

Università degli Studi di Napoli Federico II



Ph.D. Thesis in Chemical Sciences

XXXIII Cycle

PROTEOMICS FOR ANCIENT PROTEIN DEGRADATION

Georgia Ntasi

Tutor

Prof. L. Birolo

Supervisor

Prof. A. Vergara

Coordinator

Prof. A. Lombardi

Acknowledgments

University of Naples “ Federico II”

Prof. Leila Birolo

Prof. Alessandro Vergara

Prof. Gennaro Marino

Prof. Andrea Carpentieri

Prof. Paola Giardina

Prof. Angela Amoresano

Prof. Pier Paolo Petrone

Prof. Angela Lombardi

University of Copenhagen

Prof. Enrico Cappellini

University of Cambridge

Prof. Matthew Collins

University of Pisa

Prof. Maria Perla Colombini

Prof. Ilaria Bonaduce

Dr. Anna Liuveras Tenorio

University of Suor Orsola Benincasa

Museum “Del Prado”, Madrid

Museum of Fine Arts, Boston and special thanks to Dr. Daniel P. Kirby

My colleagues:

Dr. Gabriella Pinto

Dr. Anna Illiano

Dr. Chiara Melchioro

Dr. Matteo Imbelloni

Dr. Carolina Fontanarosa

Dr. Michele Spinelli

Dr. Paola Cicatiello

Dr. Ilaria Stanzione

Dr. Rosanna Pitocchi

All students that I had the honor to co-supervised in these years

All TEMPERA colleagues (<https://sites.google.com/palaeome.org/tempera>)

My family and my friends that became family



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 722606".

The science of this book is nothing but a search for the truth, aiming at the best hypothesis to open the way for the next best one.

Index

Abstract.....	11
References	12
1. Introduction	14
1.1 What is cultural heritage, and why is it important?	14
1.2 Chemistry and cultural heritage	14
1.3 Archaeometric analysis	15
1.4 Organic residues in Cultural Heritage	16
1.4.1 Works of art	16
1.4.2 Human and animal remains	17
1.4.3 Contaminants.....	17
1.5 Analytical approaches in cultural heritage studies	18
1.5.1 Mass spectrometry	18
1.5.2 Spectroscopic techniques	29
1.5.2.2 FT-IR	31
1.5.2.3 Optical microscopy.....	32
1.6 Palaeoproteomic: A window into the past	33
References	36
2. Aim of the thesis	40
Palaeoproteomic projects performed during this PhD.....	41
List of publications	42
3. Proteomics strategies for the characterization of collagen-based materials.....	44
Introduction	44
Aim of the study.....	44
Book chapter:.....	44
Manuscripts:	44
References	45
3.1 Manuscript 1: Chemical review on Palaeoproteomic analysis of bones and teeth.....	47
Abstract.....	47
Protein survival in ancient bones and dentine	47
Protein survival in ancient enamel.....	51

Palaeoproteomics and phylogeny.....	51
Contribution of palaeoproteomics to study ancient pathogens	54
Conclusions	56
Author Contributions	56
Acknowledgments.....	56
References	56
3.2 Manuscript 2: Written in bone proteins: molecular signatures in the 79 AD skeletal remains from Herculaneum and Pompeii.	61
Abstract.....	61
Introduction	61
Results.....	64
Discussion.....	76
Author Contributions	81
Acknowledgments.....	81
References	81
Supporting Information	84
References	129
3.3 Manuscript 2: Selection of peptide biomarkers in human bone collagen for the evaluation of deamidation (N, Q) and oxidation (M) by targeted proteomics.....	130
Abstract.....	130
Introduction	130
Materials and methods.....	131
Results & Discussion	138
Conclusions	148
Abbreviations	149
Author Contributions	149
Acknowledgments.....	149
References	149
Supporting Information	152
References	180
3.4 Manuscript 3: Molecular characterization of animal glues for restoration purposes.....	181
Abstract.....	181

Introduction	181
Results	183
Conclusions	195
Author Contributions	196
Acknowledgments.....	197
References	197
Supporting information	199
References	223
4. Development of micro-invasive techniques for protein identification and characterization in the field of cultural heritage.....	224
Introduction	224
Aim of study	224
Manuscripts:	225
4.1 Manuscript 1: A minimally-invasive and portable method for the identification of proteins in ancient paintings.....	226
Abstract.....	226
Novelty.....	226
Materials and Methods.....	227
Results and Discussion	229
Supporting material	234
Author Contributions	234
Acknowledgments.....	234
References	234
4.2 Manuscript 2: A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects.....	236
Introduction	236
Abstract.....	236
Significance	237
Materials and Methods.....	237
Results and discussion	240
Conclusions	250
Supporting material	250

Author Contributions	250
Acknowledgments.....	250
References	250
4.3 Development and future perspective for a micro-invasive tool for the analysis of proteins in historical objects.....	253
Introduction	253
Materials and methods.....	254
Results and Discussion	255
Supplementary material	262
References	288
5. Integration of mass spectrometric, spectroscopic and microscopic techniques for the molecular characterization of archaeological remains and works of art.	289
Introduction	289
Aim of study	289
Manuscripts	290
References	291
5.1 Manuscript 1: Novel use of Evolved Gas Analysis/Mass Spectrometry to identify the earliest organic binding medium in Aegean wall paintings from Tel Kabri, Israel, dated to the 18th C. B.C.E.....	292
Abstract.....	292
Introduction	292
Material and Methods	293
Results and discussion	294
Acknowledgments.....	300
References	300
5.2 Manuscript 2: A multidisciplinary campaign for the characterization of a XXII dynasty wooden Sarcophagus (yellow coffin).....	302
Abstract.....	302
Introduction	302
Sarcophagus description.....	304
Materials and Methods.....	305
Results and Discussion	310
Conclusions	318
Acknowledgements.....	319

References	319
5.3 Manuscript 3:Organic remains in amphorae from the temple of Hera in Paestum shed lights on solemn ceremonies.....	322
Abstract.....	322
Author’s contribution.....	322
Introduction	322
Materials and Methods.....	323
Results.....	324
Pre-screening by FT-IR spectroscopy	324
GC-MS analysis.....	328
LC-MS/MS analysis.....	334
Conclusions	335
Acknowledgements.....	335
References	336
5.4 Manuscript 4: Wall paintings from a Roman domus in Santa Maria Capua Vetere	341
Abstract.....	341
Author’s contribution.....	341
Introduction	341
Aim of study	341
Materials and methods.....	342
Results and Discussion	343
Raman Analysis	344
Analysis with stereo microscope and petrographic microscope	346
GC-MS Analysis of organic binders	355
LC-MS/MS Analysis of proteinaceous binders	357
Conclusions	360
References	361
5.5 Manuscript 5: Molecular characterization of protective layers in statues from the monumental staircase in the Royal Palace of Caserta	366
Introduction	366
Aim of the study.....	367
Materials and methods.....	367

Results	368
Conclusions	372
References	372
6.Final conclusions	377
7.PhD Course Activity Summary	381
Attended courses	381
Attended seminars	381
Attended congresses /summer schools.....	382
Periods abroad	383
Support for teaching activity.....	383
Publications.....	383
Innovations	384

Abstract

Preservation of our cultural heritage is vital as it keeps alive the past that determines who we are, where we came from, and the human path's achievements over the years. Material culture, defined as every physical object, resource, and space that defines our culture, is powerful and appealing because it is probably more stable than human identity and social communities. Its preservation includes both cultural and economic processes [1,2].

Chemistry has a very active role and contributes significantly to the preservation of our cultural heritage, helping to transmit this wealth to the next generations [3]. The molecular characterization of archaeological objects and works of art replies to crucial questions such as: HOW and WHEN a masterpiece has been created, as also WHICH are the chemical degradation mechanisms that took place through the years. The answers to these questions are what determine how it will be kept alive in the future. However, chemistry is called upon to face immense challenges concerning the micro-invasiveness of the chemical approach, the high heterogeneity of the material encountered, and the tiny amount of samples in some of the cases. As a result, there is a demand for sensitive, versatile, and minimally invasive analytical approaches that can shed light on the origin of the material and its conservation state without provoking further deterioration. Mass spectrometry is an analytical tool that gained more and more attention in the last years because it includes fast analytical techniques of high sensitivity and accuracy that are micro destructive.

Over the last 20 years, protein and mass spectrometry have been successfully adapted to the analysis of cultural heritage samples, actively contributing to the understanding of their origin and their preservation, a new scientific field known as palaeoproteomics [4]–[14]. Proteins consist of a great target because of their high resistance over the years, giving us the possibility to recover information further back in time than previously thought possible [15]. Palaeoproteomics is complementary to a variety of analytical techniques (physicochemical, geochemical, dating, microscopic etc.) That investigate components other than proteins in the archaeological object, aiming to the overall characterization of its origin, provenance and degradation state. The combination of these and proteomic approaches have obtained remarkable results in the field of cultural heritage, interpreting data related to the inorganic and organic material record of human cultural history[16][17].

This Ph.D. project focused on developing proteomic analytical strategies with the use of mass spectrometry in archaeological objects and works of art. These strategies were used as such or combined with data

obtained by other analytical approaches for a more in-depth insight of material origin and use and the chemical degradation mechanisms taking place over the years. After a brief introduction of the importance of cultural heritage, the types of cultural heritage materials and an overview on mass spectrometric analytical techniques herein used, chapter 2 explains the aim of the Ph.D. thesis and organizes the projects on the basis of the analytical techniques used in the single case of study. Chapter 3 reports a chemical review on palaeoproteomic analysis of bone and teeth as sources of collagen. Subsequently, the chapter reports proteomic strategies for the characterization of collagen-based materials ranging from archaeological bones to animal glues. Chapter 4 describes the development of novel micro-invasive techniques for the molecular characterization of proteins in archaeological objects and works of art. Chapter 5 reports the results of multidisciplinary approaches that combine proteomic strategies and mass spectrometry techniques with other analytical approaches such as microscopy and spectroscopy, with the final aim of characterizing organic materials in archaeological remains such as amphorae, historical paintings, and wooden coffins. All the data obtained from the aforementioned scientific projects have been later then organized in already published, under submission, or in preparation manuscripts.

References

- [1] A. M. Sullivan, "Cultural Heritage & New Media: A Future for the Past," *John Marshall Rev. Intellect. Prop. Law*, vol. 15, p. 604, 2016, Accessed: Sep. 09, 2020. [Online]. Available: <https://repository.jmls.edu/cgi/viewcontent.cgi?article=1392&context=ripl>.
- [2] "What is Cultural Heritage - Culture in Development." http://www.cultureindevelopment.nl/Cultural_Heritage/What_is_Cultural_Heritage (accessed Sep. 09, 2020).
- [3] R. Wiesinger and M. Schreiner, "Chemistry for cultural heritage," *Heritage Science*, vol. 3, no. 1. BioMed Central Ltd., p. 35, Dec. 01, 2015, doi: 10.1186/s40494-015-0064-z.
- [4] "Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings. - PubMed - NCBI." <https://www.ncbi.nlm.nih.gov/pubmed/30063323> (accessed Dec. 06, 2019).
- [5] E. Cappellini *et al.*, "Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins," *J. Proteome Res.*, vol. 11, no. 2, pp. 917–926, Feb. 2012, doi: 10.1021/pr200721u.
- [6] T. P. Cleland, E. R. Schroeter, and M. H. Schweitzer, "Biologically and diagenetically derived peptide modifications in moa collagens," *Proc. R. Soc. B Biol. Sci.*, vol. 282, no. 1808, pp. 20150015–20150015, 2015, doi: 10.1098/rspb.2015.0015.
- [7] R. C. Hill *et al.*, "Preserved proteins from extinct bison latifrons identified by tandem mass spectrometry; Hydroxylysine glycosides are a common feature of ancient collagen," *Mol. Cell. Proteomics*, vol. 14, no. 7, pp. 1946–1958, Jul. 2015, doi: 10.1074/mcp.M114.047787.
- [8] S. Orsini, A. Yadav, M. Dilillo, L. A. McDonnell, and I. Bonaduce, "Characterization of Degraded Proteins in Paintings Using Bottom-Up Proteomic Approaches: New Strategies for Protein Digestion and Analysis of Data," *Anal. Chem.*, vol. 90, no. 11, pp. 6403–6408, 2018, doi: 10.1021/acs.analchem.8b00281.
- [9] G. A. Castiblanco, D. Rutishauser, L. L. Ilag, S. Martignon, J. E. Castellanos, and W. Mejía, "Identification of proteins from human permanent erupted enamel," *Eur. J. Oral Sci.*, vol. 123, no. 6, pp. 390–

395, Dec. 2015, doi: 10.1111/eos.12214.

[10] S. Dallongeville, N. Garnier, C. Rolando, and C. Tokarski, "Proteins in Art, Archaeology, and Paleontology: From Detection to Identification," *Chem. Rev.*, vol. 116, no. 1, pp. 2–79, Jan. 2016, doi: 10.1021/acs.chemrev.5b00037.

[11] F. Welker *et al.*, "Ancient proteins resolve the evolutionary history of Darwin's South American ungulates," *Nature*, vol. 522, no. 7554, pp. 81–84, Jun. 2015, doi: 10.1038/nature14249.

