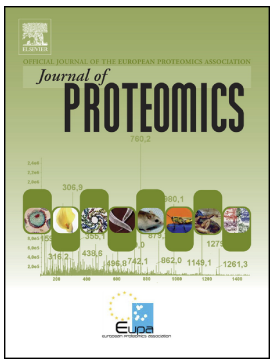
Analysis of 5000 year-old human teeth using optimized large-scale and targeted proteomics approaches for detection of sex-specific peptides

Carine Froment, Mathilde Hourset, Nancy Sáenz-Oyhéréguy, Emmanuelle Mouton-Barbosa, Claire Willmann, Clément Zanolli, Rémi Esclassan, Richard Donat, Catherine Thèves, Odile Burlet-Schiltz, et al.



Analysis of 5000 year-old human teeth using optimized large-scale and targeted proteomics approaches for detection of sex-specific peptides

Carine Froment, Mathilde Hourset, Nancy Sáenz-Oyhéréguy, Emmanuelle Mouton-Barbosa, Claire Willmann, Clément Zanolli, Rémi Esclassan, Richard Donat, Catherine Thèves, Odile Burlet-Schiltz, Catherine Mollereau

PII: S1874-3919(19)30320-3

DOI: https://doi.org/10.1016/j.jprot.2019.103548

Reference: JPROT 103548

To appear in: *Journal of Proteomics*

Received date: 24 January 2019

Revised date: 30 August 2019

Accepted date: 7 October 2019

Please cite this article as: C. Froment, M. Hourset, N. Sáenz-Oyhéréguy, et al., Analysis of 5000 year-old human teeth using optimized large-scale and targeted proteomics approaches for detection of sex-specific peptides, *Journal of Proteomics* (2018), https://doi.org/10.1016/j.jprot.2019.103548

Analysis of 5000 year-old human teeth using optimized large-scale and targeted proteomics approaches for detection of sex-specific peptides

Carine Froment[1], Mathilde Hourset[2,3], Nancy Sáenz-Oyhéréguy[2], Emmanuelle Mouton-Barbosa[1], Claire Willmann[2,3], Clément Zanolli[2,4], Rémi Esclassan[2,3], Richard Donat[2], Catherine Thèves[2], Odile Burlet-Schiltz[1] and Catherine Mollereau[2]

[1] Institut de Pharmacologie et Biologie Structurale (IPBS), Université de Toulouse, CNRS, UPS, Toulouse, France.

[2] Laboratoire d'Anthropobiologie Moléculaire et Imagerie de Synthèse (AMIS), Université de Toulouse, CNRS, UPS, Toulouse, France

[3] Faculté de chirurgie dentaire de Toulouse, Université de Toulouse, UPS, Toulouse, France.

[4] Laboratoire PACEA, UMR 5199 CNRS, Université de Bordeaux, Pessac, France.

Corresponding authors:
Dr Catherine Mollereau catherine.mollereau-manaute@ipbs.fr
Laboratoire AMIS
Faculté de médecine, 37 allées Jules Guesde
31073 Toulouse Cedex 03, France
Tel : 33 561 14 55 13
and
Dr Odile Burlet-Schiltz, odile.schiltz@ibps.fr
IPBS
205, Route de Narbonne BP 64182
31077 Toulouse Cedex 04, France
Tel: 33 561 17 55 47

Keywords: Paleoproteomics; Data dependent acquisition; Parallel reaction monitoring; Tooth; Amelogenin; Sex determination

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We are very grateful to Xavier Mata, Gabriel Renaud, and Franklin Delehelle (AMIS Toulouse) for their help with computational analyses, and to Ludovic Orlando (AMIS) and Jean Sébastien Saulnier-Blache (I2MC, Toulouse) for helpful discussion and comments during the preparation of the manuscript. The work was supported by CNRS (PEPS blanc 2016 and DefiXlife 2018-2019), in part by the Région Occitanie, European funds (Fonds Européens de Développement

Régional, FEDER), Toulouse Métropole, and by the French Ministry of Research with the Investissement d'Avenir Infrastructures Nationales en Biologie et Santé program (ProFI, Proteomics French Infrastructure project, ANR-10-INBS-08).

1. Introduction

Tooth is one of the most abundant fossil remain in archaeological sites. It contains a number of phylogenetic and life history traits that are recorded and preserved in the mineralized dental tissues. Such traits, including taxonomy, growth and development, sexual dimorphism, diet, gestation and perinatal life aspects, pathogens, etc..., can be extracted from tooth morphostructural characteristics as well as from (bio)molecular content [1-3]. Tooth thus represents a remarkable source of information for anthropologists.

Although genetic information can be retrieved from past specimens, proteins best survive at longer times making paleoproteomics an attractive approach for providing original and valuable information complementary to morphological and genetic studies [4]. Investigation of ancient proteins in fossils is now possible owing to the improvement of mass spectrometer performances in terms of high resolution and sensitivity for protein analysis [5]. Such approach has proven its potential to get an insight on past (patho)physiology, biological process, phenotype and lifestyle [4, 6] and to make protein-based phylogenetic reconstruction for samples where no ancient DNA is available [4, 7]. In the particular case of ancient human teeth, paleoproteomics offers the possibility to investigate health status [8, 9] and diet [10-12] of past populations. Another interest of ancient tooth proteome comes from the possibility to sex the individuals by detecting peptides specific to the X (AMELX) and Y (AMELY) chromosome-encoded gene products of amelogenin [13-17]. This information is essential for paleoanthropologists facing the issue of sex determination and dimorphism.

Tooth provides several advantages to conduct paleoproteomics studies. Since they are less porous than bone, the dental mineralized tissues are expected to preserve proteins from contamination and degradation [7], a main issue in ancient protein studies [18]. The hard (enamel, dentine and cement) and soft (pulp) tissues express specific and complex proteomes [19, 20], which can be retrieved in ancient specimens [9, 21]. In addition to collagens that represent the main proteins in tooth root, a large diversity of non-collageneous proteins (NCP) are potentially accessible [7]. Among them, amelogenin (AMEL), ameloblastin (AMBN), enamelin (ENAM), amelotin (AMTN), dentin sialophosphoprotein (DSPP), cementum protein (CEMP1) are tooth-specific proteins. Their identification ensures therefore the reliability of the analysed samples. Moreover, single amino acid variations detectable in some tooth proteins can also be of clinical [20, 22], morphological [23] and phylogenetic interest [7, 22, 24, 25].

The objective of the present study was to develop optimized MS-based proteomics approaches for tooth fossil identification, especially focused on the characterization of peptides showing sex-specificity or phenotypic information. For that, eleven Neolithic human teeth dated to ~5000-years ago and originating from Mont Aimé multiple burrial (France) have been analysed in comparison with modern samples by using shotgun nanoLC-MS/MS and a bioinformatics database search method adapted to the particular case of ancient proteins prone to degradation, damage and evolutionary variation [5, 26, 27]. The workflow was based

on an iterative database search strategy [28-30] to overcome the limitation of conventional single-step method in managing the presence of high-level of protein modifications, incomplete enzymatic or non-specific hydrolysis. In addition, a customized protein database consisting in the Human Uniprot database hand-upgraded with genetically variant products presenting an interest for phenotypic, taxonomic or dental diseases (Table S1) was used. This approach allowed the identification of nearly 1500 proteins in the totality of archaelogical samples. They were mostly identified in the no enzyme search modes, indicating that, when only considering the conventional semi-tryptic database search mode in shotgun analyses, a number of peptides issued from randomly degraded proteins may be missed contributing thus to a loss of information. Based on the identification of the sex-specific peptides TALVLTPLK, IALVLTPLK and WYQSIRPPYP of amelogenin, a targeted MS approach using a parallel reaction monitoring (PRM) mode [31] was set up to maximize the sensitivity and the reproducibility of detecting these unique peptides for sex estimation. This led to confirm the sex of individuals in all the samples.

2. Materials and Methods

2.1. Samples

The archaeological material originates from the French Neolithic necropole of Mont Aimé (3650-3380 cal BC) located in the Bassin Parisien [32]. Eleven teeth (Table S2) were randomly collected from individuals of the hypogea 1 (1H01, 1H05, 1H06, 1H07, 1H12, 1H15, 1H18) and the hypogea 2 (2H08, 2H10, 2H12, 2H23). Two present-day teeth (CW02, CW03) from males undergoing dental surgery were obtained by collaborators after written informed consent. Just after extraction, teeth were briefly cleaned with alcohol and kept at -20°C until use. The genetic sex of the neolithic individuals (4 females and 7 males) was previously determined by using the multiplex PCR amplification assay described in [33], in a study aiming at determining the Y-chromosome lineages of the male individuals from the Mont Aimé site (Sáenz-Oyhéréguy et al., in preparation). Teeth were manipulated under a laminar flow hood in a cleanroom laboratory dedicated to ancient DNA. The surfaces were cleaned with bleach (at 20% for 30 s), rinsed with sterile pure water and exposed to UV light (30 min on each side). After abrasion of the tooth surfaces by manual drilling with a Dremel instrument, whole teeth were totally reduced into a fine powder in liquid nitrogen using a Spex SamplePrep TM6870 Freezer/MillTM (Fisher Scientific). The grinding vials were extensively washed between samples to protect against cross-contamination. The powders were kept at -20°C until use.

2.2. Protein extraction and Trypsin digestion

Samples were prepared as four separate series for archaeological teeth (S2, S51, S52, S6), and series S7 for present-day individual, each including an extraction blank (Blk-E) sample with no material but exactly processed as the tooth samples. Protein extraction was performed by using a filter-aided sample preparation (FASP) protocol adapted from [9]. A total

of 10-50 mg of tooth powder was demineralized in 1 ml 0.5 M EDTA (pH 8) for 18 h at room temperature, under rotation. After centrifugation (5 min, 13000 rpm), the supernatant was harvested, supplemented with 100 µl of 1 M DTT, mixed with 9 ml 8 M urea in 0.1 M Tris pH8 for protein denaturation and ultra-filtered then through Amicon™ Ultra-4 (10kDa) centrifugal filter unit (4000g, swinging rotor, room temperature). The pellet was incubated in 300 µl lysis buffer (0.1 M Tris pH8, 0.1 M DTT, 4% SDS) for 2h at 60 °C and centrifuged for 5 min at 13000 rpm. The supernatant was mixed with 2 ml 8 M urea in 0.1 M Tris pH8 and ultra-filtered through the same centrifugal filter unit as the corresponding supernatant. After a wash of the ultrafiltration unit with 2 ml 8 M urea in 0.1M Tris pH8, proteins were alkylated by incubation with 500 µl of 50 mM 2-Chloroacetamide in 8M urea, 0.1 M Tris pH8, for 20-30 min at room temperature in the dark. The ultrafiltration unit was then washed 2-times with 1 ml urea 8M urea in 0.1 M Tris pH8, followed by two washes with 1 ml and then 0.5 ml 50 mM ammonium bicarbonate to replace buffer. Proteins retained on the filter were dissolved in 50 mM ammonium bicarbonate. 10 µl aliquot was harvested for quantification using the Qubit protein assay kit (Fisher Scientific). Proteins were subjected to enzymatic digestion by adding 2 µg sequencing grade modified porcine trypsin (Promega) in 100 µl 50 mM ammonium bicarbonate and overnight incubation at 37 °C. The digestion was prolonged the next day for 4-6 h with 2 µg additional trypsin. The tryptic peptide mixture was recovered by centrifugation over a new tube, followed by an additional elution with 500 µl 50 mM ammonium bicarbonate. The entire eluate was transferred to a microtube and dried by using a centrifugal vacuum concentrator and kept at -20°C until mass spectrometry analysis.

2.3. Shotgun nanoLC-MS/MS analysis

The dried peptides were resuspended with 0.05% trifluoroacetic acid in 2% acetonitrile at a concentration of 1µg/µl, and analysed by online nanoLC using an UltiMate® 3000 RSLCnano LC system (Thermo Scientific, Dionex) coupled to an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Scientific, Bremen, Germany). 1µg of the samples were loaded on a 300 µm ID x 5 mm PepMap C18 pre-column (Thermo Scientific, Dionex) at 20 µl/min in 2% acetonitrile, 0.05% trifluoroacetic acid. After 5 min of desalting, peptides were on-line separated on a 75 µm ID x 50 cm C18 column (in-house packed with Reprosil C18-AQ Pur 3 µm resin, Dr. Maisch; Proxeon Biosystems, Odense, Denmark) equilibrated in 95% of buffer A (0.2% formic acid), with a gradient of 5 to 25% of buffer B (80% acetonitrile, 0.2% formic acid) for 80 min then 25% to 50% for 30 min at a flow rate of 300 nL/min.

The instrument was operated in the data-dependent acquisition (DDA) mode using a top-speed approach (cycle time of 3s). The survey scans MS were performed in the Orbitrap over m/z 350-1550 with a resolution of 120,000 (at 200 m/z), an automatic gain control (AGC) target value of 4e5, and a maximum injection time of 50 ms. Most intense ions per survey scan were selected at 1.6 m/z with the quadrupole and fragmented by Higher Energy Collisional Dissociation (HCD). The monoisotopic precursor selection was turned on, the intensity

threshold for fragmentation was set to 50,000 and the normalized collision energy was set to 35%. The resulting fragments were analysed in the Orbitrap with a resolution of 30,000 (at 200 m/z), an automatic gain control (AGC) target value of 5e4, and a maximum injection time of 60 ms. The dynamic exclusion duration was set to 30 s with a 10 ppm tolerance around the selected precursor and its isotopes. For internal calibration the 445.120025 ion was used as lock mass.

Each sample was subjected to two independent LC-MS/MS runs (R1, R2) for assessing the identification reproducibility. To control for carry-over contamination, the MS workflow process included a washing step followed by two buffer MS runs, referred as injection blank (Blk-I), before and after each sample MS run, including blank samples.

