabeti, P.; Marquès-Bonet, T.; Navarro, A.; Willerslev, E.; Lalueza-Fox, C., Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **2014,** 507, (7491), 225-228.

38.      Child, A., Towards and understanding of the microbial decomposition of archaeological bone in the burial environment. *J Archaeol Sci* **1995,** 22, (2), 165-174.

39.      Weyrich, L. S.; Duchene, S.; Soubrier, J.; Arriola, L.; Llamas, B.; Breen, J.; Morris, A. G.; Alt, K. W.; Caramelli, D.; Dresely, V.; Farrell, M.; Farrer, A. G.; Francken, M.; Gully, N.; Haak, W.; Hardy, K.; Harvati, K.; Held, P.; Holmes, E. C.; Kaidonis, J.; Lalueza-Fox, C.; de la Rasilla, M.; Rosas, A.; Semal, P.; Soltysiak, A.; Townsend, G.; Usai, D.; Wahl, J.; Huson, D. H.; Dobney, K.; Cooper, A., Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **2017,** 544, (7650), 357-361.

40.      Procopio, N.; Buckley, M., Minimizing Laboratory-Induced Decay in Bone Proteomics. *J Proteome Res* **2017,** 16, (2), 447-458.

41.      Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J., Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol* **2014,** 10, (10), 757.

42.      Sielaff, M.; Kuharev, J. r.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U., Evaluation of FASP, SP3, and iST protocols for proteomic sample preparation in the low microgram range. *J Proteome Res* **2017,** 16, (11), 4060-4072.

43.      Moggridge, S.; Sorensen, P. H.; Morin, G. B.; Hughes, C. S., Extending the compatibility of the SP3 paramagnetic bead processing approach for proteomics. *J Proteome Res* **2018,** 17, (4), 1730-1740.

44.      Cleland, T. P., Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. *J Proteome Res* **2018,** 17, (11), 3976-3983.

45.      Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J., Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **2019,** 14, (1), 68-85.

46.      Tromp, M.; Buckley, H.; Geber, J.; Matisoo-Smith, L., EDTA decalcification of dental calculus as an alternate means of microparticle extraction from archaeological samples. *Journal of Archaeological Science: Reports* **2017,** 14, 461-466.

47.      Jeong, C.; Wilkin, S.; Amgalantugs, T.; Bouwman, A. S.; Taylor, W. T. T.; Hagan, R. W.; Bromage, S.; Tsolmon, S.; Trachsel, C.; Grossmann, J.; Littleton, J.; Makarewicz, C. A.; Krigbaum, J.; Burri, M.; Scott, A.; Davaasambuu, G.; Wright, J.; Irmer, F.; Myagmar, E.; Boivin, N.; Robbeets, M.; Rühli, F. J.; Krause, J.; Frohlich, B.; Hendy, J.; Warinner, C., Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc Natl Acad Sci U S A* **2018,** 115, (48), E11248-e11255.

48.      Balthazar, C.; Pimentel, T.; Ferrão, L.; Almada, C.; Santillo, A.; Albenzio, M.; Mollakhalili, N.; Mortazavian, A.; Nascimento, J.; Silva, M., Sheep milk: physicochemical characteristics and relevance for functional food development. *Compr Rev Food Sci F* **2017,** 16, (2), 247-262.

49.      Selvaggi, M.; Laudadio, V.; Dario, C.; Tufarelli, V., Major proteins in goat milk: an updated overview on genetic variability. *Mol Biol Rep* **2014,** 41, (2), 1035-1048.

50.      Selvaggi, M.; Laudadio, V.; Dario, C.; Tufarelli, V., Investigating the genetic polymorphism of sheep milk proteins: a useful tool for dairy production. *J Sci Food Agric* **2014,** 94, (15), 3090-3099.

51.      Wadsworth, C.; Buckley, M., Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. *Rapid Commun Mass Spectrom* **2014,** 28, (6), 605-615.

52.      Wadsworth, C.; Procopio, N.; Anderung, C.; Carretero, J.-M.; Iriarte, E.; Valdiosera, C.; Elburg, R.; Penkman, K.; Buckley, M., Comparing ancient DNA survival and proteome content in 69 archaeological cattle tooth and bone samples from multiple European sites. *J Proteomics* **2017,** 158, 1-8.

53.      Sawyer, L., β-Lactoglobulin. In *Advanced Dairy Chemistry*, McSweeney, P. L., Fox. Patrick F., Ed. Springer: New York, 2013; Vol. 1A:Proteins: Basic Aspects, pp 319-386.

54.     Paulsson, M.; Dejmek, P., Thermal denaturation of whey proteins in mixtures with caseins studied by differential scanning calorimetry. *J Dairy Sci* **1990,** 73, (3), 590-600.

55.     Farrell, H. M.; Kumosinski, T. F.; Malin, E. L.; Brown, E. M., The caseins of milk as calcium-binding proteins. In *Methods in Molecular Biology. calcium Binding Protein Protocols*, Vogel, H. J., Ed. Humana Pree: Totowa, NJ, 2002; Vol. 172, pp 97-140.

56.     Farrell Jr, H.; Malin, E.; Brown, E.; Qi, P., Casein micelle structure: What can be learned from milk synthesis and structural biology? *Curr Op Coll & Interf Sci* **2006,** 11, (2-3), 135-147.

57.     Farrell Jr, H.; Qi, P.; Brown, E.; Cooke, P.; Tunick, M.; Wickham, E.; Unruh, J., Molten globule structures in milk proteins: implications for potential new structure-function relationships. *J Dairy Sci* **2002,** 89, 459-471.

58.     Linderstrøm-Lang, K., Studies on casein. III. On the fractionation of casein. *Compt Rend Trav Lab Carlsberg Ser Chim* **1929,** 17, 1-116.

59.     McMahon, D. J.; Oommen, B., Supramolecular structure of the casein micelle. *J Dairy Sci* **2008,** 91, (5), 1709-1721.

60.     De Kruif, C.; Holt, C., Casein micelle structure, functions and interactions. In *Advanced Dairy Chemistry Proteins*, McSweeney, P. L., Fox. Patrick F., Ed. Kluwer Academic/Plenum: New York, 2003; Vol. 1A: Proteins: Basic Aspects, pp 233-276.

61.     Farrell, H.; Brown, E.; Hoagland, P.; Malin, E., Higher order structures of the caseins: a paradox? In *Advanced Dairy Chemistry Proteins*, McSweeney, P. L., Fox. Patrick F., Ed. Kluwer Academic/Plenum: New York, 2003; Vol. 1A: Proteins: Basic Aspects, pp 203-232.

62.     McMahon, D.; Oommen, B., Casein Micelle Structure, Functions, and Interactions. In *Advanced Dairy Chemistry*, McSweeney, P. L., Fox. Patrick F., Ed. Springer: New York, 2013; Vol. 1A:Proteins: Basic Aspects, pp 185-210.

63.     Kulkarni, B. V.; Mattes, R. D., Lingual lipase activity in the orosensory detection of fat by humans. *Am J Physiol Regul Integr Comp Physiol* **2014,** 306, (12), R879-R885.

64.     Des Gachons, C. P.; Breslin, P. A., Salivary amylase: digestion and metabolic syndrome. *Curr Diab Rep* **2016,** 16, (10), 102.

65.     Mathews, C.; Holde, V. K., Biochemistry. In 2nd ed.; The Benjamin/Cummings Publishing Company: USA 1996; p 1159.