[12] G. Ntasi *et al.*, "A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects," *J. Proteomics*, vol. 231, Jan. 2021, doi: 10.1016/j.jprot.2020.104039.

[13] T. P. Cleland, "Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics," *J. Proteome Res.*, vol. 17, no. 11, pp. 3976–3983, 2018, doi: 10.1021/acs.jproteome.8b00637.

[14] M. Mackie *et al.*, "Preservation of the metaproteome: variability of protein preservation in ancient dental calculus," *Sci. Technol. Archaeol. Res.*, vol. 3, no. 1, pp. 58–70, Jan. 2017, doi: 10.1080/20548923.2017.1361629.

[15] M. Buckley, "Paleoproteomics: An Introduction to the Analysis of Ancient Proteins by Soft Ionisation Mass Spectrometry," Springer, Cham, 2018, pp. 31–52.

[16] R. Linn *et al.*, "Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings," *Angew. Chemie Int. Ed.*, vol. 57, no. 40, pp. 13257–13260, Oct. 2018, doi: 10.1002/anie.201806520.

[17] P. Cennamo, M. Rosaria, B. Lumaga, G. Fatigati, A. Amoresano, and A. Carpentieri, "CONSERVATION SCIENCE A MULTIDISCIPLINARY ASSESSMENT TO INVESTIGATE A XXII," vol. 11, no. 1, pp. 25–38, 2020.

1. Introduction

1.1 What is cultural heritage, and why is it important?

Cultural heritage is usually considered as material artifacts (paintings, drawings, prints, mosaics, and sculptures), historical monuments and buildings, as well as archaeological sites. Nevertheless, the cultural heritage is not exclusively referred to material objects but also consisted of immaterial elements such as traditions, rituals, etc. For this reason, the term of cultural heritage is expressed either as tangible or intangible [1][2]

The conservation of cultural heritage is essential for the conservation of our history and its transmission from one generation to the others. Thus, it is essential to respect and safeguard the whole cultural heritage through national laws and international treaties, and UNESCO (United Nations Educational, Scientific and Cultural Organization), founded in 1954, has adopted international conventions on the protection of cultural heritage.

Briefly, the conservation of cultural heritage applies simple ethical guidelines [3][4]:

- Minimal intervention;
- Appropriate materials and reversible methods;
- Full documentation of all work undertaken.

1.2 Chemistry and cultural heritage

When we visit a museum or an art gallery, we easily forget that much chemistry lies behind the exhibits. For instance, in the case of historical painting, chemistry provides innumerable information regarding the chemical composition referring to proteinaceous binders, colors, varnishes, and inorganic materials used by the artist. This information is essential not only for restoration purposes but also provides important historical background about the pictorial style of that period. Chemistry and cultural heritage are strictly connected since the former reveals information that might have been lost through the years, basically due to the chemical damage that the archaeological samples have undergone. Briefly, the questions that chemistry is usually called upon to answer in cultural heritage studies are, among others:

- Chemical composition of archaeological remains in amphoraes, utensils etc for a deeper understanding of their use thereof, revealing food and cultural habits of the past.
- Decoration techniques of archaeological objects and important archaeological sites.

- Identification of organic residues in works of art, such as paintings, helps to shed light on the used painting techniques, to attribute unknown artworks, and even more helps to an effective restoration intervention.
- Molecular characterization of ancient skeletal tissues (human, animal) for archaeological, paleontological, phylogenetic, and taxonomic studies.
- Investigation of the chemical transformations either due to aging or due to human intervention during restoration processes.

In this context, however, it is necessary to use techniques that require the use of the minimum amount of sample and which can reduce the invasiveness in order to avoid further deterioration.

1.3 Archaeometric analysis

Archaeometric analyses can be considered the link between Natural Sciences and Humanities, including a series of chemical techniques and approaches with the ultimate goal of answering several archaeological questions and problems. These questions are about how people in the past used materials either for food processes (i.e., amphorae) or for living needs (i.e., tools), from their use to their production and finally their deposition in the fossil record. There are several chemical techniques that have contributed significantly to this goal in recent decades, such as radiocarbon dating techniques, but in this dissertation, we addressed our attention on the great potentialities of mass spectrometry, with a special focus on gas chromatographic separation coupled to mass spectrometric analysis (GC-MS). The analytical methodology requires in a first instance the extraction of organic molecules, which is one of the most difficult and challenging issues. We should remember that organic waste has been deposited in ancient objects for centuries or even thousands of years and has degraded over the time. The extracted polar and non-polar molecules are then chemically derivatized to become volatile and thus easily ionized and analyzed by mass spectrometry. Some of the biggest challenges can be summarized in:

- Difficulties that are related to sampling.
- Environmental and human contamination of the samples (also modern contamination).
- Chemical modification procedures.

The first problem is intrinsically related to the nature and type of analysis. The creative use of appropriate procedures, in combination with the technological developments of instruments, makes it possible to identify the molecules in the material under investigation. The other two questions are connected with the relationship and interaction of the object under investigation with the environment as well as with

decomposition processes. It is absolutely necessary to verify the presence of contaminants due to the specific context in which the object has been found and to verify the degree of degradation of the organic residues present in the ceramic vessel. The presence of contaminant molecules in the surrounding soil can obviously be very variable in type and quantity. The collection and analysis of samples of surrounding soil are, therefore, crucial for the proper evaluation of the results obtained. Moreover, degradation processes strongly depend on the type of residues, on the type of soil, on humidity and on the temperature. However, the penetration of organic matter into the ceramic body of the container somehow protects against decomposition and deterioration caused by time and contact with the soil [5][6]. The type of organic residues analyzed by archaeometric techniques are references in the next chapter.

1.4 Organic residues in Cultural Heritage

Organic residues can be recovered in archaeological objects, works of art, as well as, of course, in animal, human, and plant remains. Palaeoproteomics and the recent advances in mass spectrometry have allowed the analysis of organic residues despite their high degradation state. The analysis of proteins and organic molecules, in combination or alone, provides essential information about ancient societies, their life, and their customs. This chapter provides an overview of the organic materials found in archaeological and artistic materials.

1.4.1 Works of art

Organic substances have been employed in the production of various works of arts, such as paintings, textiles, books, archaeological objects, adhesive for gilding, additives for plaster, mortars, and paint grounds. The most frequently found components can be divided into seven main categories:

- Proteins: Egg, milk, animal glue, silk, wool, plant proteins, human and animal tissues [7]
- Lipids: Animal fats, vegetable & drying oils [8]
- Waxes: Beeswax, paraffin, vegetable and animal waxes [9,10]
- Resins: Pine resins, plant resins such as styrax, animal resins and resins coming from the thermal degradation of wood [11]
- Carbohydrates or polysaccharides: Starch, cellulose, plant gums (Arabic gum, tragacanth, ghatti, guar, locust bean, fruit tree gum) [12]

- Dyes: Cochineal, madder, kermes, saffron, purple, indigo, synthetic [12]

1.4.2 Human and animal remains

Bones are one of the most well-preserved remains in the fossil record. In the past, they were used for practical, hunting, symbolic, artistic, and ornamental reasons. Moreover, an enormous amount of information is stored in human bones and in animal remains and in teeth. Evidence of disease or illness is often embedded in bones or in teeth; thus, the analysis of archaeological bones reveals important information about the diseases of the past. It is often possible to determine the age at death, cause of death, sex, occupation, movement, and nutritional status. The reason why the scientific community focused on bone and teeth analysis is its resistance, that is due to its peculiar composition. Bones are constituted for up to 30% of organic material, mostly collagen proteins; teeth are usually constituted of enamel and dentine, both containing proteins. Bone and enamel, in particular, thanks to their high mineral content, are known to provide protein preservation for a long time [13,14]. A more detailed description of the palaeoproteomic investigation of bones and teeth is given in the third chapter of the thesis.

1.4.3 Contaminants

Contaminants often consist of more than 40% of the molecules identified in an archaeological sample[15] . This makes sense given the antiquity of the samples but also the human intervention. Their provenance depends on the type of the analyzed material. For instance, in the analysis of the organic molecules from archaeological remains such as amphoraes, contaminants come from the soil in the environments (natural compounds, bacteria, fungi, all irrelative to the amphorae's content and use in the past) or from human presence during excavation and sample collection or later on in the laboratory.

In the studies of skeletal contaminants are prevalently coming from skin and hair in dust particles. Moreover, skin contamination is widespread because of some handling of the objects without gloves, a practice very diffused in the past in conservation science environments. Sometimes identification of the molecules becomes quite tricky such as because of the huge presence of contaminating keratins impairing confidence attribution. Therefore, a series of precautions in the excavation process and subsequent processing of ancient remains are welcome and required. When unavoidable contaminants have to be considered, creation of contamination lists, inserted in the analytical instrument software either before analyses or during the data processing step can be adopted.

1.5 Analytical approaches in cultural heritage studies

1.5.1 Mass spectrometry

Mass spectrometry (MS) is the most used modern analytical techniques for the identification, characterization, and quantification of a big variety of molecules, ranging from small to big m/z .

The most important characteristics of mass spectrometry are reproducibility, sensitivity, resolution, and accuracy. In particular, mass spectrometry is based on the determination of the molecular mass to charge ratio. Its unique capabilities and the ability to analyze molecules even when in complex mixtures, make this methodology suitable for the characterization of cultural heritage material and works of art [14].

Major issues in a cultural heritage study are the degree of invasiveness and destructiveness, where invasiveness refers to the sampling method, while destructiveness refers to the actual analysis of that sample. The intrinsic heterogeneity and complexity of cultural heritage samples and the necessity to use a small quantity of samples, in order to preserve the entire heritage, makes vital to apply techniques that are barely invasive and very sensitive [14]. Mass spectrometry covers all of the aforementioned issues, as being sensitive since it requires a minimal amount of sample and being accurate since it gives reliable and high-resolution results.

The chemical analysis of a cultural heritage object can be divided into the analysis of the organic molecules (proteinaceous materials, sugars, lipids) and the analysis of the inorganic ones. The first can be approached with mass spectrometry techniques (LC-MS, LC-MS\MS, GC-MS), while the inorganic part is commonly approached by spectroscopic techniques (Raman, FT-IR, etc.)

1.5.1.1 Basic Principles of Mass spectrometry

A mass spectrometer is an instrument that generates ions in the gas phase and separates them according to their mass-to-charge ratio (m/z). It consists of four components: the inlet system, for the introduction of the sample; the ion source, where the ionization of analytes takes place; the mass analyzer, that separates the ions with different m/z values, and the detector, that reveals the ions separated by the mass analyzer and generates a quantifiable impulse which can be translated into a mass spectrum by the data system (Figure 1).

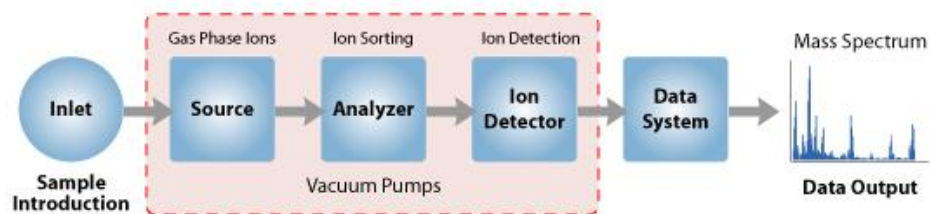


Figure 1: - General scheme of a mass spectrometer (<http://www.premierbiosoft.com/>).

Generally, mass spectrometers can be distinguished both on the type of ionization system and on the type of analyzer; the latter is an essential component to define the accessible mass range, sensitivity, and resolution of an instrument [16]. All mass spectrometers operate at low pressure (from 10^{-5} to 10^{-9}) as in case of high pressure, ion collisions can be provoked during the flight from the ion source to the mass detector. The stages of any mass spectrometric analysis are ion generation, ion acceleration, the direction of ions to the mass analyzer, and ion detection. Ionization occurs in the first part of a mass spectrometer. A single ionization technique cannot ionize all kinds of molecules, but different ionization techniques are now available. Their choice is based on the chemical-physical properties of the molecule, such as molecular weight, polarity, thermal stability, etc. However, only ions either positively or negatively charged are analyzed, although the most used ionization mode is the positive one, performed at low pH.

In any MS experiment, the first step is the introduction of the sample into the mass spectrometer that is performed with an inlet system. There are different ways to introduce the sample, depending on its purity and properties. The simplest way is the direct injection of the sample. The direct inlet design depends on the ionization techniques that will be used in the following step. However, it is essential that the analytes are not introduced into the mass spectrometer at the same time as this would provoke ion saturation, leading to a superimposed mass spectrum.

This coupling is generally on-line, and it allows the combination of complementary features: high sensitivity, selectivity, and specificity offered by MS with high performance in separation.

The term 'soft' or 'hard' ionization techniques are referred to the amount of energy deposited onto the molecule during the ionization process. In hard ionization techniques, among which electron ionization (EI) is the most common, the use of high energies results in an abundant ion fragmentation. In soft ionization methods, the excess energy deposited onto the ionized molecule is very small, and stable even-electron ions

are formed. Soft ionization methods can be classified into three different groups: (a) gas-phase soft ionization; (b) spray ionization techniques; (c) desorption ionization techniques [17].