2.4. Bioinformatics data processing of shotgun nanoLC-MS/MS data

Data obtained from the shotgun nanoLC-MS/MS analysis were processed using an iterative search method in two different search engines: Mascot 2.6.1 (Matrix Science, London, UK) in Proteome Discoverer™ software 2.1.1.21 (Thermo Fischer Scientific) and X!Tandem version X! Tandem Sledgehammer (2013.09.01.1) [34] in SearchGUI 3.2.14 [35]. Data were searched against the UniProtKB/Swiss-Prot protein database released 2016_02 with *Homo sapiens* taxonomy (20,295 sequences). The database was implemented with a number of isoforms and variants (listed in Table S1) relevant for dental diseases, morphostructure or taxonomic discrimination [22, 24], or retrieved from Ensembl.org.

The iterative database search performed with Mascot in Proteome Discoverer™ consisted of three steps (Figure 1) and was combined with the Percolator algorithm (version 2.05) for calculation of q-values and Posterior Error Probability (PEP) of peptide-spectrum matches (PSM) [36, 37]. For all steps, a target-decoy approach [38] was used for FDR estimation using reversed or randomised databases for semi-tryptic step or no enzyme steps, respectively. Mass tolerances in MS and MS/MS were set to 10 ppm and 20 mmu, respectively. Carbamidomethylation of cysteine was set as a fixed modification. Step 1: the MS raw files were pre-processed for selecting MS/MS spectra and the derived peaklists were searched using semi-tryptic enzyme specificity with a maximum of three missed cleavages. The main protein modifications affecting damaged ancient proteins were set as variable modifications: deamidation (N,Q), oxidation (M,P), carbamylation (K, N-terminal protein), and conversion to pyro-glutamic acid (N-terminal Q). Validated PSM based on Percolator q-value with a false discovery rate (FDR) worse than 1% were exported in mgf format for searching in the next step. Step 2: the pool of spectra imported from step 1 was searched with no enzyme specificity and only oxidation (M,P) as variable modifications. Validated PSM based on Percolator q-value with a false discovery rate (FDR) worse than 1% were exported in mgf format for searching in the next step. Step 3: the pool of spectra imported from step 2 was searched with no enzyme specificity and the same variable modifications as in step 1. In a final consensus step, all the processed data (msf files) were combined, validated and filtered with the following parameters:

only PSM with rank 1 and a Mascot ion score ≥ 20 were considered. PSM and peptides were validated based on Percolator PEP values at a FDR set to 1%. Then, the peptide identifications were grouped into proteins according to parsimony principles and filtered to 5% FDR. The estimated FDR and the PEP values are monitored for each step of the iterative database search and for the global consensus step (Table S3).

Using X!Tandem, protein identification was conducted against a concatenated target/decoy version of the customized human UniProtKB/Swiss-Prot protein database and the decoy sequences were created by reversing the target sequences in SearchGUI. Mass tolerances in MS and MS/MS were set to 10 ppm and 20 mmu, respectively. Carbamidomethylation of cysteine was set as a fixed modification. Mgf files were initially searched using tryptic enzyme specificity with a maximum of three missed cleavages and a limited number of variable modifications : oxidation (M,P). This method is known as a non-refined search. Then, a refined search of the "candidate proteins" identified in the non-refined search is performed allowing non-specific hydrolysis and supplemental protein modifications: deamidation (N,Q), carbamylation (K, N-terminal protein), Acetylation (N-terminal protein), Pyrolidone (E,Q, carbamidomethylated C). Peptides and proteins were inferred from the spectrum identification results using PeptideShaker version 1.16.15 [39]. Peptide Spectrum Matches (PSMs), peptides and proteins were validated at a 1.0%, 1.0%, and 5.0% False Discovery Rate (FDR) estimated using the decoy hit distribution, respectively.

For the label free approach, extracted-ion chromatograms (XIC) of the sex-specific peptides identified in archaelogical samples (TALVLTPLK, MH2+ at m/z 478.3130; IALVLTPLK, MH2+ at m/z 484.3325; MH2+ at m/z 654.3259; WYqSIRPPYP, MH2+at m/z 654.3259) were performed using Qual Browser in Xcalibur™ software 3.0.63 (Thermo Fischer Scientific) with a mass tolerance of 4 ppm and a 5 point boxcar smoothing. The ICIS algorithm was used for peak detection and peak area integration using the default parameters.

2.5. Targeted proteomic analysis for sex estimation

Based on the shotgun analysis results, three sex-specific peptides of amelogenin including TALVLTPLK, WYQSIRPPYP and the counterpart deamidated form, WYqSIRPPYP, from AMELX and IALVLTPLK from AMELY, were selected for targeted mass spectrometry analysis. Corresponding peptides used as standard references were synthetized as isotopically labelled C-terminus Lys U-$^{13}C_6$,$^{15}N_2$ or Arg U-$^{13}C_6$,$^{15}N_4$ heavy peptides. Both labelled peptides were coated together on a water-soluble biopolymer bead in a controlled amount (READYBEADS™, ANAQUANT, Villeurbanne, France). One coated READYBEADS™ was dissolved in 1ml of 0.05% trifluoroacetic acid in 2% acetonitrile for 5 min by vortexing, releasing the heavy isotope-labelled peptides at final concentrations of 1 pmol/µL for TALVLTPLK, 1.2 pmol/µL for IALVLTPLK, 3 pmol/µL for WYQSIRPPYP and 1.9 pmol/µL for WYqSIRPPYP (dilution named RB1). Subsequently, 10 µl of each sample at a concentration of 1µg/µl were spiked with 1µl of the standard heavy peptide solution. A 1.1µl volume of each spiked-in sample was analysed

by a targeted mass spectrometry approach conducted in PRM mode with the same instrument and chromatographic conditions as for shotgun nanoLC-MS/MS analyses. The PRM acquisition method combined two scan events starting with a full scan followed by targeted MS/MS of the doubly charged precursor ions for both light (at m/z 478.3130, 484.3312, 653.8326 and 654.3246, respectively) and heavy (at m/z 482.3201, 488.3383, 658.8367 and 659.3287, respectively) sex-specific peptides AMELX-TALVLTPLK, AMELX-WYQSIRPPYP, AMELX-WYqSIRPPYP and AMELY-IALVLTPLK along the complete chromatographic run. The full scan event employed a m/z 350–1550 mass range selection, an Orbitrap resolution of 60,000 (at 200 m/z), a target automatic gain control (AGC) value of 4e5, and maximum filling times of 50 ms. The targeted MS/ MS was run at an Orbitrap resolution of 60,000 (at 200 m/z), target AGC value of 1e5, and maximum filling times of 120 ms. The targeted peptides were isolated using a 2 m/z unit window with the quadrupole and fragmented by Higher Energy Collisional Dissociation (HCD) with normalized collision energy of 35 eV. All analyses were performed in triplicate for all the samples.

2.6. Bioinformatics data processing of PRM data

PRM data were processed in Skyline 3.7.0.11317 [40]. Firstly, the peptide search results (.dta files) of 1H12 and 1H15 shotgun data for which both sex-specific peptides markers were well identified, were used for spectral library building with a cut-off score of 0.95. All the extracted ion chromatograms (XICs) of selected fragments (Table S4) were manually inspected to ensure correct peak detection and integration. The MS/MS spectra of imported PRM data were matched (dotp ≥ 0.75) with that of the spectral library to confirm the sex-specific peptide identities. The heavy isotope-labelled sex-specific peptides were invoked as standard references for supplemental reliable identification and peak selection.

For the determination of linearity, response curves were generated by spiking the CW02 present-day sample with various dilutions of one coated READYBEADS^{TM}, and the LOD/LLOQ of the assays was calculated using the Quasar software [41] implemented as an external tool in Skyline (supplementary Materials and Methods, Table S5 and Figures S1-S4).

The peak areas intensity from the selected peptide transitions were exported for heavy and light sex-specific peptides (Table S6), filtered based on the calculated LOD and summed for each replicate in each sample. The mean of the summed peak area intensities with the standard deviation (SD) and coefficient of variation (CV) were calculated across triplicates for each peptide in each sample.

2.7. Data analysis

Graphic representations and statistical analysis of the data were performed using Prism 7 (GraphPad Software Inc., USA). Venn diagrams were drawn with online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/). Classification of the identified proteins

into functional categories according to GO terms was performed by using the functional annotation chart tool of DAVID (https://david.ncifcrf.gov/home.jsp, [42]).

All the RAW data files, the corresponding Proteome Discoverer™, X!tandem and Skyline output files, and the customized human database have been deposited to the ProteomeXchange Consortium [PMID 24727771] via the PRIDE partner repository [PMID 16041671] and can be accessed with the dataset identifier PXD014442.

3. Results and discussion

3.1. In-depth analysis of ancient human tooth proteomes based on shotgun nanoLC-MS/MS analysis and a dedicated iterative database search strategy using Proteome Discoverer™

The discovery phase was based on the nanoLC-MS/MS data-dependent analysis of the tryptic digestions of global protein extracts from eleven 5000 year-old human teeth (4 female and 7 male individuals) and two present-day teeth from male individuals. Working on ancient samples required a number of procedures and controls aiming at preventing and monitoring contamination from modern and ancient environments (see sample preparation). Specifically, extraction (Blk-E) and injection (Blk-I) blanks were included for the control of contamination and carry-over during the analytical process, and two independent LC-MS/MS analyses were run (R1, R2).

Ancient dental remains are expected to contain a large number of peptides showing non-tryptic cleavages due to post-mortem protein degradation by diagenesis process [18, 26], in addition to endogenous proteolysis by proteases such as MMP20 and KLK4 during the maturation and crystallization phases of the dental hard tissue formation [13, 20]. Furthermore, ancient proteins harbour a large set of putative modifications including the main damaged modifications such as deamidation of asparagine and glutamine, oxidation of methionine and proline, and carbamylation of peptide N-termini and lysine residues that could hamper trypsin digestion and contribute to a high rate of missed cleavages. The current single-step database search approaches are not very effective at identifying a large number of modifications and non-specific cleavages from complex proteome mixtures because the database search space expands exponentially as their number increases. This also increases the search time and false positive rate. Consequently, in order to overcome the drawbacks of conventional database search methods, some iterative search strategies such as ISPTM (Iterative Search for Peptide Identification with PTMs) have been already described [29] or implemented to conventional search engines such as in X!Tandem [28, 34]. We took advantage of the Proteome Discoverer™ software that offers the possibility to design iterative search methods, to tailor-make an iterative database search strategy in three steps to account for possible diagenetic amino-acid substitutions and non-enzymatic cleavages. The iterative search strategy (Figure 1) consists in making, at the end of a first semi-tryptic search with multiple dynamic modifications, additional successive searches (no enzyme and oxidation, no enzyme

and multiple dynamic modifications) on the pool of not well-identified spectra (PSM > 1% FDR) at each step. Furthermore, to get access to putative phenotypic or taxonomic information, the UniProtKB/Swiss-Prot protein reference database was implemented with 63 variants (Table S1), essentially tooth proteins. This customized human protein database was used for all the steps of the iterative database search.

Using this optimized shotgun proteomics approach, a total number of 1496 proteins were identified in the archaeological samples (Table S7), with 64.4% and 65.5 % overlap between the two injection replicates (Figure S5A). In total, 101 proteins were found in the extraction (Blk-E) and injection (Blk-I) blanks and were considered as contaminants (Table S8). They were mostly identified in the Blk-I, reflecting some background carry-over. However, only a part (47/101) of these contaminant proteins was retrieved in the samples (Figure S5B and Table S8), representating about 3% of the total number of proteins identified. They essentially included collagens and keratins, but also the alpha-2-HS-glycoprotein considered as endogenous to the samples in other paleoproteomics studies [21]. Deamidation of asparagine residues, and more specifically of glutamine residues which occurs at lower rate and is less dependent on buffer, represents another criterion to assess the endogenous origin of proteins in ancient samples [4, 7]. Among all possible glutamine and asparagine residues counted in the list of PSM from archaeological samples, 41% and 70% were found deamidated, respectively (Figure 2). Compared to modern samples, total and glutamine deamidation in ancient samples were 1.5 and 1.8 fold higher, respectively, indicating a significant decay in the 5000-year old proteins.

3.1.1. Proteins identified in the three steps of the iterative database search

As shown in figure 3A and Table S3, our iterative database search workflow allowed the identification of around 300 proteins per archaeological sample. The first step of the iterative database search workflow, corresponding to the usual semi-tryptic mode, identified 187 proteins in total (Figure S5C), with 50 to 130 proteins per sample (Figure 3A). This yield is consistent with the number of proteins identified in other studies on modern [43, 44] and ancient [17, 21] dentin, or ancient tooth root [9]. Interestingly, more than 50% of the proteins were identified in the no enzyme database searches only (Figure 3A), and particularly in the second step which only includes proline and methionine oxidations as variable modifications (Figures 3B and S5C). This is likely related to the endogenous high level of hydroxyproline in collagens and the presence of oxidised methionine in ancient collagens and non-collagenous proteins [24]. In contrast, all the extraction blanks controls show a completely reverse profile, as proteins were mostly identified after the first step of the identification process in the semi-tryptic database search mode (Figure 3A and 3B).

Compared to ancient samples, the analysis of modern teeth (Figure 3A and Table S3) gave a higher number of protein identification per sample (mean = 448) and the identification in the semi-tryptic database search mode was increased by 2.3 fold (mean per sample = 200).

However, similarly to ancient teeth, a large proportion (≥ 50%) of the proteins in modern samples were identified in the no enzyme search modes and were similarly distributed in the 3 steps (Figure 3A and 3B). This indicates thus that the high abundance of randomly fragmented and damaged proteins is not specific to archaeological samples but is also observed in present-day samples. This result suggests that beside the natural maturation of tooth-specific proteins during amelogenesis and dentinogenesis, a large part of protein degradation likely occurs at the time of death of the tissue by cellular processes, independently from the age of the specimen. The possibility that random degradation results from our biochemical extraction procedure is to be excluded since it is not observed in the extraction blanks that show a reverse profile of distribution of proteins in the iterative search modes. The results show therefore that, when only considering the classical semi-tryptic database search mode (step 1 only) in MS analyses, a number of peptides issued from randomly degraded proteins may be missed contributing to a loss of information.