In this Ph.D., the ionization techniques that were mostly used are electrospray ionization (ESI) as spray ionization technique and matrix-Assisted Laser Desorption Ionization (MALDI) as desorption ionization technique.

Electrospray Ionization (ESI)

In ESI, ions are formed from peptides and proteins by spraying a dilute solution of these analytes at atmospheric pressure from the capillary, due to the potential that is generated between capillary and inlet to a mix of charged droplets [18]. As the droplets evaporate, the peptide and protein molecules in the droplets pick up one, two, or more protons from the solvent to form singly or, more frequently, multiply charged (e.g., $[M+H]^+$, $[M+2H]^{2+}$) ions. The number of charges acquired by a molecule depends on the molecule characteristics and charges follow a Gaussian distribution. As the droplets continue to shrink, the charge density on the surface of each droplet increases to the point where charge repulsion overcomes the forces holding the droplet and the solvated ions contained within it together. Ions are then 'emitted' or 'evaporated' from the droplet surface via coulombic repulsion. The ions are sampled into the high-vacuum region of the mass spectrometer for mass analysis and detection (Figure 2). The typical solvent for peptides and proteins is a mixture of water and an organic modifier such as CH_3CN , and up to a few percent by volume of trifluoroacetic acid, formic, or another volatile acid to enhance ionization of sample constituents [19]. A significant improvement in ESI technology consists of the development of nanospray ionization. The low flow rates possible with nanospray ionization reduce the amount of sample consumed and increase the time available for analysis [20].

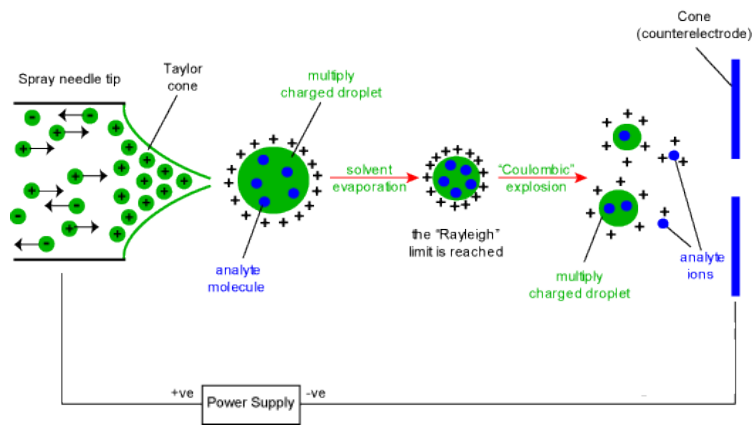


Figure 2: A scheme of ESI ionization technique (www.macscientists.com)

Desorption Ionization Techniques

Matrix-Assisted Laser Desorption Ionization (MALDI)

In this method, the analyte should be mixed with a large excess of an aromatic 'matrix compound' that has a strong absorption at the wavelength of the laser used to excite the molecules. For instance, the matrix compound *o*-cyano-4-hydroxycinnamic acid can absorb the energy from a nitrogen UV laser (337 nm). The addition of the matrix compound ensures that the the laser pulse energy will be absorbed by the matrix and transferred to the analyte in the sample. By this way, chemical decomposition of the analyte is avoided [21].

The analyte and matrix are dissolved in an organic solvent and then spotted on a metallic target. Later, the solvent evaporates, leaving the matrix and analyte co-crystallized. The target is placed in the vacuum chamber of the mass spectrometer, and a high voltage is applied. At the same time, the crystals are bombarded with a short laser pulse (1–20 ns), which causes the sample and matrix to be volatilized. The formation of ions occurs by an acid-base reaction between the ionized matrix molecules and the analyte molecules. Desorption ionization processes form finally singly charged (protonated or deprotonated) molecules of the analyte with a range of dimensions from a few hundred to several hundred thousand Daltons [22]. This process is simply illustrated in Figure 3.

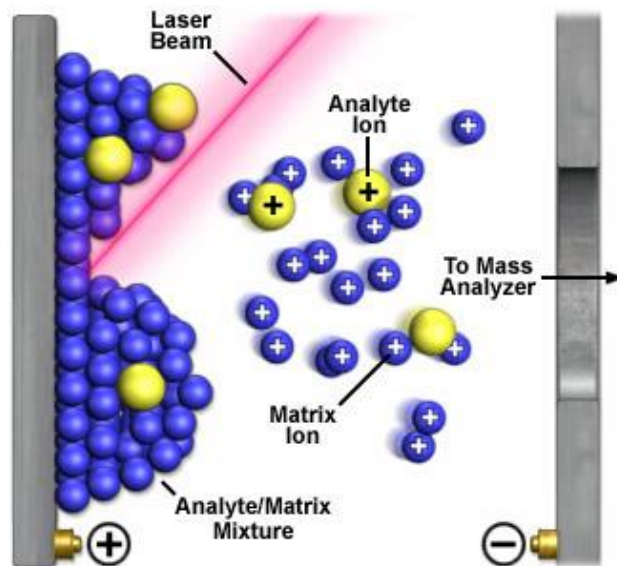


Figure 3: A scheme of MALDI ionization technique (nationalmaglab.org).

The Analyzer

Ions formed in the ions source are then accelerated to the mass analyzer, where they are separated according to their m/z ratio. The mass analyzer plays an essential role as it defines the performance of the instrument, including accuracy, resolution, sensitivity, and MS/MS capability. The main types of mass analyzers are: Analyzers are classified according on ion separation mode. Quadrupole (Q), Time Of Flight (TOF), as also the magnetic and electrostatic double sector mass analyzers, separate in space, while Orbitrap, ion traps, and Fourier circular ion transformation (FTICR) analyzers, separate the ions in time [17].

Quadrupole Analyzer

The quadrupole mass filter consists of a linear array of four symmetrically arranged rods to which radio frequency (RF) and DC voltages are supplied. Forces are exerted in a plane normal to the direction (z-direction) in which the ions drift. Ions oscillate in the x,y-plane and the frequency of oscillations depends on their m/z values and on the applied potentials. If the oscillations of an ion within this plane are constant, the ions will continue to drift down the assembly and reach the detector. Constant oscillations are achieved only with ions given m/z value for a given rod assembly, oscillation frequency, RF voltages, and DC voltages. All other ions do not have a fixed orbit and will never reach the detector. The RF is varied to bring ions of different m/z into focus on the detector and thus build up a mass spectrum.

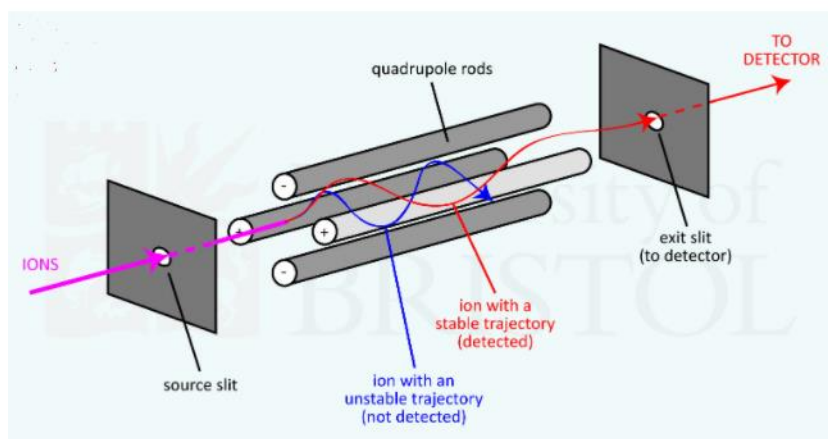


Figure 4: A scheme of quadrupole mass analyzer (<http://www.chm.bris.ac.uk/> © Paul. J gates 2014)

Time of Flight Analyzer

In Time-of-flight mass spectrometry (TOF-MS) an ion's mass-to-charge ratio is calculated after a time of flight measurement. In particular, ions are accelerated in an electric field, and those that are charged have gained kinetic energy.

The time the ions take to travel to the detector depends on the mass-to-charge ratio (the lower the m/z the faster). The time for each ion to reach a detector at a known distance is measured, a value that strictly depends on the ion velocity and thus is a measure of its mass-to-charge ratio [23].

Many improvements have been made through the years, the most important being the introduction of “reflector” in the instrumentation of TOF analyzer, that allow all ions with the same m/z values but different kinetic energy to arrive at the same time to the detector, resulting in narrow peaks and higher resolution (Figure 5).

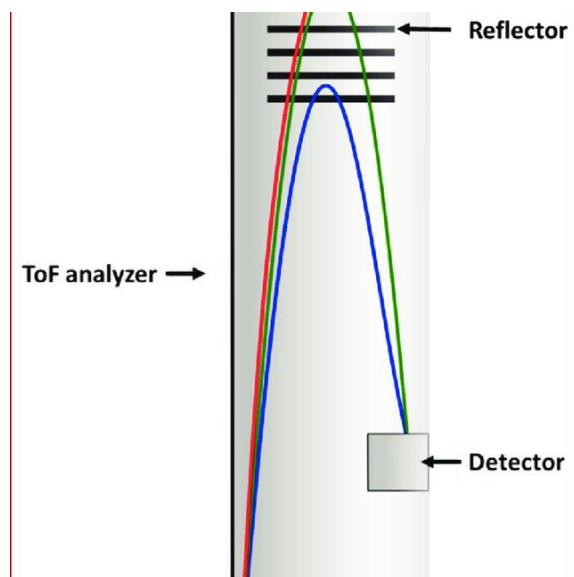


Figure 5:- A scheme of TOF analyzer with reflector [24].

One of the significant advantages of the TOF analyzer is that it has no restriction on the m/z range, which enables it to analyze ions with m/z values ranging from a few hundred to 500000 and higher. This is an essential characteristic for its coupling with MALDI in the study of large molecules. In fact, as ions produced by MALDI are generally singly charged species, their m/z values can be very high. Due to its high resolution and mass accuracy capabilities, TOF is currently coupled with all ionization techniques.

Orbitrap Analyzer

In Orbitrap, ions are trapped in an electrostatic field produced by two electrodes: a central spindle-shaped and an outer barrel-like electrode (Figure 6). Ions move in harmonic, complex spiral-like motions around the central electrode while moving back and forth over its longitudinal axis in harmonic oscillation with a frequency that depends only on their values m/z [25]. The oscillating ions create an electric current in the outer electrode, and the mass spectrum is obtained by a Fourier transform of the recorded current. The longer the current recording period, the better is the resolution [26].

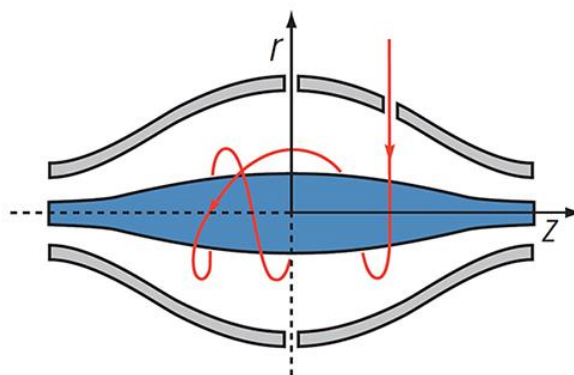


Figure 6:- A scheme of Orbitrap. Ions enter Orbitrap through the opening in the outer electrode. Under the influence of an electric field, they start to move around and along the central electrode creating concentric rings. The motion of ions induces an image current in the outer electrode, which is recorded and used for the determination of m/z ratios of ions by Fourier transform (www.creative-proteomics.com)

1.5.1.2 Tandem Mass Spectrometry

Tandem mass spectrometry is used to produce structural information about a compound by fragmenting specific analytes inside the mass spectrometer and analyzing the fragment ions. Tandem mass spectrometry also enables the identification of specific compounds in complex mixtures because of their specific fragmentation patterns. MS/MS is based on two stages of mass analysis, one to select a precursor (or *parent* ion) and the second to analyze fragment (or *daughter* ions). Instruments for tandem mass spectrometry can be classified as a *tandem in space* or *tandem in time* [27]. Tandem in space means that ion selection, ion fragmentation, and fragments analysis are events that occur in three different regions of the spectrometer; instruments of this type are, for example, the triple quadrupole (QqQ) and hybrid instruments such as QqTof or TOFTOF. In tandem in time instruments, those three steps of analysis occur in the same region of the spectrometer but at different times; QIT and LIT are instruments classified as a tandem in time.

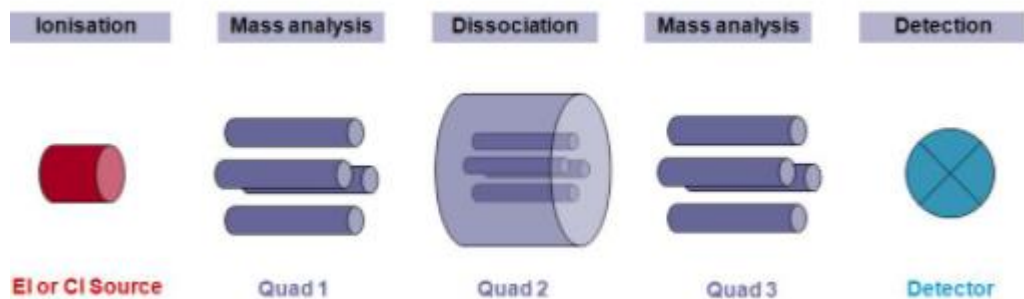


Figure 7: Scheme of tandem in space and in a triple quadrupole mass spectrometer (www.biologie.hu-berlin.de).

In tandem in the space mass spectrometer, the first mass analyzer is used to selectively pass an ion into a reaction region, while the second mass analyzer is used to record the m/z values of the dissociation products. Generally, the reaction region, or collision cell, is the region into which an inert gas (e.g. nitrogen, argon, helium) is admitted to collide with the selected ions causing the fragmentation. This process is called *Collision Induced Dissociation* (CID) [23]. Inelastic collisions between the precursor ion with high translational energy and a neutral target gas convert part of ion translational energy into internal energy, leading, therefore, to its decomposition. All CID processes can be separated into two categories based on the energy of the precursor ion. Low-energy collisions are common in triple quadrupoles (QqQ), hybrid devices, such as QqTof [28], and trapping devices, such as ion traps (IT). They occur in the 1–100 eV range of collision energy, while high-energy collisions, as in the sector and TOF/TOF instruments are in the keV collision energy range. The precursor ion beam, having a kinetic energy of a few keV, can enter the collision cell, usually causing single collisions before mass analysis of the product ions. Moreover, high-energy CID spectra usually show increased side-chain fragmentation.

Peptide precursor ions, dissociated by the most usual low-energy collision conditions, fragment along the backbone at the amide bonds, forming structurally informative sequence ions and less useful non-sequence ions by losing small neutral molecules like water, ammonia, etc. The amino acid backbone has three different types of bonds (NH- C-H, C-H-CO, and CO-NH), and each of them can be fragmented, originating different fragment ions. Each bond breakage leads to the creation of two species, one neutral and one charged. However, only the charged molecules will be later monitored by the mass spectrometer. At each amino acid residue, there are six possible fragment ions: the a, b, and c ions that are charged on the N-terminal

fragment, and the x, y, and z ions that are charged on the C-terminal fragment (Figure 8). The most common cleavage sites are at the CONH bonds, which give rise to the b and/or y ions [20,29].

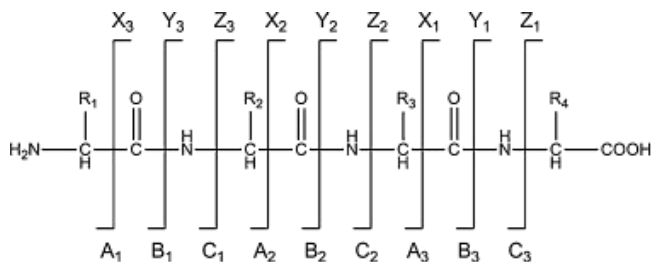


Figure 8: Peptide ion fragmentation nomenclature. The charge on the C-terminus designs a y-ion, whereas charging on the N-terminus designs a b-ion. The other additional types of fragmentation are also indicated[30].

The Mass Spectrum

The mass spectrum is a two-dimensional plot that reports the m/z ratio of ions and their relative intensity. This means that the most abundant ions are assigned as 100%. The appearance of the mass spectrum is determined by the amount of energy deposited on the molecule during the ionization. If it is high, the mass spectrum is characterized by molecular and fragment ions. However, in the case of soft ionization techniques, due to the low energy, the mass spectrum reports only the protonated/deprotonated molecule and a few adduct ions but not fragment ions [22].

Database Searching and Interpretations

Once a mass spectrum has been obtained, protein identification can be performed with the use of several bioinformatics tools [12]. In particular, MS data are compared to peptide mass values contained in the selected primary sequence database and that have been subjected to virtual hydrolysis. Software that are based on scoring algorithms return a list of proteins, each of which is associated with a probability index (score), calculated as $-10 \cdot \log P$, where P is the probability that the observed event of matching is random. Typically, the research is carried out with sequences that are present in NCBI and SwissProt databases and Chordata (Vertebrates and Relatives)/ All Entries as taxonomic restrictions. There are several bioinformatic tools in the literature to identify proteins, such as MASCOT [31], PEAKS [32], pfind [33], MaxQuant [34]. The choice of the search engine should be based on the user's specific aim since very often they do not contain the same features. This is particularly true in case of the discovery of protein chemical modifications. [35][36].

1.5.1.3 GC-MS

The GC-MS technique combines the separation capacity of gas chromatography (GC) with the potential of mass spectrometry (MS) for the qualitative and quantitative determination of a wide range of substances.

The advantages of this technique are:

- Greater sensitivity than detectors commonly coupled to GC (like the flame detector → FID)
- They offer more possibilities to identify the analyzed substances by comparison in databases of the mass spectra of each analyte.

To separate the different species that are present in a mixture, the GC uses their chemical properties for separation. Once the individual compounds of the mixture have been separated, they pass inside the MS, where they are subjected to ionization and consequent fragmentation (Figure 9).

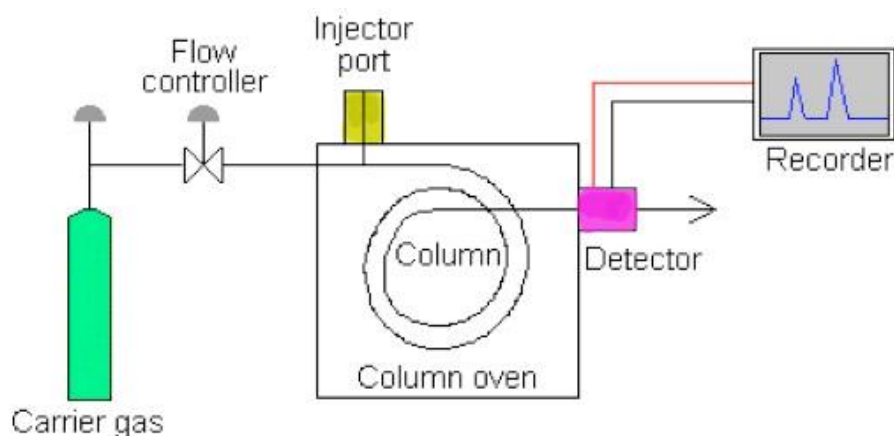


Figure 9: Simplified schematic diagram of a gas chromatograph (<http://diydiagram.academia-archi.it/>).

The mixture to be analyzed is introduced into the gas-chromatograph injection system and vaporized inside a high-temperature chamber so as to obtain a gaseous mixture which passes through a capillary column, which allows the separation of the components taking advantage of the different attitudes that each molecule or ions possess in the distribution in two phases: mobile and stationary phases.

The mobile phase, called "carrier," is an inert gas (He_2) that flows through the column in which it is located. The stationary phase is generally constituted by a capillary column of variable length between 30 and 60 meters and a diameter between 1 and $0.25\mu\text{m}$.

The capillary column is contained within an adjustable thermal system; in practice, the temperature gradients are constructed to allow the separation of the analytes on the basis of the boiling points and the polarity. It is possible to improve the resolution of a chromatogram by optimizing the temperature gradient. Once the boiling point has been reached, an analyte detaches from the column and, thanks to the thrust of the carrier gas, is brought into the source.

The ideal interface for the GC-MS is the electronic ionization source that allows the formation of ions through the impact of the analytes with a high energy electron beam (about 70eV) produced by an incandescent tungsten filament. When these electrons touch the electronic sphere of a molecule, the phenomenon of electronic impact occurs. The energy supplied by the electron beam also produces a breakdown of the molecules into smaller fragments, which can be positively or negatively charged.

The ions pass into a quadrupole that can operate in two modes: SIM, where a constant voltage is applied on the quadrupole bars, and therefore only some ions will be allowed to transit with a given m/z ratio; SCAN, where a whole mass range is scanned.

Finally, ions are detected by the energy of collision of ions on the detector surface; when a change in current is observed, the peak is recorded. The obtained chromatogram shows the ionic current (TIC) as a function of the retention time; in addition, we also get the spectrum of fragmentation that allows us to identify the substance under investigation. The spectra are acquired, stored, and presented; the identification is performed by comparing the spectra produced with known spectra contained in the database.[37,38]

To be able to analyze a mixture, compounds must be derivatized to become more volatile. There are several ways to derivatize:

- Silylation,
- Acylation,
- Cyclic derivatives formation,
- Alkylation,
- Transesterification. This allows us to identify fatty acids. It allows to carry out hydrolysis of the triglyceride, and to form the methyl-ester by reaction with alcohol, in particular with methanol.

1.5.2 Spectroscopic techniques

1.5.2.1 Raman spectroscopy

Raman spectroscopy is a rapid and effective technique for characterizing materials, giving essentially unambiguous results for inorganic and organic materials. It gives both structural and chemical information. It is a spectroscopic technique based upon the Raman effect, or scattering, discovered in 1928 by C.V. Raman [39,40], an Indian physicist Nobel laureate in 1931. *Raman scattering*, or *Raman Effect* [40], is an inelastic scattering of light: incident light can interact with atoms and molecules in different ways such as absorption, and both elastic and inelastic scattering:

- In the first case, the light incident is completely absorbed, and the system is transferred to an excitation state from which it can reach various lower levels. Decay to lower energy states comes with energy emission whose form is related with resonance lifetime length: most common decay mechanisms come with heat emission while resonance lifetime from 10^{-9} to 10^{-7} seconds order are typically related to fluorescent emission consisting in light emission at a higher wavelength ($\lambda_e > \lambda_i$) than incident radiation; longer decay times (10^{-3} to 10^{-2} seconds) are associated to phosphorescence emission which differs from fluorescence just for decay time. Other decay paths include induced resonance or an important biochemical process involved in photosynthesis, known as photo-oxidation.
- Elastic scattering, also known as *Rayleigh scattering*, is the elastic scattering of electromagnetic radiations by particles smaller than the wavelength of light itself (Figure 10). This common phenomenon is responsible, for example, for the blue and light blue coloration of the sky. Scattered photons present the same wavelength and frequency as incident photons resulting in no energy loss.
- In that case, a small fraction of the light is scattered following the Raman Effect mechanism (inelastic scattering), and the frequency among the photons and incident photons to be different (Figure 10).

Describing this phenomenon in terms of energy exchanges, a photon at a fixed frequency (ν_0), from a laser source, with an energy content described by $h\nu_0$, interacting with a molecule brings this latter from its fundamental state (a ground state in image) to an exciting *virtual energy level*. In the case of elastic scattering, decay brings to the emission of a photon with $\nu_e = \nu_0$, and the molecule returns to its actual state; in Raman Stokes scattering cases, instead, decay does not bring the molecule to its ground level but to a different vibration energy state (Figure 10): emitted photon will have *Stoke* frequency defined as:

$$\nu_0 - \nu_e = \nu_{st}$$

If the starting condition of the sample molecule is an energized state and, after excitation, decay brings it to ground level, we talk about Raman anti-Stokes scattering (Figure 2.2.1). In anti-Stokes energy, content is defined as:

$$\nu_0 + \nu_e = \nu_{ast}$$

These differences between the incident and emitted frequencies are referred to as **Raman Shift** that in spectroscopic and microscopic Raman applications are usually expressed in terms of wavenumber (cm^{-1}):

$$\Delta\omega_{\text{cm}^{-1}} = (\nu_0 - \nu_{st})/c$$

Where c represents the speed of light. Raman spectroscopy is widely used in chemistry since vibration information is specific to the chemical bonds and symmetry of the molecules.

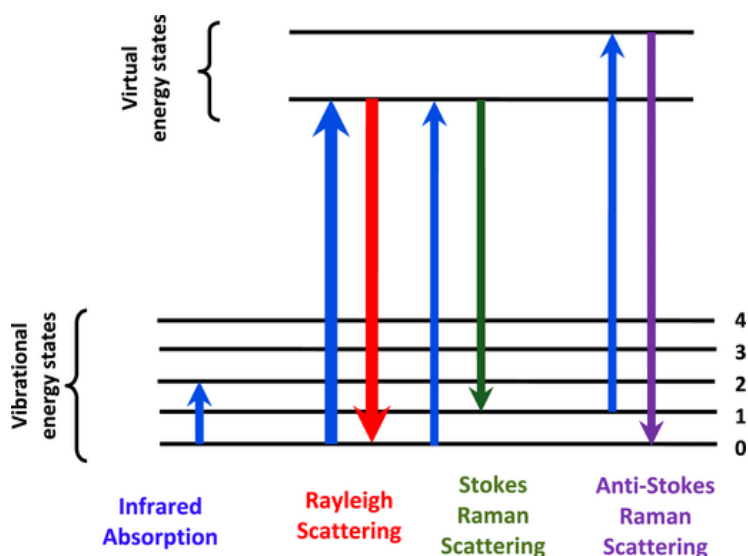


Figure 10: the different possibilities of visual light scattering: Rayleigh scattering, Stokes scattering, and anti-Stokes scattering (the atom or molecule loses energy; the emitted photon has more energy than the absorbed photon)[41].