Archaelogical and present-day proteomes shared 564 proteins (Figure S5D and Table S7). The classification of the identified proteins in each proteome into functional categories according to GO terms shows similar profiles between Neolithic (Figure 4A and 4B) and modern proteomes (Figure 4C and 4D), but different profiles between the semi-tryptic (Figure 4A and 4C) and the no enzyme (Figure 4B and 4D) search modes. More than 50% of the categories identified in the semi-tryptic search mode (55% and 60%, Figure 4A and 4C, respectively) were attributed to categories related to the mineralized dental and bone tissues (extracellular matrix and adhesion proteins, tooth, biomineralization, bone) and to pulp tissue (blood, immune and inflammatory systems). This proportion is reduced to 22% (Figure 4B) and 27 % (Figure 4D) in favor to categories related to cell cycle and transcription and to signaling and transduction processes in the no enzyme search modes. This indicates that proteins belonging to these later categories are likely more sensitive to death-induced proteolysis than structural or blood proteins. The proteins identified in the tooth proteomes of female and male individuals have been also compared by principal component analysis (data not shown), but no evidence for a differential distribution according to sex was observed, indicating no qualitative sex-dependency of total tooth proteomes.

In order to evaluate our iterative database search workflow, in terms of analysis depth, the shotgun nanoLC-MS/MS data were also processed by X!Tandem, a currently used search engine which allows to run an iterative database search using a double pass strategy. The double pass strategy identifies proteins in the sample using unmodified peptides (or minimally modified peptides) and assuming perfect proteolytic cleavage in a first pass (non-refined search).Then the database is reduced to include only those identified proteins and search it for a wide selection of modified peptides and missed/non-specific proteolytic cleavages in a second pass (refined search). Thus, our iterative database search approach differs from this double pass approach by refining the MS/MS spectra instead of refining the database. The

search parameters were set to match as much as possible our strategy, meaning that it included the same variable modifications and allowed non-enzymatic cleavage at both end. As shown in the figure S6 and Table S9, X!Tandem allowed to identify a total of 187 and 330 proteins in archeological and present-day samples, respectively. This number is far lower than that identified by our iterative search strategy, but is however very close to the number identified after the first step of the iterative search. Indeed, 76% and 73% of the proteins identified by X!Tandem in ancient and modern samples, respectively, are in common with the proteins identified by the semi-tryptic search mode in the corresponding samples (Figure S6B). In fact, even if a non-enzymatic cleavage was allowed in the refinement parameters of X!Tandem, the algorithm essentially identified semi-tryptic peptides, as verified by manual examination of the list of peptides that contained few non-enzymatic peptides. Moreover, the number of tooth-specific proteins identified by X!Tandem was lower (see below). These results demonstrate therefore that our iterative search strategy was more powerful to allow a maximum number of identification.

3.1.2.Tooth-specific proteins

The characteristic proteins of enamel (AMBN, AMEL, AMTN, ENAM), dentine (DSPP), cementum (CEMP1) and mineralized tissues (MGP, MMP20) were identified in nearly all samples by the iterative search mode, while some of them (ENAM, AMELY isoform 1, AMTN and CEMP1) were not detected by X!Tandem (Table 1), evidencing again the advantage of our iterative approach. However, probably because the signal from the enamel tissue was diluted into the whole tooth proteome, the tooth-specific proteins were identified with a low number of peptides compared to other proteomics analyses restricted to enamel [15, 45] and particularly focusing on endogenously cleaved peptides [13].

As shown in Figure 5 and Table S10, the amelogenin (AMEL) protein was identified by different peptides common to AMELX and AMELY, and by peptides specific to the species Q99217 and Q99217-3 of amelogenin X (AMELX), and specific to the species Q99218 and Q99218-1 of amelogenin Y (AMELY). All these peptides are located in the N-terminal tyrosine-rich amelogenin protein (TRAP) domain (Figure 5A), a proteolytic peptide released during enamel maturation [13]. Regarding AMELX, the protein was identified in all samples except in 1H18, which also gave poor results for other proteins (Table 1). The protein was consistently identified by the female-specific peptide TALVLTPLK, unique to the long form AMELX-Q99217-3. The other female-specific peptides common to all species of AMELX were less frequently detected in samples and corresponded to C-terminal non-specifically cleaved peptides of different lengths and carrying modifications such as deamination and/or oxidation (Figure 5B and Table S10). Regarding the AMELY protein characteristic of male individuals, it was identified in only 6 out of 9 male specimens (Figure 5B and Table S10). It is interesting to note that in archaelogical samples, AMELY was identified by the sex-specific peptide IALVLTPLK, unique to the long form AMELY-Q99218 (4/7 male individuals), while in the two

modern samples it was identified by the peptide WYQSMIRPPY common to all species of AMELY (Figure 5B). The identification of sex-specific peptides permitted to envisage sex determination by using a label free approach as reported in [15, 16]. XIC of sex-specific peptides identified in archaelogical samples (Figure 7A) show that the approach improved the detection of the female-specific peptides, in particular the deamidated WYQSMIRPPYP peptide, but it was not sensitive enough to detect AMELY-specific peptides in all the male samples (only 5/9 individuals confirmed with reproducibility). As mentioned above, the use of whole tooth instead of enamel alone may have hampered the detection of additional AMEL peptides described in other enamel proteomics studies [13, 15, 45] and may account for the inconstant identification of sex-specific peptides in samples. Also, in the case of AMELY, the protein is less expressed than AMELX [20]. In order to improve the detection of sex-specific amelogenin peptides in all samples, a targeted MS analysis, which provides high selectivity and sensitivity with confident targeted peptide sequence confirmation, was set up based on the identification of the two AMELX specific peptides TALVLTPLK and WYQSIRPPYP, and the peptide IALVLTPLK specific of AMELY.

3.2.Targeted MS analysis of amelogenin peptides for sex estimation

Parallel reaction monitoring (PRM) assay [31] was performed based on the analysis of sex-specific peptides and their corresponding heavy isotope-labelled synthetic peptides. The sex-specific peptides TALVLTPLK (AMELX-TALVLTPLK) and IALVLTPLK (AMELY-IALVLTPLK) were considered as well-adapted candidates for PRM. Indeed, they were identified in most samples, always as tryptic peptides without missed cleavage, and were rarely found modified (Figure 5B and Table S10). Moreover, the HCD MS/MS spectra (Figure 6A and 6B) of the AMELX-TALVLTPLK and the AMELY-IALVLTPLK peptides displayed similar series of y-ions, but the N-terminal sex-specific single amino-acid variation between peptides was clearly distinguishable using the series of b-ions, allowing an accurate identification of AMELX-TALVLTPLK and AMELY-IALVLTPLK peptides. The peptides AMELX-WYQSIRPPYP and AMELY-WYMQSIRPPY described in other studies [15-17] were always found deamidated, and oxidized on methionine residue in the case of the Y-peptide (Figure 5B and Table S10). As shown in Figure 6C and 6D, the HCD MS/MS spectra of these peptides exhibited a poor spectral quality and also less discriminative ion series, especially AMELY-WYMQSIRPPY which is deamidated and oxidized.

Therefore, the peptides AMELX-TALVLTPLK and AMELY-IALVLTPLK, as well as the deamidated peptide AMELX-WYqSIRPPYP and its non-deamidated counterpart, were selected for designing a targeted proteomics PRM assay. For each peptide, the transitions were chosen using as much as possible the most discriminative and intense fragments (Table S4) and the signal response curves were performed to assess the linearity and LOD determination (Figure S1-S4). PRM measurements were retrieved from the AMELX peptides

TALVLTPLK and WYqSIRPPYP, and AMELY peptide IALVLTPLK in all samples (Table S6) and compared to those of the label free analysis (Figure 7A). The PRM assay improved the detection of the three sex-specific peptides in all the samples, in particular the Y-specific peptide that is detected in all male individuals.

Although it is usually recommended to use several proteotypic peptides per protein to obtain more reliable and precise quantification results, a single pair of peptides (AMELX-TALVLTPLK and AMELY-IALVLTPLK) with acceptable quality standards was nevertheless finally selected for PRM-based sex-estimation, based on the comparative analysis of the CVs of the peak area intensities across triplicates (Table S6). An example of representative PRM traces using the selected transitions derived from endogenous and heavy isotope-labelled AMELX-TALVLTPLK and AMELY-IALVLTPLK peptides is given for the male individual 1H07 and the female individual 1H06 (Figure 7B). The retention time and peptide fragment patterns of the endogenous AMELX-TALVLTPLK and AMELY-IALVLTPLK peptides well matched with those of their corresponding heavy isotope-labelled peptides. The Figure 7C shows that the peak area intensities of the AMELX-TALVLTPLK and AMELY-IALVLTPLK heavy peptides were roughly similar giving a mean X/Y ratio of 0.99 ± 0.88 (SD, n=13). As the molar calibration given by the manufacturer for both heavy isotope-labelled peptides coated on READYBEADS^{TM} is X/Y = 0.83, it can be considered that the two peptides respond similarly to MS, allowing a direct comparison of their intensities. In the case of the endogenous peptides in samples, the ratio between AMELX-TALVLTPLK and AMELY-IALVLTPLK was 91 ± 53 (SD, n=9) in the male samples, indicating that the Y-specific peptide was around 100 fold less abundant than the X-specific peptide. This is lower than the estimation of 10 % previously obtained from transcriptomic studies [20], and of 8 ± 6% reported in a recent proteomic analysis using a combination of amelogenin peptides for quantitation [15]. This discrepancy may be explained by the use in our PRM assay of a pair of peptides unique to the long forms of AMELX and AMELY that may be differently expressed than the other species of amelogenin splice variants. However, because they are unique to the long forms, the X/Y peptide ratio likely reflects the X/Y gene expression ratio for this variant.

The AMELX-TALVLTPLK peptide was detected in all the female and male samples, as expected since derived from the amelogenin encoded by chromosome X. The AMELY-IALVLTPLK peptide was successfully detected in all the male individuals and not in the female individuals (Figure 7A), indicating a specific detection in males and the possibility to confidently assign the sex of males based on the detection of the peptide. In contrast, the absence of the Y-specific peptide is not sufficient to predict the sex of female individuals, as it could be due to a lower detection of the Y-specific peptide. This constitutes a limitation of our PRM assay based on a single pair of peptides. However, considering the sum of the LOD values for the AMELY-IALVLTPLK peptide ($1.12x10^5$, Table S5) and a 91 ± 53 less fold abundance compared to the AMELX-TALVLTPLK peptide, it can be assumed that for X-specific peptide

values below $5.58x10^6$ (lower 95% limit), the sex of an unknown sample cannot be assigned when the AMELY-IALVLTPLK peptide is not detected. Conversely, for AMELX-TALVLTPLK peak area values higher than $5.58x10^6$, the absence of AMELY-IALVLTPLK may indicate a female. The targeted proteomics PRM assay thus allowed the confirmation of the sex in all the samples. This demonstrates the capacity of the assay to be applied in unfavorable conditions where the signal is diluted (whole tooth instead of enamel), which may be advantagous since archaeological remains are not always optimized samples.

4. Conclusion

The present study describes the potentiality of a shotgun MS-based proteomics approach with a dedicated bioinformatics iterative database search pipeline to deeply explore ancient proteomes. Compared to the more conventional approaches, our iterative database search method yields higher identification, especially in the no enzyme search, indicating the presence of randomly degraded proteins.This high number of identification in the non-tryptic searches was also observed in present-day teeth, suggesting that a large part of proteins fragmentation is not entirely due to diagenesis but probably also results from proteolysis of tissue at the time of death. These results indicate therefore that, when only considering the classical semi-tryptic database search mode in MS analyses, a number of peptides issued from randomly degraded proteins may be missed contributing to a loss of information. Moreover, the targeted proteomics PRM assay using the isotope-labelled AMELX-TALVLTPLK and AMELY-IALVLTPLK peptides developed in the present study was efficient to successfully confirm the sex in all samples. While the assay is limited by the use of a single couple of peptides, it could however represent a protein-based method alternative to genetic analysis for estimating sex when DNA is not exploitable, as it especially occurs in very old samples.

References

[1] Macchiarelli R, Bayle P, Bondioli L, Mazurier A, Zanolli C. From outer to inner structural morphology in dental anthropology. Integration of the third dimension in the visualization and quantitative analysis of fossil remains. . Scott GR, Irish JD, editors: Cambridge University Press; 2013 2013. 250-77 p.

[2] Smith TM. Teeth and Human Life-History Evolution. Annu Rev Anthropol. 2013;42:191-208. doi: 10.1146/annurev-anthro-092412-155550. PubMed PMID: WOS:000326694800013.

[3] Smith TM, Austin C, Green DR, Joannes-Boyau R, Bailey S, Dumitriu D, et al. Wintertime stress, nursing, and lead exposure in Neanderthal children. Sci Adv. 2018;4(10):eaau9483. doi: 10.1126/sciadv.aau9483. PubMed PMID: 30402544; PubMed Central PMCID: PMCPMC6209393.

[4] Cappellini E, Prohaska A, Racimo F, Welker F, Pedersen MW, Allentoft ME, et al. Ancient Biomolecules and Evolutionary Inference. Annu Rev Biochem. 2018;87:1029-60. doi: 10.1146/annurev-biochem-062917-012002. PubMed PMID: 29709200.

[5] Cleland TP, Schroeter ER. A Comparison of Common Mass Spectrometry Approaches for Paleoproteomics. Journal of proteome research. 2018. doi: 10.1021/acs.jproteome.7b00703. PubMed PMID: 29384680.