Stokes bands are more intense than anti-Stokes bands and hence are measured in conventional Raman spectroscopy, while anti-Stokes bands are measured with fluorescing samples because fluorescence causes interference with Stokes bands [20].

1.5.2.2 FT-IR

Infrared spectroscopy is the analysis of the interaction of the material with electromagnetic fields in the IR region. In this region, a molecule can be excited to a higher vibrational state by absorbing IR radiation. The probability of a particular IR frequency being absorbed depends on the actual interaction between the electromagnetic radiation at these frequencies and the specific molecule. The frequencies of the absorption bands ν , as reported in Bohr's law, are proportional, through Plank's constant (h), to the energy difference between the vibrational ground and excited state (E) [42]

$$E = h\nu$$

FT-IR microscopy is an almost universal vibration technique, susceptible to the molecular structure and, therefore, to the chemical composition of the sample. The spectrum is expressed in terms of Transmittance or Absorbance expressed in function of the wavenumber in cm^{-1} . Transmittance (T) is defined as:

$$T = I/I_0$$

Where I is the radiation intensity transmitted and I_0 is the intensity of radiation incident. is expressed as the reciprocal of transmittance by the following equation

$$A = \log(1/T) = \log(I_0/I)$$

The general scheme of the FT-IR spectrometer is as follows:

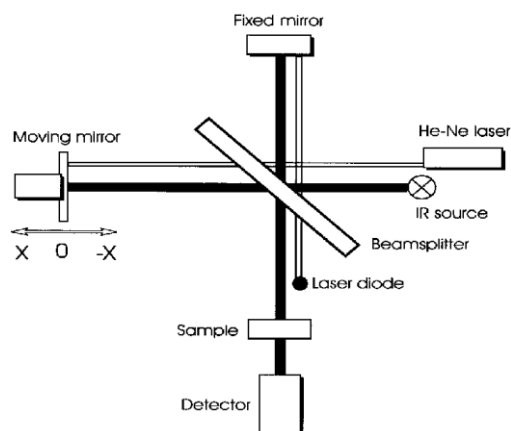


Figure 11: Diagram of an FT-IR spectrometer with Michelson interferometer[41].

The Michelson interferometer contains a beamsplitter and two flat mirrors. One of the mirrors is fixed in one interferometer arm, whereas the other one is movable in the second interferometer arm. Most common IR beamsplitters are made of KBr with a multilayer coating. The beamsplitter should have a reflectivity of 50 % and no absorption across its range of use. The functionality of a Michelson interferometer is based on a collimated IR beam. The latter is directed to the beamsplitter, which divides the beam into two parts of equal intensity (in the ideal case). The divided beams are reflected, by the fixed and the movable mirrors, back to the beamsplitter, where they recombine and interfere. The recombined IR beam passes the sample (or the reference) and reaches the detector. The mathematical procedure, which is employed to convert the IR interferogram (intensity versus time, also called time domain) to an IR spectrum (intensity versus frequency, also called frequency domain), is called Fourier transformation [42]

The main limitation of the FT-IR technique is the onset of strong IR bands due to water, which overlaps with important regions that make up the chemical fingerprint of the sample. In this sense, and by virtue of different selection rules, the IR microscope is the natural completion of Raman microscopy.

1.5.2.3 Optical microscopy

The petrographic microscope provides one of the primary means of studying both inorganic and organic materials. It is a specialized instrument that utilizes polarized light to allow the measurement of a variety of optical properties.

Optical microscopy allows to study and observe two-dimensional or three-dimensional samples. Especially in the two-dimensional microscopy is based on materials' thin section observation which optical properties are bound to structural, chemical traits. Petrological microscope takes advantage of polarized light's properties and its interactions with materials. In general, we can do observation in-plane light and crossed polarizers to investigate the optical phenomenon, such as Pleochroism, extinction, and color interference.[43,44]

The 3D observations allow morphological and textural characterizations, observation, and study of physical properties (growth effect, luster, color, fracture, inclusion, etc.). These observations can be realized on minerals and, of course, rocks, gems, cultural heritage objects, films, polymers, etc.

1.6 Palaeoproteomic: A window into the past

The advent of high-sensitivity mass spectrometry (MS) has allowed palaeoproteomics to become increasingly relevant in the fields of archaeology and evolutionary biology[45]. For instance, in the study of ancient bones and teeth, many MS-based techniques have been implemented to allow protein identification, species distinction, investigation of protein degradation pathways, and post-translational and taphonomic modifications. Given that nucleic acids are much less stable than proteins, the potentiality of recovering information from ancient proteins has intrigued scientists for over half a century. Ancient proteins are likely to suffer diagenetic alterations, and recent research has therefore largely focused on bottom-up, peptide-based proteomics also to characterize the molecular effects of aging and degradation. Soft ionization techniques such as ESI and MALDI have, therefore, been extensively used for the analysis of ancient proteins [46–48]. MS screening of species-specific peptides in specific proteins (mainly collagen) has recently been used as a low-cost, rapid alternative to DNA sequencing for taxonomic attribution of morphologically unidentifiable small bone fragments and teeth from diverse archaeological remains [49,50].

Amongst the techniques to approach the analysis of ancient objects, mass spectrometry (MS) has gained popularity because of its ability to handle the complexities associated with the proteome.

Palaeoproteomics aim to protein identification and to the characterization of post-translational modifications occurred either *in vivo* or during ageing. There are two proteomic approaches known as bottom-up and top-down. In bottom-up, proteins, or complex protein mixtures, are subjected to enzymatic hydrolysis with the use of a proteolytic enzyme (most frequently trypsin) and the peptide mixture is analyzed by MS. In top-down proteomics, intact proteins or large protein fragments are directly analysed and fragmented for MS analysis [51]. Post-translational modifications (PTMs) or diagenetically induced chemical modifications of ancient proteins are part of what makes palaeoproteomics so much challenging. Chemical modifications include both those derived from biological and diagenetic pathways. Biological PTMs (i.e., methylation, di-methylation, alkylation, hydroxylation, fucosylation, phosphorylation, to name a few) are related to physiology of the specific organism and might reveal evolutionary adaptations of the organism. On the other hand, diagenetically derived modifications (i.e., deamidation, oxidation, carboxymethylation) are the result of post-mortem decay and indicate how proteins degrade over geological time. The information that can be obtained in the study of protein chemical modifications helps, firstly, to obtain a more confident protein identification, often also increasing the protein sequence coverage. Most importantly, however, it

provides information regarding protein degradation that can be precious for palaeoenvironmental studies or characterizing protein diagenetic alteration over time, including a possible (relative) quantification to discriminate potential contaminating from endogenous proteins.

Among the specific modifications of collagen, a high number of hydroxyproline and hydroxylysine are observed in fibrillar collagens. Most importantly, more confident identifications are commonly obtained for these proteins by taking into account these extensive post-translational modifications when collagen is present [52]. Biochemical analysis has indicated that the extent of lysine hydroxylation of collagen is highly variable in comparison with proline hydroxylation[53][54]. Approximately 50% of proline residues in collagen are hydroxylated in various genetic types of collagen in different tissues, and it does not significantly change under physiological conditions. The extent of lysine hydroxylation, however, can vary from 15 to 90%, depending on the collagen types. In addition, within type I collagen, it varies significantly from tissue to tissue and under the physiological/pathological condition of the tissue [53][55], highlighting the variability at which hydroxylation occurs within living tissues. Hydroxylation is restricted to collagens *in vivo*, but given the longevity of these proteins within bone and dentine, this biological PTM can also be observed and analyzed over long time periods.

Similarly, enamel proteomes are characterized by extensive, proteome-wide phosphorylation. This regular PTM receives significant attention in clinical proteome studies of modern proteomes and has recently been observed across enamel proteomes of 2 million years old [56]. There is, therefore, an opportunity to study protein modification in a significant component of the mammalian skeleton for extended periods of time.

The most frequently observed chemical modification is the oxidation of methionine (mass shift of +16 Da) that can occur both *in vivo* and *in vitro* as well as during sample manipulation and is often checked during routine protein identification searches. Moreover, recently, a particular focus was given to deamidation that has been described as a molecular clock that regulates the timing of *in vivo* processes. Deamidation is a spontaneous chemical modification that plays a vital role in protein degradation [57]. It occurs at Asn, and Gln amino acids resulting, respectively, in aspartic/isoaspartic acid and glutamic/ γ -glutamic acid, and it represents a potential aging marker (Figure 12)[20]. Conversions of Asn and Gln to aspartic and glutamic acids can be detected by MS techniques and MS/MS methods, despite the fact that the modification results in a small mass shift of only 0.984 Da. Deamidation analysis has, therefore, become an interesting modification to be looked at in the characterization of aging in proteins from fossils to works of art [58–62].

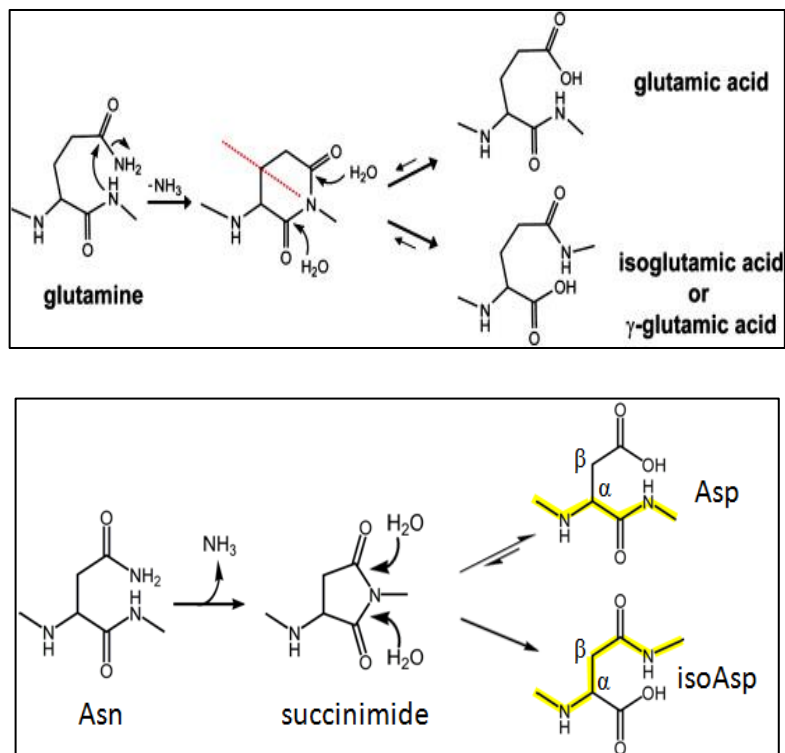


Figure 12: Schematic representation of deamidation of Asn and Gln [63,64]