[6] Cappellini E, Collins MJ, Gilbert MT. Biochemistry. Unlocking ancient protein palimpsests. Science. 2014;343(6177):1320-2. Epub 2014/03/22. doi: 10.1126/science.1249274. PubMed PMID: 24653025.

[7] Welker F. Palaeoproteomics for human evolution studies. Quaternary Science Reviews. 2018;190:137-47.

[8] Jersie-Christensen RR, Lanigan LT, Lyon D, Mackie M, Belstrom D, Kelstrup CD, et al. Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. Nature communications. 2018;9(1):4744. doi: 10.1038/s41467-018-07148-3. PubMed PMID: 30459334; PubMed Central PMCID: PMCPMC6246597.

[9] Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, et al. Pathogens and host immunity in the ancient human oral cavity. Nature genetics. 2014;46(4):336-44. Epub 2014/02/25. doi: 10.1038/ng.2906. PubMed PMID: 24562188; PubMed Central PMCID: PMC3969750.

[10] Hendy J, Warinner C, Bouwman A, Collins MJ, Fiddyment S, Fischer R, et al. Proteomic evidence of dietary sources in ancient dental calculus. Proc Biol Sci. 2018;285(1883). doi: 10.1098/rspb.2018.0977. PubMed PMID: 30051838; PubMed Central PMCID: PMCPMC6083251.

[11] Jeong C, Wilkin S, Amgalantugs T, Bouwman AS, Taylor WTT, Hagan RW, et al. 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-E55. doi: 10.1073/pnas.1813608115. PubMed PMID: 30397125; PubMed Central PMCID: PMCPMC6275519.

[12] Warinner C, Hendy J, Speller C, Cappellini E, Fischer R, Trachsel C, et al. Direct evidence of milk consumption from ancient human dental calculus. Scientific reports. 2014;4:7104. Epub 2014/11/28. doi: 10.1038/srep07104. PubMed PMID: 25429530; PubMed Central PMCID: PMC4245811.

[13] Castiblanco GA, Rutishauser D, Ilag LL, Martignon S, Castellanos JE, Mejia W. Identification of proteins from human permanent erupted enamel. Eur J Oral Sci. 2015;123(6):390-5. doi: 10.1111/eos.12214. PubMed PMID: 26432388.

[14] Nielsen-Marsh CM, Stegemann C, Hoffmann R, Smith T, Feeney R, Toussaint M, et al. Extraction and sequencing of human and Neanderthal mature enamel proteins using MALDI-TOF/TOF MS. J Archaeol Sci. 2009;36(8):1758-63. doi: 10.1016/j.jas.2009.04.004. PubMed PMID: ISI:000267562900011.

[15] Parker GJ, Yip JM, Eerkens JW, Salemi M, Durbin-Johnson B, Kiesow C, et al. Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel. Journal of Archaelogical Science. 2019;101:169-80. doi: https://doi.org/10.1016/j.jas.2018.08.011.

[16] Stewart NA, Gerlach RF, Gowland RL, Gron KJ, Montgomery J. Sex determination of human remains from peptides in tooth enamel. Proc Natl Acad Sci U S A. 2017;114(52):13649-54. doi: 10.1073/pnas.1714926115. PubMed PMID: 29229823; PubMed Central PMCID: PMCPMC5748210.

[17] Wasinger VC, Curnoe D, Bustamante S, Mendoza R, Shoocongdej R, Adler L, et al. Analysis of the Preserved Amino Acid Bias in Peptide Profiles of Iron Age Teeth from a Tropical Environment Enable Sexing of Individuals Using Amelogenin MRM. Proteomics. 2019;19(5):e1800341. doi: 10.1002/pmic.201800341. PubMed PMID: 30650255.

[18] Hendy J, Welker F, Demarchi B, Speller C, Warinner C, Collins MJ. A guide to ancient protein studies. Nat Ecol Evol. 2018;2(5):791-9. doi: 10.1038/s41559-018-0510-x. PubMed PMID: 29581591.

[19] Jagr M, Eckhardt A, Pataridis S, Broukal Z, Duskova J, Miksik I. Proteomics of human teeth and saliva. Physiol Res. 2014;63 Suppl 1:S141-54. PubMed PMID: 24564654.

[20] Lacruz RS, Habelitz S, Wright JT, Paine ML. Dental Enamel Formation and Implications for Oral Health and Disease. Physiol Rev. 2017;97(3):939-93. doi: 10.1152/physrev.00030.2016. PubMed PMID: 28468833.

[21] Procopio N, Chamberlain AT, Buckley M. Exploring Biological and Geological Age-related Changes through Variations in Intra- and Intertooth Proteomes of Ancient Dentine. Journal of proteome research. 2018;17(3):1000-13. doi: 10.1021/acs.jproteome.7b00648. PubMed PMID: 29356547.

[22] Zanolli C, Hourset M, Esclassan R, Mollereau C. Neanderthal and Denisova tooth protein variants in present-day humans. PLoS One. 2017;12(9):e0183802. doi:

10.1371/journal.pone.0183802. PubMed PMID: 28902892; PubMed Central PMCID: PMCPMC5597096.

[23] Daubert DM, Kelley JL, Udod YG, Habor C, Kleist CG, Furman IK, et al. Human enamel thickness and ENAM polymorphism. Int J Oral Sci. 2016;8(2):93-7. doi: 10.1038/ijos.2016.1. PubMed PMID: 27357321; PubMed Central PMCID: PMCPMC4932773.

[24] Welker F, Hajdinjak M, Talamo S, Jaouen K, Dannemann M, David F, et al. Palaeoproteomic evidence identifies archaic hominins associated with the Chatelperronian at the Grotte du Renne. Proc Natl Acad Sci U S A. 2016. Epub 2016/09/18. doi: 10.1073/pnas.1605834113. PubMed PMID: 27638212.

[25] Chen F, Welker F, Shen CC, Bailey SE, Bergmann I, Davis S, et al. A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. Nature. 2019. doi: 10.1038/s41586-019-1139-x. PubMed PMID: 31043746.

[26] Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RA, Stafford TW, et al. Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. Journal of proteome research. 2012;11(2):917-26. Epub 2011/11/23. doi: 10.1021/pr200721u. PubMed PMID: 22103443.

[27] Hendy J, Colonese AC, Franz I, Fernandes R, Fischer R, Orton D, et al. Ancient proteins from ceramic vessels at Catalhoyuk West reveal the hidden cuisine of early farmers. Nature communications. 2018;9(1):4064. doi: 10.1038/s41467-018-06335-6. PubMed PMID: 30283003.

[28] Craig R, Beavis RC. A method for reducing the time required to match protein sequences with tandem mass spectra. Rapid communications in mass spectrometry : RCM. 2003;17(20):2310-6. doi: 10.1002/rcm.1198. PubMed PMID: 14558131.

[29] Huang X, Huang L, Peng H, Guru A, Xue W, Hong SY, et al. ISPTM: an iterative search algorithm for systematic identification of post-translational modifications from complex proteome mixtures. Journal of proteome research. 2013;12(9):3831-42. doi: 10.1021/pr4003883. PubMed PMID: 23919725; PubMed Central PMCID: PMCPMC3786209.

[30] Wang H, Tang HY, Tan GC, Speicher DW. Data analysis strategy for maximizing high-confidence protein identifications in complex proteomes such as human tumor secretomes and human serum. Journal of proteome research. 2011;10(11):4993-5005. doi: 10.1021/pr200464c. PubMed PMID: 21955121; PubMed Central PMCID: PMCPMC3221390.

[31] Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol Cell Proteomics. 2012;11(11):1475-88. doi: 10.1074/mcp.O112.020131. PubMed PMID: 22865924; PubMed Central PMCID: PMCPMC3494192.

[32] Donat R, Sohn M, Langry-François F, Polloni A, Maingaud A, Mazière G, et al. Le mobilier de l'hypogée II du Mont Aimé au Val-des-Marais (Marne) dans son cadre régional :

nouvelles données. Revue archeologique de l'Est et d'Ile de France. 2014;RAE,34 RAIF,1:389-410.

[33] Theves C, Cabot E, Bouakaze C, Chevet P, Crubezy E, Balaresque P. About 42% of 154 remains from the "Battle of Le Mans", France (1793) belong to women and children: Morphological and genetic evidence. Forensic Sci Int. 2016;262:30-6. Epub 2016/03/12. doi: 10.1016/j.forsciint.2016.02.029. PubMed PMID: 26968017.

[34] Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004;20(9):1466-7. doi: 10.1093/bioinformatics/bth092. PubMed PMID: WOS:000222125600019.

[35] Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. Proteomics. 2011;11(5):996-9. doi: 10.1002/pmic.201000595. PubMed PMID: 21337703.

[36] Kall L, Storey JD, MacCoss MJ, Noble WS. Posterior error probabilities and false discovery rates: two sides of the same coin. Journal of proteome research. 2008;7(1):40-4. doi: 10.1021/pr700739d. PubMed PMID: 18052118.

[37] Sinitcyn P, Rudolph JD, Cox J. Computational methods for understanding mass spectrometry-based shotgun proteomic data. Annula Review of Biomedical Data Science. 2018;1:28. doi: https://doi.org/10.1146/annurev-biodatasci-080917-013516.

[38] Elias JE, Gygi SP. Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol Biol. 2010;604:55-71. doi: 10.1007/978-1-60761-444-9_5. PubMed PMID: 20013364; PubMed Central PMCID: PMCPMC2922680.

[39] Vaudel M, Burkhart JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, et al. PeptideShaker enables reanalysis of MS-derived proteomics data sets. Nat Biotechnol. 2015;33(1):22-4. doi: 10.1038/nbt.3109. PubMed PMID: 25574629.

[40] MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966-8. doi: 10.1093/bioinformatics/btq054. PubMed PMID: 20147306; PubMed Central PMCID: PMCPMC2844992.

[41] Mani DR, Abbatiello SE, Carr SA. Statistical characterization of multiple-reaction monitoring mass spectrometry (MRM-MS) assays for quantitative proteomics. BMC Bioinformatics. 2012;13 Suppl 16:S9. doi: 10.1186/1471-2105-13-S16-S9. PubMed PMID: 23176545; PubMed Central PMCID: PMCPMC3489552.

[42] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57. doi: 10.1038/nprot.2008.211. PubMed PMID: 19131956.

[43] Jagr M, Eckhardt A, Pataridis S, Miksik I. Comprehensive proteomic analysis of human dentin. Eur J Oral Sci. 2012;120(4):259-68. doi: 10.1111/j.1600-0722.2012.00977.x. PubMed PMID: 22813215.

[44] Park ES, Cho HS, Kwon TG, Jang SN, Lee SH, An CH, et al. Proteomics analysis of human dentin reveals distinct protein expression profiles. Journal of proteome research. 2009;8(3):1338-46. doi: 10.1021/pr801065s. PubMed PMID: 19193101.

[45] Stewart NA, Molina GF, Issa JPM, Yates NA, Sosovicka M, Vieira AR, et al. The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction. Rsc Adv. 2016;6(66):61673-9. doi: 10.1039/c6ra05120k. PubMed PMID: WOS:000379485200081.

Figure legends:

Figure 1: Iterative database search strategy used for the bioinformatics data analysis of shotgun MS-based proteomics data.

Figure 2: Deamidation of glutamine (Q) and asparagine (N) residues in Neolithic and present-day samples.
For each sample in the two injection replicates R1 and R2, the number of deamidated Q and/or N residues was counted in the list of PSM and reported to the sum of all Q and/or N residues in the same PSM list. Bars represent mean ± SEM of the ratios for each sample and replicate. * Statistical significance determined using the Holm-Sidak method, with alpha = 0.05 in a multiple unpaired t-test (GraphPad Prism v7, La Jolla CA).

Figure 3: Number of proteins identified using the optimized three steps data analysis workflow based on iterative database searches.
Data are represented as mean ± SEM of the two injection replicates per sample.
(A) Total number of proteins per sample. The number of identified proteins after the first step (usual semi-tryptic database search) is displayed in black. The number of proteins exclusively identified in steps 2 and 3 (no enzyme database search) is displayed in white.
(B) Percentage distribution of the number of proteins identified after each step (step 1 in black, step 2 in white, step 3 in grey) of the iterative database search workflow.

Figure 4: Classification of proteins into GO terms annotations.
The lists of proteins from Neolithic (A, B) and present-day (C, D) samples either identified in the semi-tryptic search mode or exclusively identified in the no enzyme search modes, were classified using the functional annotation chart tool of DAVID (https://david.ncifcrf.gov/home.jsp, [42]).

Figure 5: Amelogenin peptides.
(A) Alignement of the N-terminal part of AMELX and AMELY proteins, showing the TRAP peptide. Amino acid variation between X and Y amelogenins are indicated by white characters highlighted in black, boxes indicate the tryptic peptides used in the targeted MS analysis, bold characters indicate potential trypsin cleavage. (B) List of AMELX and AMELY peptides identified with high confidence in Neolithic and Present-day samples by the iterative database search workflow; including their modifications. In italic: non-specific peptides, in bold: sex-specific peptides used in the targeted MS. A more detailed table is given in supplementary data (Table S9).

Figure 6: HCD MS/MS spectra of the sex-specific peptides used in the targeted MS (A) AMELX peptide, TALVLTPLK (doubly charged precursor ion, MH2+, at m/z 478.3130). (B) AMELY peptide, IALVLTPLK (doubly charged precursor ion, MH2+, at m/z 484.3325). (C) AMELX peptide, WYqSIRPPYP (doubly charged precursor ion, MH2+, at m/z 654.3259) and (D) AMELY peptide, WYqSmIRPPY (doubly charged precursor ion, MH2+, at m/z 679.3165). Series of y- and b-ions are highlighted in blue and red, respectively. q: deamidated glutamine residue; m: oxidized methionine residue.

Figure 7: Label free and PRM-based targeted analyses for detection of sex-specific peptides. (A) Label free and PRM measurements of the peak areas with the standard deviations (SD) for the female-specific peptides TALVLTPLK and the male-specific peptide IALVLTPLK in all the samples. Peak areas correspond to the mean of the summed peak area intensities across the triplicate and the XIC integration across the two replicates for the PRM-based targeted and the label free analyses, respectively. (B) Typical examples for transition chromatograms of AMELX-TALVLTPLK and AMELY-IALVLTPLK peptides from male individual 1H07 and female individual 1H06 using PRM mode with corresponding RT and fragment patterns of the heavy counterparts. (C) Plot of the peak area intensities (converted to Log values) of the endogenous and heavy AMELX-TALVLTPLK and AMELY-IALVLTPLK peptides in all samples. An arbitrary value of 1 was attributed when no peak was detected, as it is the case for the Y peptide in the 4 female individuals. The genetic sex of each individual is represented by the corresponding symbols. The synthetic heavy peptides spiked in all samples are represented by squares. The dotted lines illustrate the experimental X/Y ratio 0.99 ± 0.88 (SD, n=13) for the standard peptides. The two ticks on the X-axis indicate the 95 % confidence interval surrounding the limit of detection value for AMELX-TALVLTPLK enabling the identification of a female individual in the absence of detection of the AMELY-IALVLTPLK. For AMELX-TALVLTPLK values below the lower limit, the sex of an unknown sample cannot be predicted.

Table 1: Proteins characteristic of the tooth hard tissues identified per sample by using Proteome Discoverer™ (PD) or X!Tandem (X!T) softwares.

Accession	Gene Name		# Peptides (PSM)												
			Males									Females			
			1H12	1H01	1H07	2H08	2H10	1H05	2H12	CW02	CW03	1H06	1H15	1H18	2H23
Q9NP70	AMBN	PD	1(3)	1(7)	1(2)	1(7)	1(4)	3(11)	2(6)	1(14)	1(16)	1(6)	1(4)		2(8)
		X!T	1(2)	2(5)					3(5)	1(10)	4(8)				1(5)
Q99217	AMELX	PD		3(7)		2(16)		2(4)		2(49)	3(4)	3(6)			
		X!T		2(14)		2(17)		2(12)			4(7)				1(19)
Q99217-3	Iso3 AMELX	PD	1(11)	3(13)	2(4)	2(22)	2(12)	2(13)	1(11)	2(126)	3(58)	2(8)	1(11)		2(21)
		X!T	1(11)			2(36)	2(12)	2(31)	2(30)	2(136)	5(54)	2(12)	1(18)		
Q99218	AMELY	PD		1(2)		1(2)		1(1)	1(2)						
		X!T		1(1)		1(2)									
Q99218-1	Iso1 AMELY	PD								2(14)	3(5)				
		X!T													
Q6UX39	AMTN	PD						1(1)		1(1)					1(1)
		X!T													
Q6PRD7	CEMP1	PD	1(1)							1(1)	1(1)			1(1)	
		X!T													
Q13316	DMP1	PD													
		X!T									2(2)				
Q9NZW4	DSSP	PD		3(7)	5(9)	3(6)	1(1)	1(1)	2(2)	6(26)	11(44)	2(2)	2(2)		4(8)
		X!T		4(8)					1(2)	14(36)	20(38)		4(7)		
Q9NRM1	ENAM	PD		4(4)		1(1)		3(3)	1(1)	4(11)	3(8)	1(2)			
		X!T													
P08493	MGP	PD													
		X!T								4(6)	8(13)				
P08493-2	Iso2 MGP	PD	3(11)	9(61)	7(30)	4(11)	8(63)	2(12)	11(55)	6(85)	8(120)	5(18)	2(9)	4(18)	2(7)
		X!T													
O60882	MMP-20	PD	2(7)	13(28)	11(20)	9(26)	5(7)	8(12)	7(15)	11(34)	8(15)	7(12)	3(5)	2(9)	12(26)
		X!T	2(33)	16(40)	10(28)	8(35)	6(15)	7(23)	10(19)	8(29)	10(30)	4(15)	4(13)	2(16)	8(21)

Figure 1:

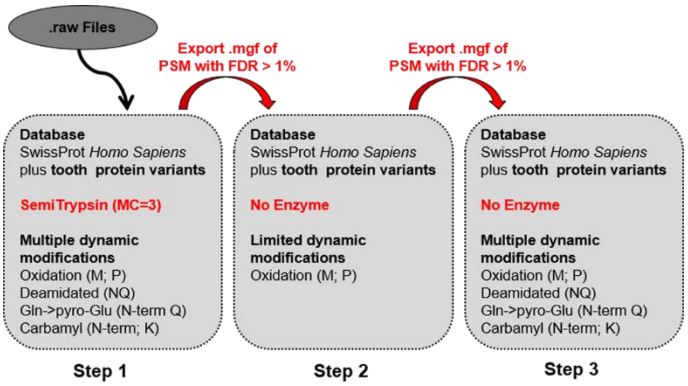


Figure 2:

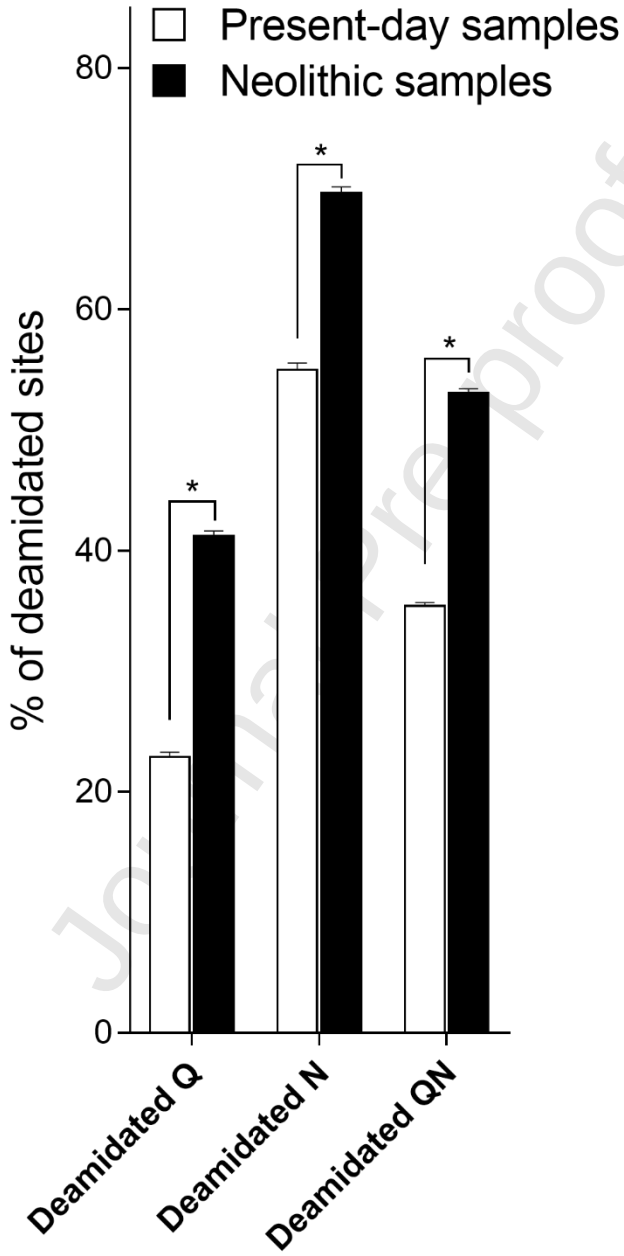


Figure 3:

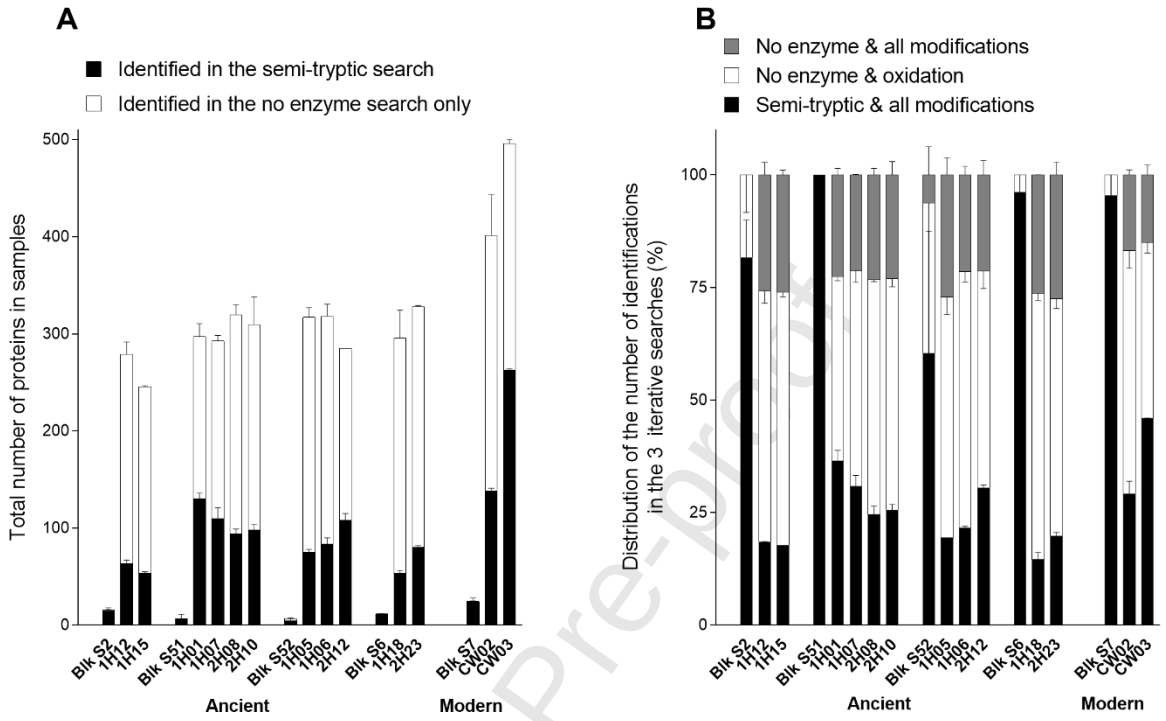


Figure 4:

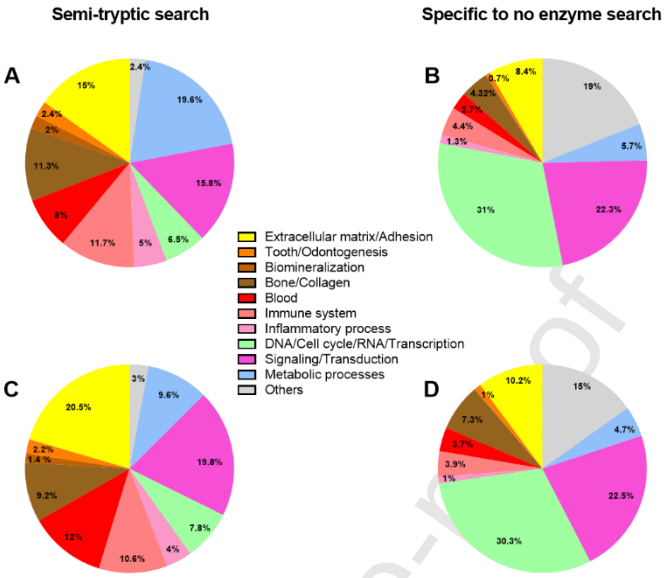
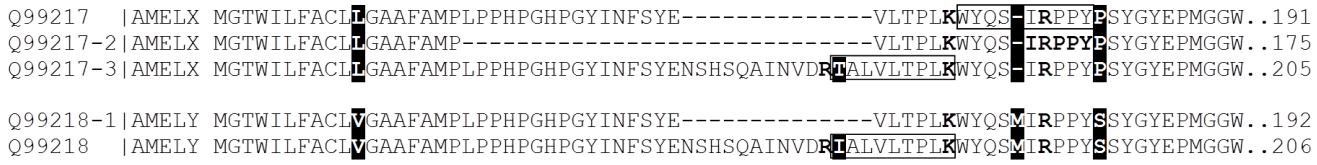


Figure 5:

A



B

						# PSM																					
						Males																Females					
						1H12		1H01		1H07		2H08		2H10		1H05		2H12		CW02		CW03		1H06		1H15		1H18		2H23	
Description	Accession	Peptides	Positions	Modifications	Theo. MH+ [Da]	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
AMELX	**Q99217**	*[N].FSYEVLTPLK.[W]*	[31-40]		1196.6562			1	1			1	1			1		1				1		1	1						
		[F].SYEVLTPLK.[W]	[32-40]		1049.5877															1		1									
		[S].YEVLTPLK.[W]	[33-40]		962.5557	1										1								1							
		[K].WYQSIRPPYP.[S]	[41-50]	1xDeam [Q3]	1307.6419			5	1	1		15		4		3		4		48		2		4						5	2
		[K].WYQSIRPPYPSYGYEPMG.[G]	[41-58]	1xOx [M17]; 1xDeam [Q3]	2207.9743			1																							
AMELX (isoform 3)	**Q99217-3**	**[R].TALVLTPLK.[W]**	[46-54]		955.6186	**11**	**11**	**7**	**5**	**9**	**3**	**7**	**20**	**10**	**8**	**10**	**18**	**11**	**9**	**21**	**22**	**21**	**18**	**4**	**11**	**11**	**1**			**16**	**15**
		[R].TALVLTPLK.[W]	[46-54]	1×Carbamyl [N-Term]	998.6245																	**1**									
		[K].WYQSIRPPYP.[S]	[55-64]	1xDeam [Q3]	1307.6419			**5**	**1**	**1**		**15**		**4**		**3**		**4**		**48**	**35**	**15**	**2**	**4**						**5**	**2**
		[K].WYQSIRPPYPS.[Y]	[55-65]	1×Deam [Q3]	1394.6739																	1									
		[K].WYQSIRPPYPSYGYEPMG.[G]	[55-72]	1xOx [M17]; 1xDeam [Q3]	2207.9743			1																							
AMELY	**Q99218**	**[R].IALVLTPLK.[W]**	[46-54]		967.6550			**2**	**2**			**2**	**2**			**1**		**2**													
AMELY (isoform 1)	**Q99218-1**	*[N].FSYEVLTPLK.[W]*	[31-40]		1196.6562																										
		[F].SYEVLTPLK.[W]	[32-40]		1049.5877															12		2									
		[S].YEVLTPLK.[W]	[33-40]		962.5557																										
		[K].WYQSMIRPPY.[S]	[41-50]	1×Ox [M5]; 1×Deam [Q3]	1357.62454															1		1									