References

- [1] What is Cultural Heritage - Culture in Development, (n.d.). http://www.cultureindevelopment.nl/Cultural_Heritage/What_is_Cultural_Heritage (accessed September 9, 2020).
- [2] ICOMOS World Report 2002-2003 on monuments and sites in danger - International Council on Monuments and Sites, (n.d.). <https://www.icomos.org/en/what-we-do/risk-preparedness/heritage-at-risk-reports/116-english-categories/resources/publications/210-icomos-world-report-2002-2003-on-monuments-and-sites-in-danger> (accessed September 9, 2020).
- [3] R. Wiesinger, M. Schreiner, Chemistry for cultural heritage, *Herit. Sci.* 3 (2015) 35. <https://doi.org/10.1186/s40494-015-0064-z>.
- [4] J. Hendy, C. Warinner, A. Bouwman, M.J. Collins, S. Fiddyment, R. Fischer, R. Hagan, C.A. Hofman, M. Holst, E. Chaves, L. Klaus, G. Larson, M. Mackie, K. McGrath, A.Z. Mundorff, A. Radini, H. Rao, C. Trachsel, I.M. Velsko, C.F. Speller, Proteomic evidence of dietary sources in ancient dental calculus, *Proc. R. Soc. B Biol. Sci.* 285 (2018) 20180977. <https://doi.org/10.1098/rspb.2018.0977>.
- [5] R. Vinciguerra, A. Illiano, A. De Chiaro, A. Carpentieri, A. Lluveras-Tenorio, I. Bonaduce, G. Marino, P. Pucci, A. Amoresano, L. Birolo, Identification of proteinaceous binders in paintings: A targeted proteomic approach for cultural heritage, *Microchem. J.* 144 (2019) 319–328. <https://doi.org/10.1016/j.microc.2018.09.021>.
- [6] S. Orsini, A. Yadav, M. Dilillo, L.A. McDonnell, I. Bonaduce, Characterization of Degraded Proteins in Paintings Using Bottom-Up Proteomic Approaches: New Strategies for Protein Digestion and Analysis of Data, *Anal. Chem.* 90 (2018) 6403–6408. <https://doi.org/10.1021/acs.analchem.8b00281>.
- [7] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: From bones to paintings ☆, *Microchem. J.* 126 (2016) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.
- [8] P. Cennamo, M. Rosaria, B. Lumaga, G. Fatigati, A. Amoresano, A. Carpentieri, CONSERVATION SCIENCE A MULTIDISCIPLINARY ASSESSMENT TO INVESTIGATE A XXII, 11 (2020) 25–38.
- [9] M. Regert, C. Rolando, Identification of archaeological adhesives using direct inlet electron ionization mass spectrometry, *Anal. Chem.* 74 (2002) 965–975. <https://doi.org/10.1021/ac0155862>.
- [10] M.P. Colombini, G. Giachi, F. Modugno, E. Ribechini, Characterisation of organic residues in pottery vessels of the Roman age from Antinoe (Egypt), *Microchem. J.* 79 (2005) 83–90. <https://doi.org/10.1016/j.microc.2004.05.004>.
- [11] J. Mazurek, M. Svoboda, M. Schilling, GC/MS Characterization of Beeswax, Protein, Gum, Resin, and Oil in Romano-Egyptian Paintings, *Heritage.* 2 (2019) 1960–1985. <https://doi.org/10.3390/heritage2030119>.
- [12] A. Lluveras-Tenorio, J. Mazurek, A. Restivo, M.P. Colombini, I. Bonaduce, Analysis of plant gums and saccharide materials in paint samples: Comparison of GC-MS analytical procedures and databases, *Chem. Cent. J.* 6 (2012) 115. <https://doi.org/10.1186/1752-153X-6-115>.
- [13] S. Walston, The Preservation And Conservation Of Aboriginal And Pacific Cultural Material In Australian Museums, *ICCM Bull.* 4 (1978) 9–21. <https://doi.org/10.1179/iccm.1978.4.4.002>.
- [14] Archaeological Chemistry (RSC Publishing) A Mark Pollard, Carl Heron, (n.d.). <https://pubs.rsc.org/en/content/ebook/978-0-85404-262-3> (accessed October 12, 2020).
- [15] Mazow, Grieve, Kennedy, Contamination in Organic Residue Analysis, *J. East. Mediterr. Archaeol. Herit. Stud.* 2 (2014) 90. <https://doi.org/10.5325/jeasmedarcherstu.2.2.0090>.
- [16] S. Pan, R. Aebersold, R. Chen, J. Rush, D.R. Goodlett, M.W. McIntosh, J. Zhang, T.A. Brentnall, Mass spectrometry based targeted protein quantification: Methods and applications, *J. Proteome Res.* 8 (2009) 787–797. <https://doi.org/10.1021/pr800538n>.
- [17] G. Giorgi, Overview of Mass Spectrometric Based Techniques Applied in the Cultural Heritage Field,

- in: *Org. Mass Spectrom. Art Archaeol.*, John Wiley & Sons, Ltd, 2009: pp. 37–74. <https://doi.org/10.1002/9780470741917.ch2>.
- [18] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science* (80-.). 246 (1989) 64–71. <https://doi.org/10.1126/science.2675315>.
- [19] G. Zhang, R.S. Annan, S.A. Carr, T.A. Neubert, Overview of peptide and protein analysis by mass spectrometry, *Curr. Protoc. Protein Sci. Chapter 16* (2010). <https://doi.org/10.1002/0471140864.ps1601s62>.
- [20] P.R. Graves, T.A.J. Haystead, *Molecular Biologist’s Guide to Proteomics*, *Microbiol. Mol. Biol. Rev.* 66 (2002) 39–63. <https://doi.org/10.1128/mnbr.66.1.39-63.2002>.
- [21] *Principles of Proteomics* - Richard Twyman - Google Libri, (n.d.). <https://books.google.it/books?id=a9uZFP7mvbYC&printsec=frontcover&dq=Twyman+R.,+Principles+of+Proteomics&hl=it&sa=X&ved=2ahUKEwj08ouA9aLtAhVInKQKHQdBB4MQ6wEwAHoECAAAQ#v=onepage&q=Twyman+R.%2C+Principles+of+Proteomics&f=false> (accessed November 27, 2020).
- [22] T.F.O.R. Biochemistry, *Mass Spectrometry and Protein Analysis*, 312 (2006) 212–217.
- [23] I. V. Chernushevich, A. V. Loboda, B.A. Thomson, An introduction to quadrupole-time-of-flight mass spectrometry, *J. Mass Spectrom.* 36 (2001) 849–865. <https://doi.org/10.1002/jms.207>.
- [24] C. Cushman, B.M. Lunt, N.J. Smith, C. Incorporated, M.R. Linford, Sample Charging in ToF-SIMS: How it Affects the Data that are Collected and How to Reduce it, *Vac. Technol. Coat.* (2018) 1–7.
- [25] E.S. Hecht, M. Scigelova, S. Eliuk, A. Makarov, *Fundamentals and Advances of Orbitrap Mass Spectrometry*, in: *Encycl. Anal. Chem.*, Wiley, 2019: pp. 1–40. <https://doi.org/10.1002/9780470027318.a9309.pub2>.
- [26] R.A. Zubarev, A. Makarov, Orbitrap mass spectrometry, *Anal. Chem.* 85 (2013) 5288–5296. <https://doi.org/10.1021/ac4001223>.
- [27] J. V. Johnson, R.A. Yost, P.E. Kelley, D.C. Bradford, Tandem-in-Space and Tandem-in-Time Mass Spectrometry: Triple Quadrupoles and Quadrupole Ion Traps, *Anal. Chem.* 62 (1990) 2162–2172. <https://doi.org/10.1021/ac00219a003>.
- [28] P. Roepstorff, J. Fohlman, Letter to the editors, *Biol. Mass Spectrom.* 11 (1984) 601–601. <https://doi.org/10.1002/bms.1200111109>.
- [29] H. Steen, M. Mann, The ABC’s (and XYZ’s) of peptide sequencing, *Nat. Rev. Mol. Cell Biol.* 5 (2004) 699–711. <https://doi.org/10.1038/nrm1468>.
- [30] I.K. Chu, C.K. Siu, J.K.C. Lau, W.K. Tang, X. Mu, C.K. Lai, X. Guo, X. Wang, N. Li, Y. Xia, X. Kong, H. Bin Oh, V. Ryzhov, F. Tureček, A.C. Hopkinson, K.W.M. Siu, Proposed nomenclature for peptide ion fragmentation, *Int. J. Mass Spectrom.* 390 (2015) 24–27. <https://doi.org/10.1016/j.ijms.2015.07.021>.
- [31] T. Koenig, B.H. Menze, M. Kirchner, F. Monigatti, K.C. Parker, T. Patterson, J.J. Steen, F.A. Hamprecht, H. Steen, Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics, *J. Proteome Res.* 7 (2008) 3708–3717. <https://doi.org/10.1021/pr700859x>.
- [32] J. Zhang, L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G.A. Lajoie, B. Ma, PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification, *Mol. Cell. Proteomics.* 11 (2012). <https://doi.org/10.1074/mcp.M111.010587>.
- [33] D. Li, Y. Fu, R. Sun, C.X. Ling, Y. Wei, H. Zhou, R. Zeng, Q. Yang, S. He, W. Gao, pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry, *Bioinformatics.* 21 (2005) 3049–3050. <https://doi.org/10.1093/bioinformatics/bti439>.
- [34] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372. <https://doi.org/10.1038/nbt.1511>.
- [35] S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, *Nat. Protoc.* 11 (2016) 2301–2319. <https://doi.org/10.1038/nprot.2016.136>.
- [36] M. Brosch, L. Yu, T. Hubbard, J. Choudhary, Accurate and sensitive peptide identification with mascot

- percolator, *J. Proteome Res.* 8 (2009) 3176–3181. <https://doi.org/10.1021/pr800982s>.
- [37] M.H. Gordon, Principles of gas chromatography, in: *Princ. Appl. Gas Chromatogr. Food Anal.*, Springer US, 1990: pp. 11–58. https://doi.org/10.1007/978-1-4613-0681-8_1.
- [38] E.R. Adlard, R.S. Juvet, A Review of Detectors for Gas Chromatography Part I: Universal Detectors, *Crit. Rev. Anal. Chem.* 5 (1975) 1–11. <https://doi.org/10.1080/10408347508085684>.
- [39] An Introduction to Raman Spectroscopy: Introduction and Basic Principles - 2014 - Wiley Analytical Science, (n.d.). <https://analyticalscience.wiley.com/do/10.1002/sepspec.1882education/full/> (accessed November 27, 2020).
- [40] G.S. Bumbrah, R.M. Sharma, Raman spectroscopy – Basic principle, instrumentation and selected applications for the characterization of drugs of abuse, *Egypt. J. Forensic Sci.* 6 (2016) 209–215. <https://doi.org/10.1016/j.ejfs.2015.06.001>.
- [41] V.I. Mikla, J.M. Turovci, V. V. Mikla, N. Mehta, Molecular structure of Se-rich amorphous films, *Prog. Solid State Chem.* 49 (2018) 1–15. <https://doi.org/10.1016/j.progsolidstchem.2017.10.001>.
- [42] M.A. Ganzoury, N.K. Allam, T. Nicolet, C. All, Introduction to Fourier Transform Infrared Spectrometry, *Renew. Sustain. Energy Rev.* 50 (2015) 1–8. <https://doi.org/10.1016/j.rser.2015.05.073>.
- [43] Basic Structure and Principle of Microscopes | KEYENCE biological fluorescence microscopes, (n.d.). <https://www.keyence.com/ss/products/microscope/bz-x/study/principle/structure.jsp> (accessed November 27, 2020).
- [44] Light Microscope- definition, principle, types, parts, magnification, (n.d.). <https://microbenotes.com/light-microscope/> (accessed November 27, 2020).
- [45] E. Cappellini, A. Prohaska, F. Racimo, F. Welker, M.W. Pedersen, M.E. Allentoft, P. De Barros Damgaard, P. Gutenbrunner, J. Dunne, S. Hamman, M. Roffet-Salque, M. Ilardo, J.V. Moreno-Mayar, Y. Wang, M. Sikora, L. Vinner, J. Cox, R.P. Evershed, E. Willerslev, Ancient Biomolecules and Evolutionary Inference, *Annu. Rev. Biochem.* 87 (2018) 1029–1060. <https://doi.org/10.1146/annurev-biochem-062917-012002>.
- [46] M. Buckley, R.A. Fariña, C. Lawless, P.S. Tambusso, L. Varela, A.A. Carlini, J.E. Powell, J.G. Martinez, Collagen Sequence Analysis of the Extinct Giant Ground Sloths *Lestodon* and *Megatherium*, *PLoS One.* 10 (2015) e0139611. <https://doi.org/10.1371/journal.pone.0139611>.
- [47] R.C. Hill, M.J. Wither, T. Nemkov, A. Barrett, A. D’Alessandro, M. Dzieciatkowska, K.C. Hansen, Preserved proteins from extinct bison *latifrons* identified by tandem mass spectrometry; Hydroxylysine glycosides are a common feature of ancient collagen, *Mol. Cell. Proteomics.* 14 (2015) 1946–1958. <https://doi.org/10.1074/mcp.M114.047787>.
- [48] P.H. Ostrom, M. Schall, H. Gandhi, T.L. Shen, P. V. Hauschka, J.R. Strahler, D.A. Gage, New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry, *Geochim. Cosmochim. Acta.* 64 (2000) 1043–1050. [https://doi.org/10.1016/S0016-7037\(99\)00381-6](https://doi.org/10.1016/S0016-7037(99)00381-6).
- [49] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A.R. Da Fonseca, T.W. Stafford, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, M.T.P. Gilbert, J. V. Olsen, Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, *J. Proteome Res.* 11 (2012) 917–926. <https://doi.org/10.1021/pr200721u>.
- [50] T.-N.-N. Tran, G. Aboudharam, A. Gardeisen, B. Davoust, J.-P. Bocquet-Appel, C. Flaudrops, M. Belghazi, D. Raoult, M. Drancourt, Classification of Ancient Mammal Individuals Using Dental Pulp MALDI-TOF MS Peptide Profiling, *PLoS One.* 6 (2011) e17319. <https://doi.org/10.1371/journal.pone.0017319>.
- [51] Z.R. Gregorich, Y.H. Chang, Y. Ge, Proteomics in heart failure: Top-down or bottom-up?, *Pflugers Arch. Eur. J. Physiol.* 466 (2014) 1199–1209. <https://doi.org/10.1007/s00424-014-1471-9>.
- [52] W.J. Marshall, M. Lapsley, A.P. Day, R.M. Ayling, Clinical biochemistry: metabolic and clinical aspects, n.d. <https://www.sciencedirect.com/book/9780702051401/clinical-biochemistry-metabolic-and-clinical-aspects> (accessed March 18, 2019).
- [53] K. Uzawa, H.N. Yeowell, K. Yamamoto, Y. Mochida, H. Tanzawa, M. Yamauchi, Lysine hydroxylation of

collagen in a fibroblast cell culture system, *Biochem. Biophys. Res. Commun.* 305 (2003) 484–487. [https://doi.org/10.1016/S0006-291X\(03\)00799-X](https://doi.org/10.1016/S0006-291X(03)00799-X).

[54] M. Yamauchi, M. Sricholpech, Lysine post-translational modifications of collagen., *Essays Biochem.* 52 (2012) 113–33. <https://doi.org/10.1042/bse0520113>.

[55] E.J. Miller, The structure of fibril-forming collagens., *Ann. N. Y. Acad. Sci.* 460 (1985) 1–13. <http://www.ncbi.nlm.nih.gov/pubmed/3868937> (accessed March 19, 2019).

[56] C. Walsh, *Posttranslational modification of proteins : expanding nature’s inventory*, Roberts and Co. Publishers, 2006. <http://www.roberts-publishers.com/walsh/chapter5.pdf> (accessed October 14, 2016).

[57] N.L. van Doorn, J. Wilson, H. Hollund, M. Soressi, M.J. Collins, Site-specific deamidation of glutamine: a new marker of bone collagen deterioration, *Rapid Commun. Mass Spectrom.* 26 (2012) 2319–2327. <https://doi.org/10.1002/rcm.6351>.

[58] T.P. Cleland, E.R. Schroeter, M.H. Schweitzer, Biologically and diagenetically derived peptide modifications in moa collagens, *Proc. R. Soc. B Biol. Sci.* 282 (2015) 20150015–20150015. <https://doi.org/10.1098/rspb.2015.0015>.

[59] M. Mackie, P. R  ther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Pegg  , H. Howard, L. Harrison, L.J. Jensen, J. V. Olsen, E. Cappellini, Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting, *Angew. Chemie - Int. Ed.* 57 (2018) 7369–7374. <https://doi.org/10.1002/anie.201713020>.

[60] R. Sawafuji, E. Cappellini, T. Nagaoka, A.K. Fotakis, R. Rakownikow Jersie-Christensen, J. V Olsen, K. Hirata, S. Ueda, Proteomic profiling of archaeological human bone, (n.d.). <https://doi.org/10.1098/rsos.161004>.

[61] F. Welker, M. Hajdinjak, S. Talamo, K. Jaouen, M. Dannemann, F. David, M. Julien, M. Meyer, J. Kelso, I. Barnes, S. Brace, P. Kamminga, R. Fischer, B.M. Kessler, J.R. Stewart, S. P  abo, M.J. Collins, J.J. Hublin, Palaeoproteomic evidence identifies archaic hominins associated with the Ch  telperronian at the Grotte du Renne, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 11162–11167. <https://doi.org/10.1073/pnas.1605834113>.

[62] J. Wilson, N.L. Van Doorn, M.J. Collins, Assessing the extent of bone degradation using glutamine deamidation in collagen, *Anal. Chem.* 84 (2012) 9041–9048. <https://doi.org/10.1021/ac301333t>.

[63] L. Jia, Y. Sun, Protein asparagine deamidation prediction based on structures with machine learning methods, *PLoS One.* 12 (2017). <https://doi.org/10.1371/journal.pone.0181347>.

[64] E.R. Schroeter, T.P. Cleland, Glutamine deamidation: An indicator of antiquity, or preservational quality?, *Rapid Commun. Mass Spectrom.* 30 (2016) 251–255. <https://doi.org/10.1002/rcm.7445>.

2. Aim of the thesis

This Ph.D., was part of the EU funded European training network *Tempera* and was focused on the development and application of proteomic strategies for the molecular characterization of archaeological objects and works of art with the use of mass spectrometry. Moreover, this work was also focused on the study of the degradation processes that the proteins have undergone over the years in an attempt to discover molecular signatures of their conservation state. Taking as a starting point the study of collagen, one of the most persistent bone proteins in the fossil record, proteomic bottom-up approaches regarding protein extraction and discovery of chemical modifications with the use of several bioinformatics tools were developed. Once the proteomic protocols had been established, we complemented our knowledge in search of other types of organic molecules such as lipids and sugars by mass spectrometry. In this way, we managed to approach each case of the study by several perspectives, leading to more integrated results.

Chapter 3 describes the development of proteomic strategies for the characterization of collagen-based materials. Initially, as a small introduction to the chapter, a chemical review on primary collagen sources such as bones and teeth reports the potentiality of ancient skeletal tissues for palaeoproteomic studies. Subchapter 3.1 deals with the analysis of ancient human bones from the excavation sites in the Vesuvius area. In particular, a bottom-up proteomic approach was applied, and screening of the obtained MS data was performed in search for molecular signatures imprinted in the protein component of bones. Moreover, a screening of the MS and MS/MS data was performed to select the most frequently detected peptides that can be used for a quantification of deamidation (Asn, Gln) and oxidation (Met). Furthermore, at the end of this chapter, a wide-broad protocol for the extraction of lipids, sugars, and proteins was applied for molecular characterization of animal glues used for restoration purposes. The analytical methodology includes the use of gas and liquid chromatography-mass spectrometry.

Chapter 4 was focused on the development of micro-invasive techniques for the characterization of proteins in archaeological objects and works of art. More specifically, innovative micro-invasive kits (subchapters 4.1 & 4.2) were developed with the immobilization of trypsin taking advantage of the adhesive properties of a specific type of hydrophobins, known as Vmh2. The kits have already been applied to several types of archaeological objects, and the obtained results have been published. Subchapter 4.3 reports some improvements and future perspectives in the development of other kits by using a different type of polymer-based materials.

In **Chapter 5**, a combination of mass spectrometric with microscopic and spectroscopic techniques was used for a more in-depth characterization of archaeological amphoraes and historical paintings. More specifically, in-depth characterization of the pigments in the surface layer or in the several layers of the archaeological paintings was performed with the use of microscopy or Raman spectroscopy. The use of organic binders such as natural gums waxes and oils or proteinaceous binders such as collagen, egg etc was confirmed with the use of mass spectrometric approaches. In this way, a more integrated approach revealed important information regarding the painting techniques in the case of historical paintings, whereas, in the case of amphoraes, spectroscopic and spectrometric techniques demonstrated the content and their possible use. Also, important information regarding the degradation processes undergone over the years was demonstrated too. A summary of all projects performed in this Ph.D. is reported below based on the scientific approach. Also, a list of the publications as outcome of the single projects is also reported.

Palaeoproteomic projects performed during this PhD

3. Proteomic strategies for the characterization of collagen-based materials

3.1. Chemical review on Paleoproteomic analysis of bones and teeth

3.2. Written in bone proteins: molecular signatures in the 79 AD skeletal remains from Herculaneum and Pompeii.

3.3. Selection of peptide biomarkers in human bone collagen for the evaluation of deamidation (N, Q and oxidation (M), by targeted proteomics.

3.4. Molecular characterization of animal glues for restoration purposes

4. Development of micro-invasive techniques for protein identification and characterization in the field of cultural heritage

4.1 Minimally Invasive and Portable Method for the identification of Proteins in Ancient Paintings

4.2 A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects.

4.3 Development and future perspective for a micro-invasive tool for the analysis of proteins in historical objects.

5. Combination of mass spectrometric, spectroscopic and microscopic techniques for the molecular characterization of archaeological remains and works of art

- 5.1. Novel use of Evolved Gas Analysis/Mass Spectrometry to identify the earliest organic binding medium in Aegean wall paintings from Tel Kabri, Israel, dated to the 18th C. B.C.E
- 5.2. A multidisciplinary campaign for the characterization of a XXII dynasty wooden Sarcophagus (yellow coffin).
- 5.3. Organic remains in amphorae from the temple of Hera in Paestum shed lights on solemn ceremonies.
- 5.4. Molecular characterization of wall paintings from Roman Domus in Santa Maria Capua Vetere
- 5.5. Molecular characterization of protective layers in statues from the monumental staircase in the Royal Palace of Caserta

List of publications

Chapter 3

- 3.1. Palaeoproteomic analysis of bone and teeth, Ntasi G, Welker F, a chapter of the upcoming book: Proteomic analysis in cultural heritage, chemical review under submission, Springer Nature
- 3.2. Written in bone proteins: molecular signatures in the 79 AD skeletal remains from Herculaneum and Pompeii; Ntasi G, Palomo I, R, Cappellini E, Birolo L, Petrone P.P. under submission
- 3.3. Selection of peptide biomarkers in human bone collagen for the evaluation of deamidation (N, Q and oxidation (M), by targeted proteomics; Ntasi G, Palomo I, R, De Chiaro A, Marino G, Petrone P.P, Birolo L ,under submission
- 3.4. Molecular Characterization of animal glues for restoration purposes;. Ntasi G; Bonaduce I, Sbriglia S, Marino G, Birolo L, under submission

Chapter 4

4.1. Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings, Ntasi G, Cicatiello P, Rossi M, Marino G, Giardina P, Birolo L, *Anal. Chem.* 90 (2018) 10128–10133. <https://doi.org/10.1021/acs.analchem.8b01718>.

4.2. A versatile and user-friendly approach for the analysis of ancient proteins. Ntasi G, Kirby D, Stanzione I, Carpentieri A, Somma P, Cicatiello P, Marino G, Giardina P, Birolo L. *J. Proteomics.* 231 (2021). <https://doi.org/10.1016/j.jprot.2020.104039>.

Chapter 5

5.1. Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings. (Linn R, Bonaduce I, Ntasi G, Birolo L, Yasur-Landau A, Cline EH, Nevin A, Lluveras-Tenorio A) *Angew Chem Int Ed Engl. Oct Ed.* 57 (2018) 13257–13260. <https://doi.org/10.1002/anie.201806520>.

5.2. A multidisciplinary assessment to investigate an XXII dynasty wooden coffin, Melchiorre C, Iorio L, Ntasi G, Birolo L, Trojsi G, Cennamo P, Barone MR, Fatigati G, Amoresano A, Carpentieri A, *International journal of conservation science*, Volume 11, Issue 1, January-March 2020: 25-38. ISSN:2067-533X

5.3. Organic remains in amphorae from the temple of Hera in Paestum shed some light on solemn ceremonies; Vergara A, Ntasi G P, Birolo L, manuscript in preparation.

3. Proteomics strategies for the characterization of collagen-based materials

Introduction

Collagen is one of the most resistant proteins in the fossil record due to its mechanical, structural and chemical properties. Its longevity lies on the Gly-X-Y repetitive sequence and a uniquely high content of Pro and Hyp that make it quite stable [1,2]. For this reason, most of the paleoproteomic studies are based on collagen identification and characterization of its diagenetic alteration over the years. Collagen identification in archeological bones and teeth has contributed significantly in the fields of evolutionary biology explaining phylogenetic relationships of human and animal species [3–10]. Animal glues consist of another collagen source that was used in the past as proteinaceous binders in paintings [11] and therefore its characterization has contributed both to the history of the artworks as also in the restoration processes regarding their preservation to the next generations [12–17]. Due to the natural or diagenetic alteration over the years the characterization of collagen chemical modifications is crucial. In this respect, deamidation can constitute an important modification to start with, since it is now widely accepted as an important biomolecular marker of the deterioration and natural aging of proteins in artistic and archaeological materials [18–20]. Moreover, the effect of aging, environmental and storage conditions at the molecular level can be analysed and studied looking at chemical modifications induced in proteins, possibly leading to the discovery of other potential biomarkers to be searched for in ancient samples.

Aim of the study

The results have been reported in the chapter of a book on palaeoproteomics and three manuscripts, currently in preparations.

Book chapter:

Chemical review on Palaeoproteomic analysis of bones and teeth. Georgia Ntasi, Frido Welker

Manuscripts:

- Written in bone proteins: molecular signatures in the 79 AD skeletal remains from Herculaneum and Pompeii. Georgia Ntasi, Ismael Rodriguez Palomo, Gennaro Marino, Francesco Sirano, Enrico Cappellini, Leila Birolo, and Pierpaolo Petrone. Manuscript in preparation.

- Selection of peptide biomarkers in human bone collagen for the evaluation of deamidation (N, Q) and oxidation (M) by targeted proteomics. Georgia Ntasi, Ismael Rodriguez Palomo, Addolorata De Chiaro, Gennaro Marino, Paolo Petrone and Leila Birolo. Manuscript in preparation.