Figure 6 :

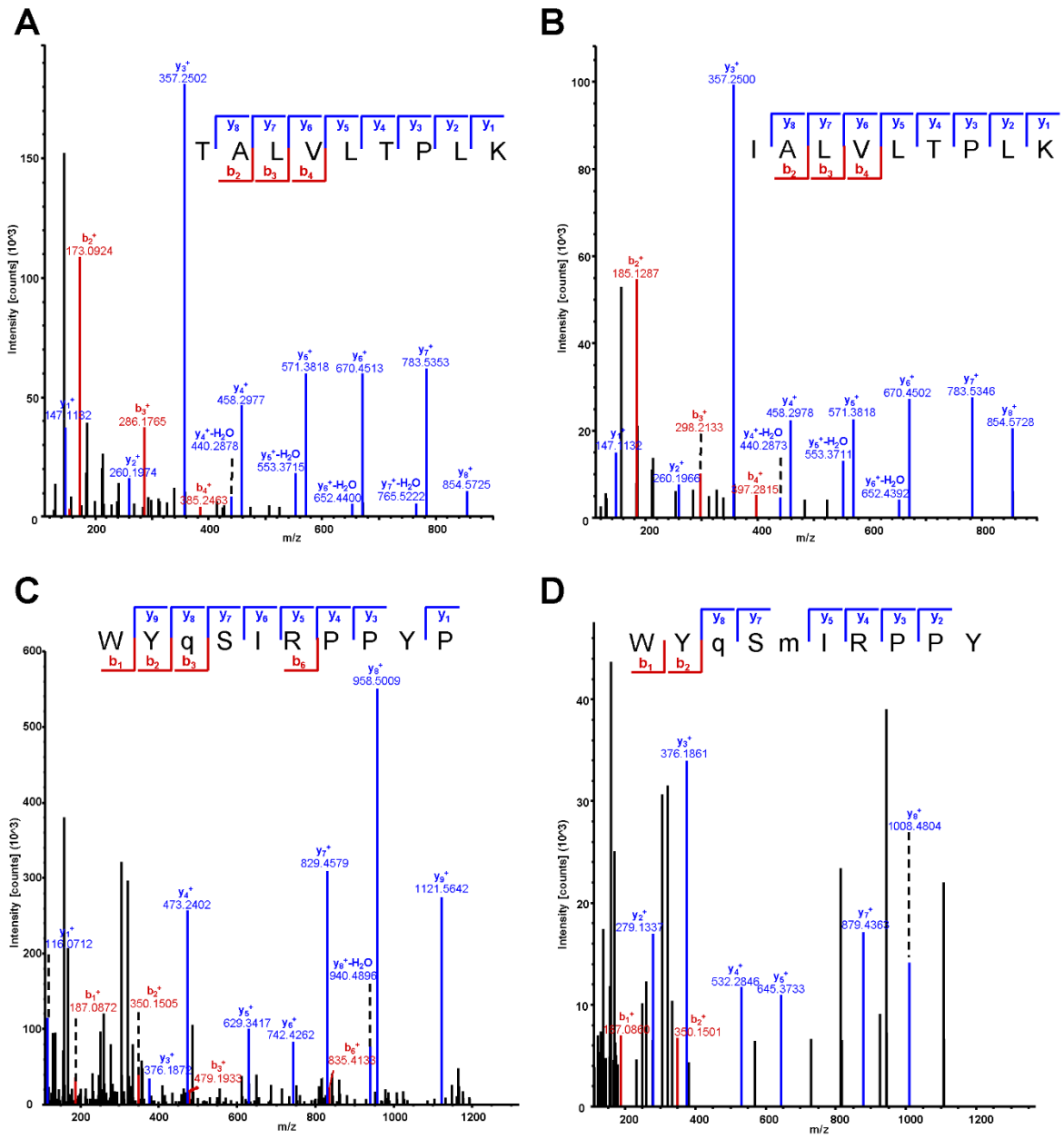
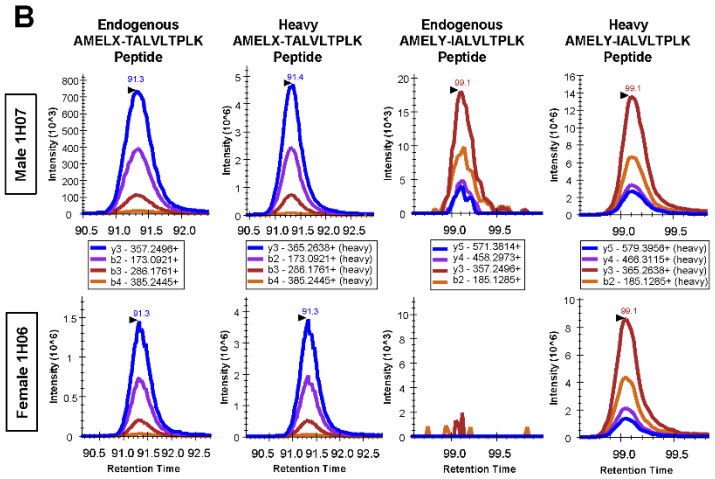

A
T A L V L T P L K
Intensity [counts] (10^3)
m/z
B
I A L V L T P L K
Intensity [counts] (10^3)
m/z
C
W Y q S I R P P Y P
Intensity [counts] (10^3)
m/z
D
W Y q S m I R P P Y
Intensity [counts] (10^3)
m/z

Figure 7:

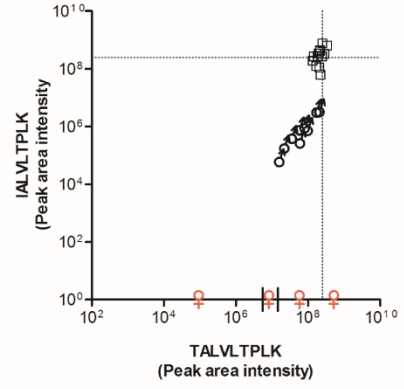
A

Sample	Sex	Peak area measurements ± SD					
		TALVLTPLK		WYqSIRPPYP		IALVLTPLK	
		Label free	PRM	Label free	PRM	Label free	PRM
1H01	M	5.45E+08 ± 3.61E+08	2.06E+08 ± 1.25E+07	1.14E+08 ± 2.39E+07	1.33E+06 ± 1.71E+05	6.74E+06 ± 4.81E+06	4.47E+06 ± 2.11E+05
1H05	M	5.74E+08 ± 1.34E+08	9.72E+07 ± 3.32E+07	4.28E+07 ± 1.93E+07	1.52E+05 ± 9.40E+04	2.89E+06 ± 1.73E+06	9.93E+05 ± 3.47E+05
1H07	M	8.40E+07 ± 6.60E+07	3.54E+07 ± 2.42E+06	6.48E+07 ± 2.13E+07	5.67E+05 ± 3.91E+04		5.31E+05 ± 4.23E+04
1H12	M	3.43E+07 ± 7.45E+06	1.59E+07 ± 1.20E+06	2.86E+06 ± 4.87E+05	6.15E+04 ± 1.47E+04		8.27E+04 ± 2.93E+03
2H08	M	3.94E+08 ± 2.13E+08	1.72E+08 ± 3.05E+07	1.53E+08 ± 6.61E+07	1.72E+08 ± 3.05E+07	4.73E+06 ± 3.04E+06	4.28E+06 ± 5.70E+05
2H10	M	4.50E+07 ± 2.57E+07	2.18E+07 ± 1.05E+06	3.07E+07 ± 8.79E+06	2.18E+07 ± 1.05E+06		2.43E+05 ± 2.21E+04
2H12	M	2.60E+08 ± 1.37E+08	8.00E+07 ± 2.16E+07	3.92E+07 ± 9.03E+06	8.00E+07 ± 2.16E+07	2.41E+06 ± 2.67E+06	1.22E+06 ± 3.29E+05
CW02	M	2.12E+08 ± 1.09E+07	5.96E+07 ± 2.51E+06	1.67E+08 ± 7.59E+06	5.96E+07 ± 2.51E+06	3.74E+05	3.63E+05 ± 1.79E+05
CW03	M	2.23E+08 ± 8.28E+06	5.83E+07 ± 1.00E+07	1.89E+08 ± 2.67E+06	5.83E+07 ± 1.00E+07	1.25E+06 ± 2.55E+05	1.05E+06 ± 2.87E+05
1H06	F	2.13E+08 ± 3.97E+07	5.77E+07 ± 7.87E+06	6.11E+07 ± 2.77E+06	5.77E+07 ± 7.87E+06		
1H15	F	1.49E+07 ± 3.83E+05	8.23E+06 ± 3.96E+05		8.23E+06 ± 3.96E+05		
1H18	F		9.26E+04 ± 1.30E+04	6.71E+06 ± 2.93E+06	9.26E+04 ± 1.30E+04		
2H23	F	1.63E+09 ± 2.53E+08	5.19E+08 ± 3.34E+07	8.57E+07 ± 7.72E+07	5.11E+05 ± 4.48E+04		

B



C



Conflict of interest

The authors declare that they have no conflict of interest.

Significance

The study demonstrates the high potential of MS-based proteomics coupled to an iterative database search strategy for the in-depth investigation of ancient proteomes. An efficient targeted PRM MS-based approach, although limited to the detection of a single pair of sex-specific amelogenin peptides, allowed confirming the sex of individuals in ancient dental remains, an essential information for paleoanthropologists facing the issue of sex determination and dimorphism.

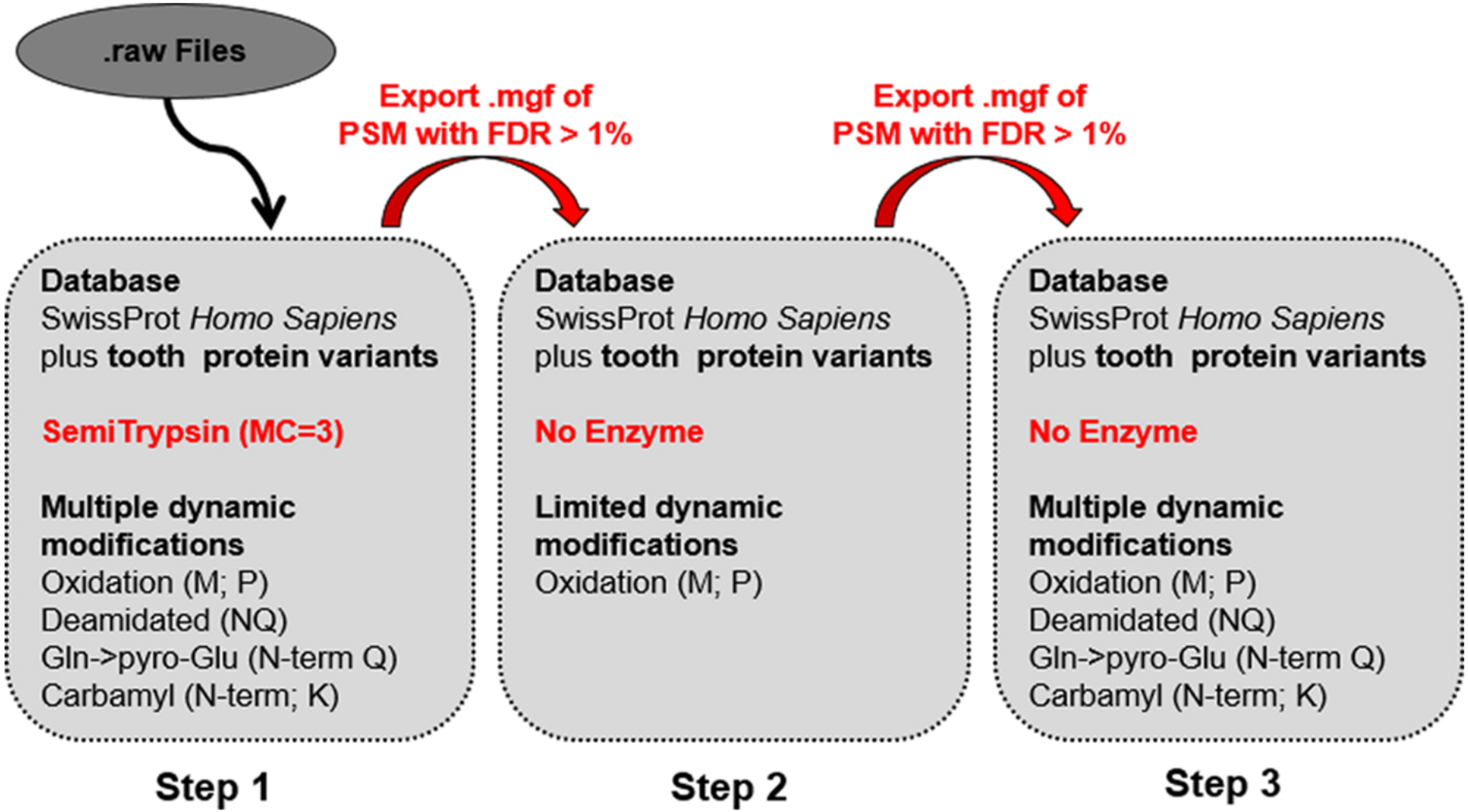

.raw Files
Export .mgf of
PSM with FDR > 1%
Export .mgf of
PSM with FDR > 1%
Database
SwissProt Homo Sapiens
plus tooth protein variants
SemiTrypsin (MC=3)
Multiple dynamic
modifications
Oxidation (M; P)
Deamidated (NQ)
Gln->pyro-Glu (N-term Q)
Carbamyl (N-term; K)
Step 1
Database
SwissProt Homo Sapiens
plus tooth protein variants
No Enzyme
Limited dynamic
modifications
Oxidation (M; P)
Step 2
Database
SwissProt Homo Sapiens
plus tooth protein variants
No Enzyme
Multiple dynamic
modifications
Oxidation (M; P)
Deamidated (NQ)
Gln->pyro-Glu (N-term Q)
Carbamyl (N-term; K)
Step 3

Figure 1

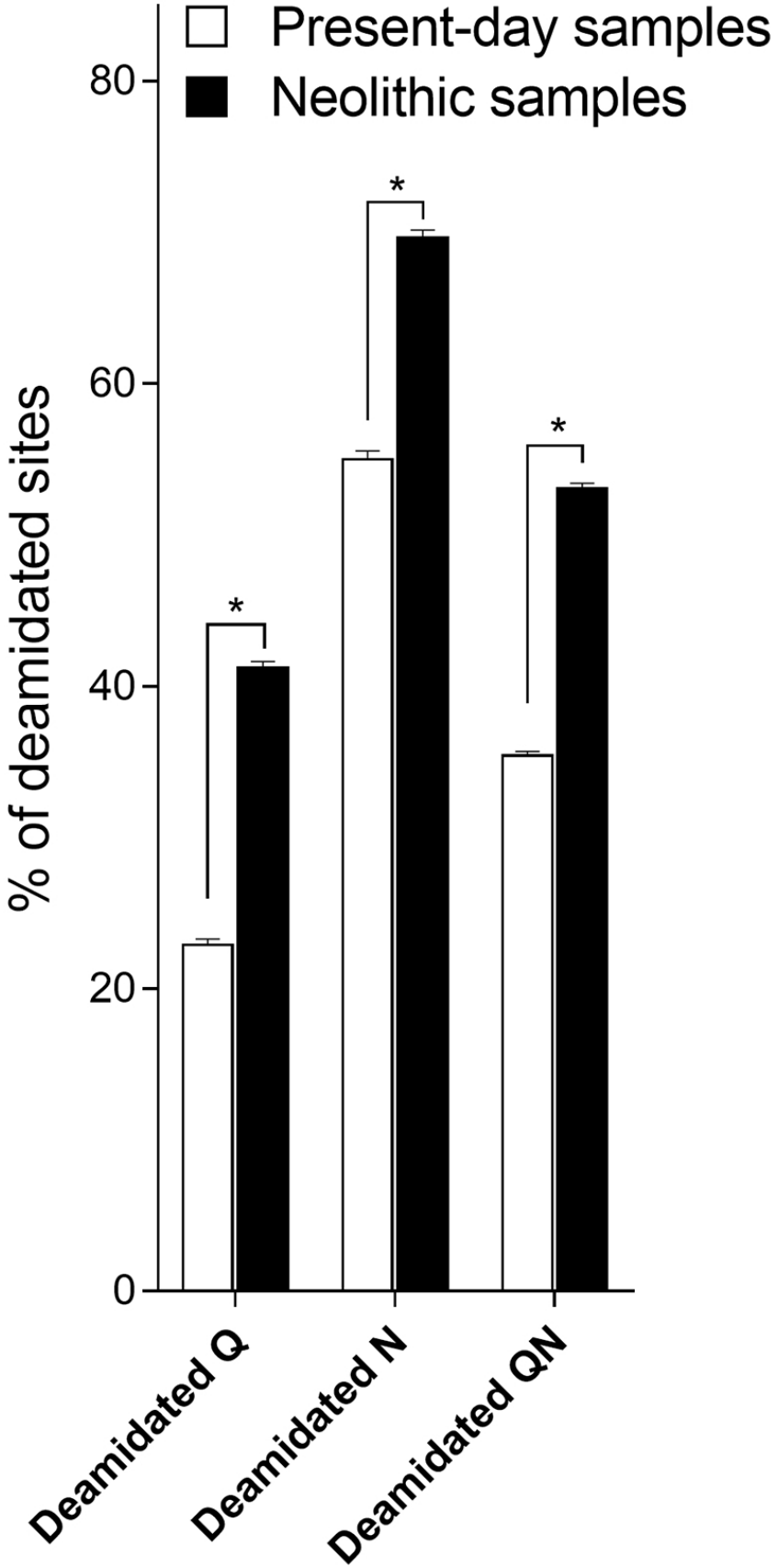

Present-day samples
Neolithic samples
% of deamidated sites
80
60
40
20
0
*
*
*
Deamidated Q
Deamidated N
Deamidated QN

Figure 2

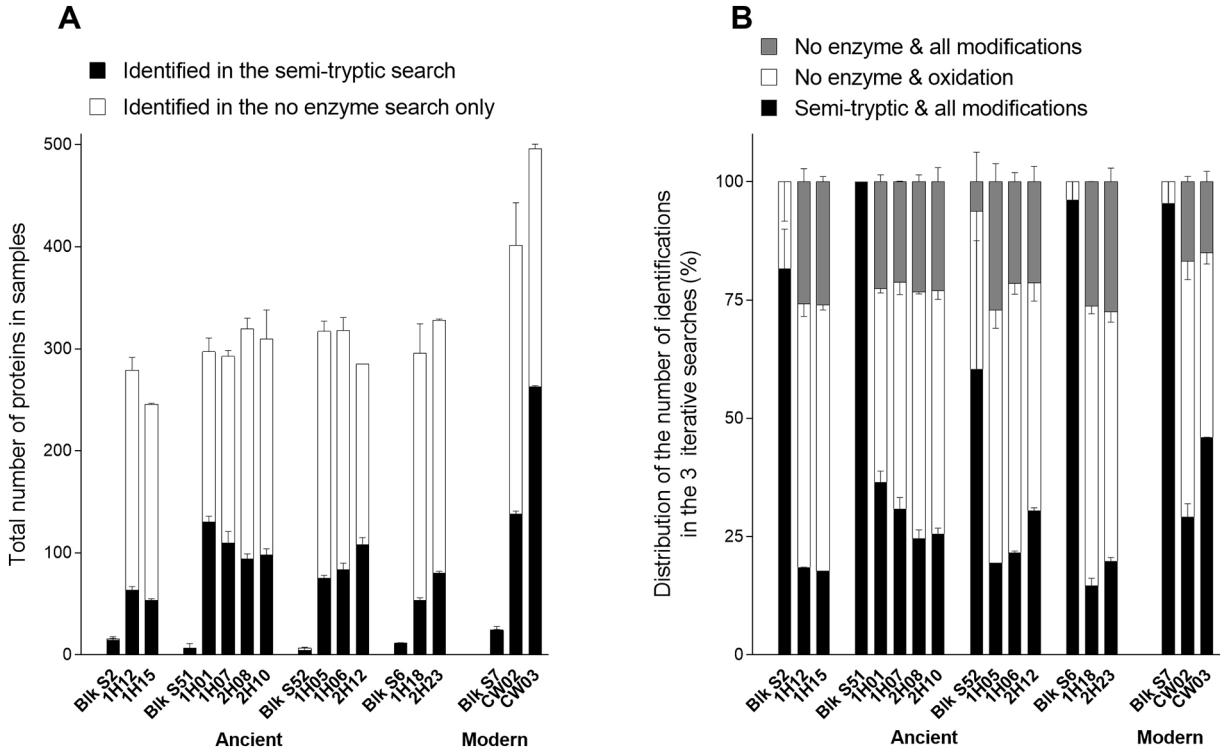

A
Identified in the semi-tryptic search
Identified in the no enzyme search only
Total number of proteins in samples
500
400
300
200
100
0
Blk S2
1H12
1H15
Blk S51
1H01
1H07
2H08
2H10
Blk S52
1H05
1H06
2H12
Blk S6
1H18
2H23
Blk S7
CW02
CW03
Ancient
Modern
B
No enzyme & all modifications
No enzyme & oxidation
Semi-tryptic & all modifications
Distribution of the number of identifications in the 3 iterative searches (%)
100
75
50
25
0
Blk S2
1H12
1H15
Blk S51
1H01
1H07
2H08
2H10
Blk S52
1H05
1H06
2H12
Blk S6
1H18
2H23
Blk S7
CW02
CW03
Ancient
Modern

Figure 3

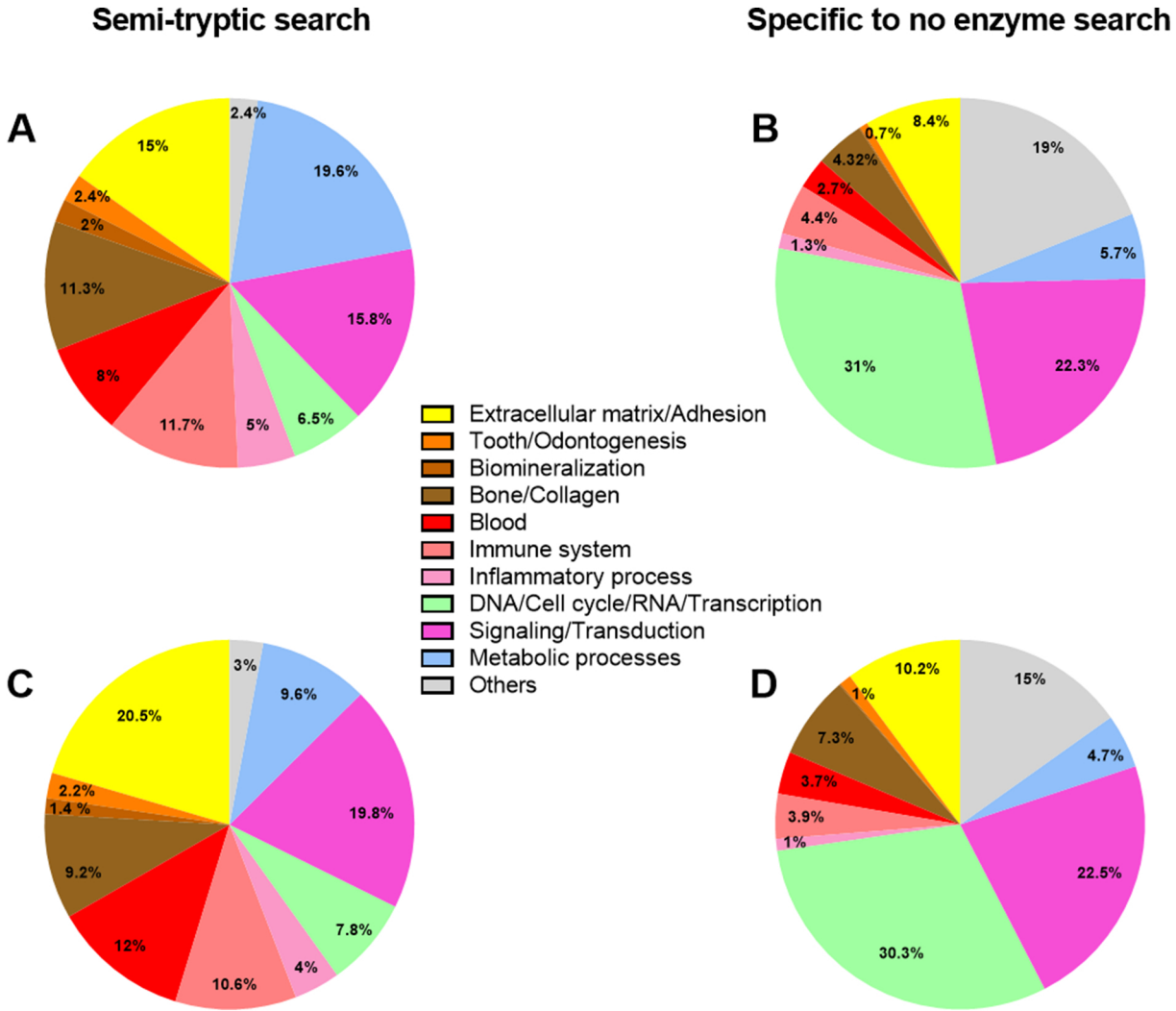

Semi-tryptic search
Specific to no enzyme search
A
B
C
D
2.4%
19.6%
15%
2.4%
2%
11.3%
15.8%
8%
11.7%
5%
6.5%
8.4%
0.7%
4.32%
2.7%
4.4%
1.3%
19%
5.7%
31%
22.3%
3%
9.6%
20.5%
2.2%
1.4 %
19.8%
9.2%
7.8%
12%
4%
10.6%
10.2%
1%
7.3%
3.7%
3.9%
1%
15%
4.7%
22.5%
30.3%
Extracellular matrix/Adhesion
Tooth/Odontogenesis
Biomineralization
Bone/Collagen
Blood
Immune system
Inflammatory process
DNA/Cell cycle/RNA/Transcription
Signaling/Transduction
Metabolic processes
Others

Figure 4

A

```
Q99217   |AMELX MGTWILFACLLGAAFAMPLPPHPGHPGYINFSYE--------------VLTPLKWYQS-IRPPYPSYGYEPMGGW..191
Q99217-2|AMELX MGTWILFACLLGAAFAMP-----------------------------VLTPLKWYQS-IRPPYPSYGYEPMGGW..175
Q99217-3|AMELX MGTWILFACLLGAAFAMPLPPHPGHPGYINFSYENSHSQAINVDRTALVLTPLKWYQS-IRPPYPSYGYEPMGGW..205

Q99218-1|AMELY MGTWILFACLVGAAFAMPLPPHPGHPGYINFSYE--------------VLTPLKWYQSMIRPPYSSYGYEPMGGW..192
Q99218   |AMELY MGTWILFACLVGAAFAMPLPPHPGHPGYINFSYENSHSQAINVDRIALVLTPLKWYQSMIRPPYSSYGYEPMGGW..206
```

B

						# PSM																					
						Males																Females					
						1H12		1H01		1H07		2H08		2H10		1H05		2H12		CW02		CW03		1H06		1H15		1H18		2H23	
Description	Accession	Peptides	Positions	Modifications	Theo. MH+ [Da]	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
AMELX	**Q99217**	*[N].FSYEVLTPLK.[W]*	[31-40]		1196.6562			1	1			1	1			1		1				1		1	1						
		[F].SYEVLTPLK.[W]	[32-40]		1049.5877															1		1									
		[S].YEVLTPLK.[W]	[33-40]		962.5557	1										1								1							
		[K].WYQSIRPPYP.[S]	[41-50]	1xDeam [Q3]	1307.6419			5	1	1		15		4		3		4		48		2		4						5	2
		[K].WYQSIRPPYPSYGYEPMG.[G]	[41-58]	1xOx [M17]; 1xDeam [Q3]	2207.9743			1																							
AMELX	**Q99217-3**	**[R].TALVLTPLK.[W]**	[46-54]		955.6186	**11**	**11**	**7**	**5**	**9**	**3**	**7**	**20**	**10**	**8**	**10**	**18**	**11**	**9**	**21**	**22**	**21**	**18**	**4**	**11**	**11**	**1**			**16**	**15**
(isoform 3)		**[R].TALVLTPLK.[W]**	[46-54]	1×Carbamyl [N-Term]	998.6245																	**1**									
		[K].WYQSIRPPYP.[S]	[55-64]	1xDeam [Q3]	1307.6419			**5**	**1**	**1**		**15**		**4**		**3**		**4**		**48**	**35**	**15**	**2**	**4**						**5**	**2**
		[K].WYQSIRPPYPS.[Y]	[55-65]	1×Deam [Q3]	1394.6739																	1									
		[K].WYQSIRPPYPSYGYEPMG.[G]	[55-72]	1xOx [M17]; 1xDeam [Q3]	2207.9743			1																							
AMELY	**Q99218**	**[R].IALVLTPLK.[W]**	[46-54]		967.6550			**2**	**2**			**2**	**2**			**1**		**2**													
AMELY	**Q99218-1**	*[N].FSYEVLTPLK.[W]*	[31-40]		1196.6562																										
(isoform 1)		*[F].SYEVLTPLK.[W]*	[32-40]		1049.5877															12		2									
		[S].YEVLTPLK.[W]	[33-40]		962.5557																										
		[K].WYQSMIRPPY.[S]	[41-50]	1×Ox [M5]; 1×Deam [Q3]	1357.62454															1		1									

Figure 5

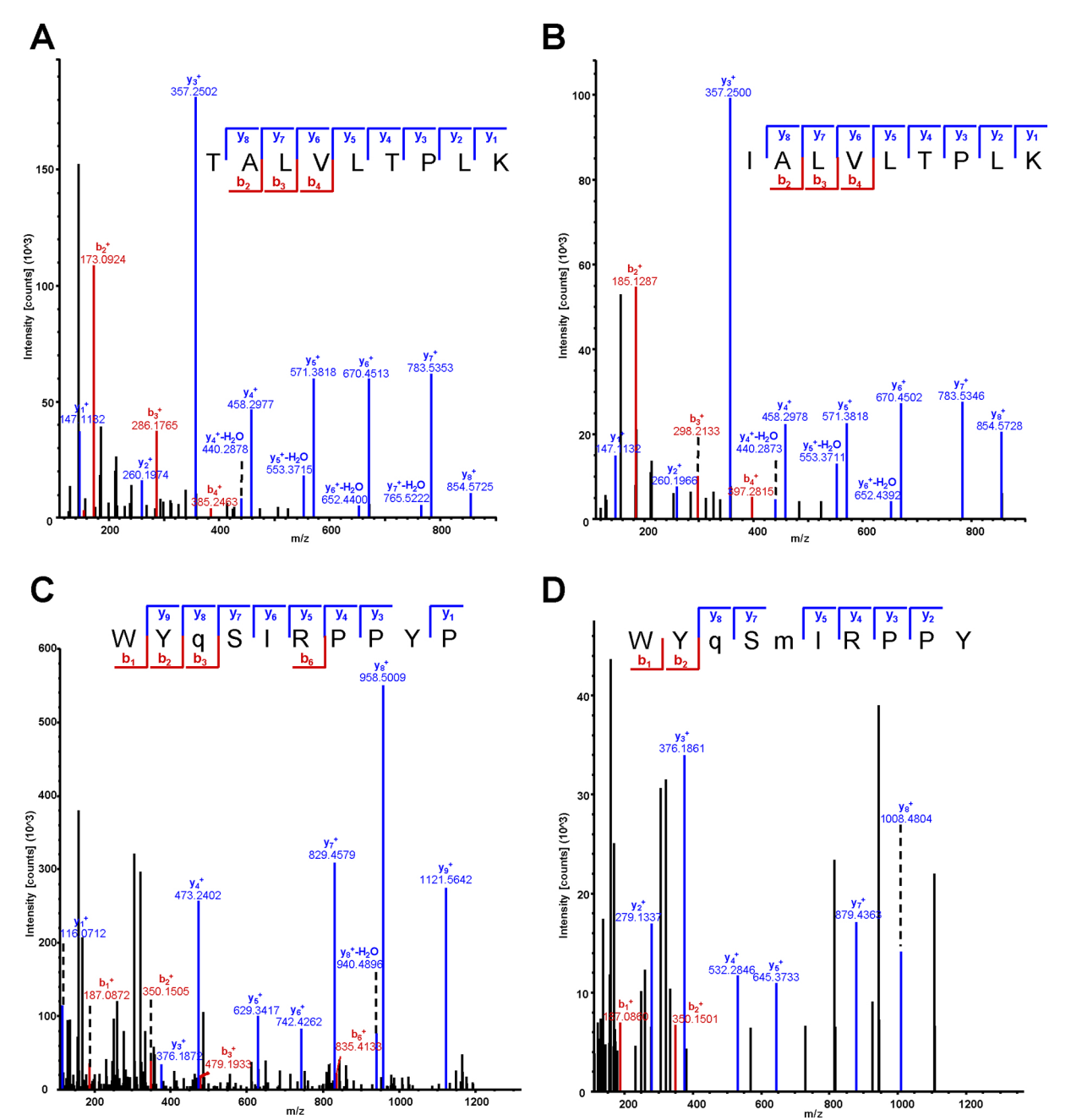
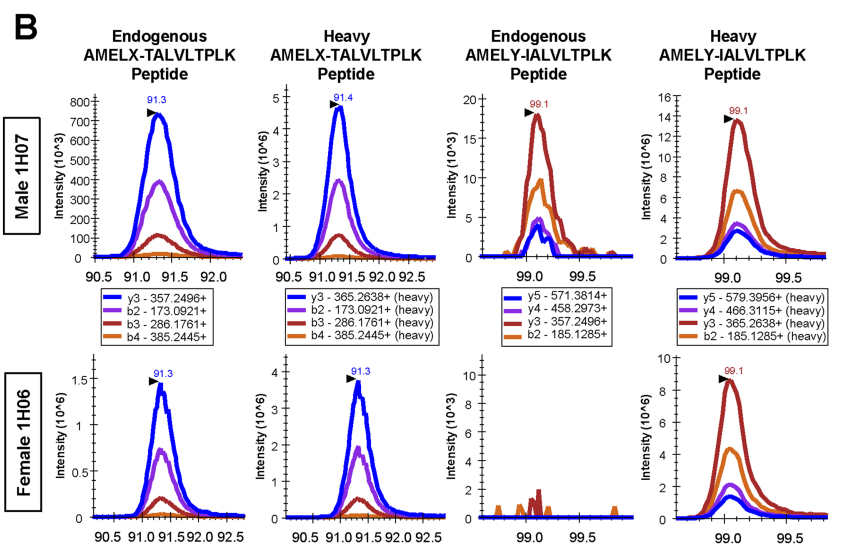

A
T A L V L T P L K
y3+ 357.2502
b2+ 173.0924
y1+ 147.1132
b3+ 286.1765
y2+ 260.1974
b4+ 385.2463
y4+ 458.2977
y4+-H2O 440.2878
y5+-H2O 553.3715
y5+ 571.3818
y6+-H2O 652.4400
y6+ 670.4513
y7+-H2O 765.5222
y7+ 783.5353
y8+ 854.5725
Intensity [counts] (10^3)
m/z
B
I A L V L T P L K
y3+ 357.2500
b2+ 185.1287
y1+ 147.1132
b3+ 298.2133
y2+ 260.1966
b4+ 397.2815
y4+ 458.2978
y4+-H2O 440.2873
y5+-H2O 553.3711
y5+ 571.3818
y6+-H2O 652.4392
y6+ 670.4502
y7+ 783.5346
y8+ 854.5728
Intensity [counts] (10^3)
m/z
C
W Y q S I R P P Y P
y8+ 958.5009
y7+ 829.4579
y9+ 1121.5642
y4+ 473.2402
y1+ 116.0712
b1+ 187.0872
b2+ 350.1505
y3+ 376.1873
b3+ 479.1933
y6+ 629.3417
y6+ 742.4262
y8+-H2O 940.4896
b6+ 835.4138
Intensity [counts] (10^3)
m/z
D
W Y q S m I R P P Y
y3+ 376.1861
y8+ 1008.4804
y7+ 879.4363
y2+ 279.1337
y4+ 532.2846
y5+ 645.3733
b1+ 187.0866
b2+ 350.1501
Intensity [counts] (10^3)
m/z

Figure 6

A

Sample	Sex	Peak area measurements ± SD					
		TALVLTPLK		WYqSIRPPYP		IALVLTPLK	
		Label free	PRM	Label free	PRM	Label free	PRM
1H01	M	5.45E+08 ± 3.61E+08	2.06E+08 ± 1.25E+07	1.14E+08 ± 2.39E+07	1.33E+06 ± 1.71E+05	6.74E+06 ± 4.81E+06	4.47E+06 ± 2.11E+05
1H05	M	5.74E+08 ± 1.34E+08	9.72E+07 ± 3.32E+07	4.28E+07 ± 1.93E+07	1.52E+05 ± 9.40E+04	2.89E+06 ± 1.73E+06	9.93E+05 ± 3.47E+05
1H07	M	8.40E+07 ± 6.60E+07	3.54E+07 ± 2.42E+06	6.48E+07 ± 2.13E+07	5.67E+05 ± 3.91E+04		5.31E+05 ± 4.23E+04
1H12	M	3.43E+07 ± 7.45E+06	1.59E+07 ± 1.20E+06	2.86E+06 ± 4.87E+05	6.15E+04 ± 1.47E+04		8.27E+04 ± 2.93E+03
2H08	M	3.94E+08 ± 2.13E+08	1.72E+08 ± 3.05E+07	1.53E+08 ± 6.61E+07	1.72E+08 ± 3.05E+07	4.73E+06 ± 3.04E+06	4.28E+06 ± 5.70E+05
2H10	M	4.50E+07 ± 2.57E+07	2.18E+07 ± 1.05E+06	3.07E+07 ± 8.79E+06	2.18E+07 ± 1.05E+06		2.43E+05 ± 2.21E+04
2H12	M	2.60E+08 ± 1.37E+08	8.00E+07 ± 2.16E+07	3.92E+07 ± 9.03E+06	8.00E+07 ± 2.16E+07	2.41E+06 ± 2.67E+06	1.22E+06 ± 3.29E+05
CW02	M	2.12E+08 ± 1.09E+07	5.96E+07 ± 2.51E+06	1.67E+08 ± 7.59E+06	5.96E+07 ± 2.51E+06	3.74E+05	3.63E+05 ± 1.79E+05
CW03	M	2.23E+08 ± 8.28E+06	5.83E+07 ± 1.00E+07	1.89E+08 ± 2.67E+06	5.83E+07 ± 1.00E+07	1.25E+06 ± 2.55E+05	1.05E+06 ± 2.87E+05
1H06	F	2.13E+08 ± 3.97E+07	5.77E+07 ± 7.87E+06	6.11E+07 ± 2.77E+06	5.77E+07 ± 7.87E+06		
1H15	F	1.49E+07 ± 3.83E+05	8.23E+06 ± 3.96E+05		8.23E+06 ± 3.96E+05		
1H18	F		9.26E+04 ± 1.30E+04	6.71E+06 ± 2.93E+06	9.26E+04 ± 1.30E+04		
2H23	F	1.63E+09 ± 2.53E+08	5.19E+08 ± 3.34E+07	8.57E+07 ± 7.72E+07	5.11E+05 ± 4.48E+04		

B



C

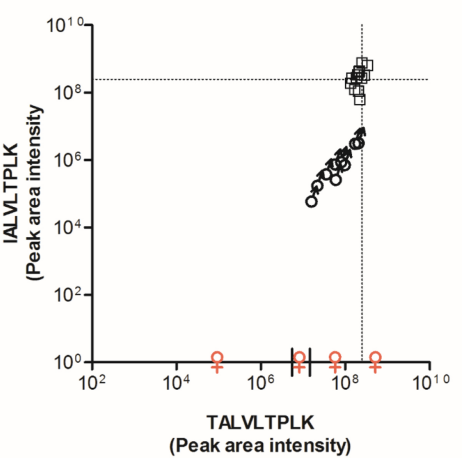


Figure 7