- Molecular characterization of animal glues for restoration purposes. Georgia Ntasi, Sara Sbriglia, Rosanna Pitocchi, Chiara Melchiorre, Laura dello Ioio, Giancarlo Fatigati, Ilaria Bonaduce, Andrea Carpentieri, Gennaro Marino and Leila Birolo. Manuscript in preparation

References

- [1] M.J. Collins, J. Hiller, C.I. Smith, J.P. Roberts, R. V Prigodich, T.J. Wess, A.R. Millard, the Survival of Organic Matter in Bone: a Review, *Archaeometry*. 3 (2002) 383–394.
- [2] A.L. Boskey, R. Coleman, Aging and Bone, 89 (2010) 1333–1348. <https://doi.org/10.1177/0022034510377791>.
- [3] J. Hendy, C. Warinner, A. Bouwman, M.J. Collins, S. Fiddyment, R. Fischer, R. Hagan, C.A. Hofman, M. Holst, E. Chaves, L. Klaus, G. Larson, M. Mackie, K. McGrath, A.Z. Mundorff, A. Radini, H. Rao, C. Trachsel, I.M. Velsko, C.F. Speller, Proteomic evidence of dietary sources in ancient dental calculus, *Proc. R. Soc. B Biol. Sci.* 285 (2018) 20180977. <https://doi.org/10.1098/rspb.2018.0977>.
- [4] S. Charlton, A. Ramsøe, M. Collins, O.E. Craig, R. Fischer, M. Alexander, C.F. Speller, New insights into Neolithic milk consumption through proteomic analysis of dental calculus, *Archaeol. Anthropol. Sci.* 11 (2019) 6183–6196. <https://doi.org/10.1007/s12520-019-00911-7>.
- [5] M. Mackie, J. Hendy, A.D. Lowe, A. Sperduti, M. Holst, M.J. Collins, C.F. Speller, Preservation of the metaproteome: variability of protein preservation in ancient dental calculus, *Sci. Technol. Archaeol. Res.* 3 (2017) 58–70. <https://doi.org/10.1080/20548923.2017.1361629>.
- [6] T.P. Cleland, E.R. Schroeter, R.S. Feranec, D. Vashishth, Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics, *Proc. R. Soc. B Biol. Sci.* 283 (2016). <https://doi.org/10.1098/rspb.2016.0593>.
- [7] F. Welker, M. Hajdinjak, S. Talamo, K. Jaouen, M. Dannemann, F. David, M. Julien, M. Meyer, J. Kelso, I. Barnes, S. Brace, P. Kamminga, R. Fischer, B.M. Kessler, J.R. Stewart, S. Pääbo, M.J. Collins, J.J. Hublin, Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 11162–11167. <https://doi.org/10.1073/pnas.1605834113>.
- [8] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J. V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.-J. Hublin, L. Orlando, I. Barnes, R.D.E. MacPhee, Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, *Nature*. 522 (2015) 81–84. <https://doi.org/10.1038/nature14249>.
- [9] F. Welker, Palaeoproteomic analysis of Early Pleistocene *Gigantopithecus blacki*, *Am. J. Phys. Anthropol.* 168 (2019).
- [10] E. Cappellini, F. Welker, L. Pandolfi, J. Ramos-Madrugal, D. Samodova, P.L. Rùther, A.K. Fotakis, D. Lyon, J.V. Moreno-Mayar, M. Bukhsianidze, R. Rakownikow, Jersie-Christensen, M. Mackie, A. Ginolhac, R. Ferring, M. Tappen, E. Palkopoulou, M.R. Dickinson, T.W. Stafford, Y.L. Chan, A. Götherström, S.K.S.S. Nathan, P.D. Heintzman, J.D. Kapp, I. Kirillova, Y. Moodley, J. Agusti, R.D. Kahlke, G. Kiladze, B. Martínez-Navarro, S. Liu, M. Sandoval Velasco, M.H.S. Sinding, C.D. Kelstrup, M.E. Allentoft, L. Orlando, K. Penkman, B. Shapiro, L. Rook, L. Dalén, M.T.P. Gilbert, J. V. Olsen, D. Lordkipanidze, E. Willerslev, Early Pleistocene enamel

proteome from Dmanisi resolves *Stephanorhinus* phylogeny, *Nature*. 574 (2019) 103–107. <https://doi.org/10.1038/s41586-019-1555-y>.

[11] D.P. Kirby, M. Buckley, E. Promise, S.A. Trauger, T.R. Holdcraft, Identification of collagen-based materials in cultural heritage, *Analyst*. 138 (2013) 4849–4858. <https://doi.org/10.1039/c3an00925d>.

[12] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: From bones to paintings ☆, *Microchem. J.* 126 (2016) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.

[13] R. Linn, I. Bonaduce, G. Ntasi, L. Birolo, A. Yasur-Landau, E.H. Cline, A. Nevin, A. Lluveras-Tenorio, Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings, *Angew. Chemie - Int. Ed.* 57 (2018). <https://doi.org/10.1002/anie.201806520>.

[14] P. Cicatiello, G. Ntasi, M. Rossi, G. Marino, P. Giardina, L. Birolo, Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings, *Anal. Chem.* 90 (2018) 10128–10133. <https://doi.org/10.1021/acs.analchem.8b01718>.

[15] R. Vinciguerra, A. Illiano, A. De Chiaro, A. Carpentieri, A. Lluveras-Tenorio, I. Bonaduce, G. Marino, P. Pucci, A. Amoresano, L. Birolo, Identification of proteinaceous binders in paintings: A targeted proteomic approach for cultural heritage, *Microchem. J.* 144 (2019) 319–328. <https://doi.org/10.1016/j.microc.2018.09.021>.

[16] S. Orsini, A. Yadav, M. Dilillo, L.A. McDonnell, I. Bonaduce, Characterization of Degraded Proteins in Paintings Using Bottom-Up Proteomic Approaches: New Strategies for Protein Digestion and Analysis of Data, *Anal. Chem.* 90 (2018) 6403–6408. <https://doi.org/10.1021/acs.analchem.8b00281>.

[17] M. Mackie, P. Rüther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Peggie, H. Howard, L. Harrison, L.J. Jensen, J. V. Olsen, E. Cappellini, Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting, *Angew. Chemie - Int. Ed.* 57 (2018) 7369–7374. <https://doi.org/10.1002/anie.201713020>.

[18] G. Leo, I. Bonaduce, A. Andreotti, G. Marino, P. Pucci, M.P. Colombini, L. Birolo, Deamidation at asparagine and glutamine As a major modification upon deterioration/aging of proteinaceous binders in mural paintings, *Anal. Chem.* 83 (2011) 2056–2064. <https://doi.org/10.1021/ac1027275>.

[19] E.R. Schroeter, T.P. Cleland, Glutamine deamidation: An indicator of antiquity, or preservational quality?, *Rapid Commun. Mass Spectrom.* 30 (2016) 251–255. <https://doi.org/10.1002/rcm.7445>.

[20] N.L. van Doorn, J. Wilson, H. Hollund, M. Soressi, M.J. Collins, Site-specific deamidation of glutamine: a new marker of bone collagen deterioration, *Rapid Commun. Mass Spectrom.* 26 (2012) 2319–2327. <https://doi.org/10.1002/rcm.6351>.

3.1 Manuscript 1: Chemical review on Palaeoproteomic analysis of bones and teeth

Georgia Ntasi, Frido Welker

Abstract

The high-sensitivity of current mass spectrometry instruments has permitted palaeoproteomics to become relevant in the fields of archaeology and evolutionary biology. Taking advantage of the continuous development of MS techniques, it is now possible to extract ancient skeletal proteins, characterize the preserved proteome, and discover diagenetic or biological derived modifications. This information, by itself, or combined with other kinds of analysis such as DNA analysis, expands the application of palaeoproteomics to paleopathology and phylogenetic and taxonomic studies. Here we illustrate the potential of ancient skeletal tissues for palaeoproteomic studies. We explore the differences in the protein survival, define the difference between *in vivo* and *in vitro* protein chemical modifications, which occur either because of protein diagenetic alteration over time or during the life of the organism. Finally, we will indicate some future challenges for the field of ancient skeletal proteomics that require addressing in the coming years.

Protein survival in ancient bones and dentine

Bones consist of 70% inorganic material, predominantly hydroxyapatite. The remaining 30% is organic material, of which the protein component is largely composed of collagenous proteins (85-90%, primarily collagen type I) and 10-15% of non-collagenous proteins (NCPs). The last category includes serum proteins, proteoglycans, glycosylated proteins (including small integrin-binding ligand, N-linked glycoprotein, so-called SIBLINGS), glycosylated proteins with potential cell-attachment activities, and a variety of γ -carboxylated (gla) proteins such as *MGP*[1].

The inorganic phase of bone gives resistance to compression forces and is composed of hydroxyapatite crystals with a formula often generalized as $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$. Bone collagen is composed largely from collagen type I (COL1). COL1 is a triple helix composed out of three α -chains: two identical $\alpha 1$ (I) chains and one genetically different $\alpha 2$ (I) chain[2]. All fibrillar collagens are synthesized in the form of soluble precursor molecules, called procollagens, with large N- and C-terminal propeptide domains. The C-propeptides are removed during subsequent stages of biosynthesis, leaving the short C-telopeptides. The extent of N-terminal processing depends on the collagen type. For procollagens I, II and III the N-propeptides are completely removed, leaving short N-telopeptides, typically about 20 residues in length at both the N- and C-

termini. In contrast, N-terminal processing of procollagens V and XI leaves large N-terminal domains, which help to modulate fibril formation. The mature triple helical region, characteristic of collagens, is located between these mature telopeptides. For COL1, the mature helical region of the alpha chains contain 338 triplets of amino acids constructed from repeating Gly-Xaa-Aaa triplets, where Aaa and Yaa can be any amino acid but are frequently the amino acids proline and hydroxyproline, respectively[3]. The glycine (Gly) at every third position of each chain is a prerequisite for the folding of the three chains into a tight triple helix. In addition to collagen type I, several other forms of collagen are frequently present in bone as well, like the above-mentioned collagens II, III, V and XI.

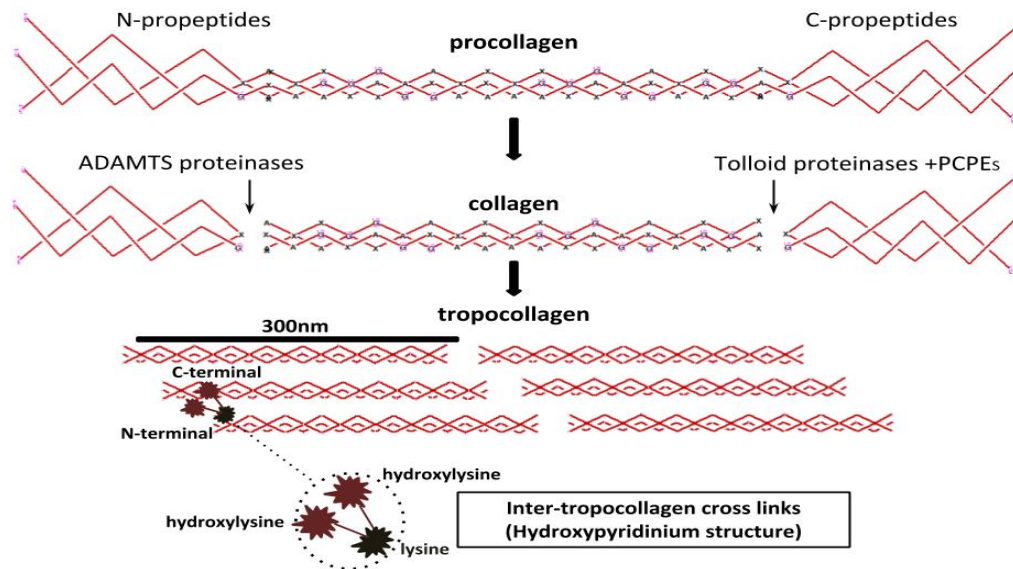


Figure 1: Collagen fibril formation. After transcription and translation, the processing of the N- and C-termini of procollagen lead to spontaneous assembly of collagen fibrils. These are subsequently stabilized by the formation of covalent cross-links.

The bone and dentine proteomes are not solely composed of collagens, however. A minor but highly diverse proteome component is composed of a variety of non-collagenous proteins (NCPs). Serum proteins that compose 25% of the NCPs are exogenously derived and may be related to matrix mineralization or cell proliferation during bone growth. The remaining NCPs are largely synthesized from osteoblasts. Among the glycosylated proteins (leucine rich-repeat), some regulate collagen fibril diameter (decorin, osteonectin), some are related to matrix organization (aggrecan, versican), and others, such as biglycan, interact with collagen type II and type I. Furthermore, SIBLING proteins (osteopontin, bone sialoprotein) are responsible for bone mineralization and remodeling. Glycoproteins with arginyglycylaspartic acid-containing (RGD) cell

attachments motifs are also frequently present (thrombospondin, fibronectin, vitronectin, fibrillin). Finally, γ -carboxylated (gla) proteins (matrix Gla protein, osteocalcin, protein S) may inhibit mineralization through the regulation of osteoclasts. Bone and dentine proteomes are similar, consisting primarily of type I collagen and NCPs. The NCPs are less abundant and for this reason most research has focused on collagen identification. However, dentine contains three unique proteins apparently absent from bone and other tissue: dentine phosphophoryn (DPP), dentine matrix protein 1 (DMP1) and dentine sialoprotein (DSP). However, just a few studies have focused on the analysis of dentine [4]. This might be because bones are the most abundant mineralized tissue in the fossil record, as well the enhanced morphological identifiability of (even highly fragmented) teeth.

Furthermore, phylogenetic studies of ancient proteomes based on collagen sequences have given a deeper insight into the evolutionary relationships of several species [5–7]. In addition, bones have been used to investigate specific cases such as that of the Vesuvius eruption, where in 79 AD two ancient cities, Pompeii and Herculaneum were totally destroyed, within several hours killing, almost instantaneously, their inhabitants [8].

A major focus of ancient bone proteomics is species identification with the use of mass spectrometry since the exploitation of protein and genome databases allows the taxonomic characterization of the species. A classical bone extraction protocol starts with homogenization, sometimes in tandem with sample clean up. Subsequently, a demineralization or decalcification step is normally performed, with the primary aim to remove hydroxyapatite. The rate of demineralization depends mainly on the temperature and concentration of the chelating (complexes) or organic (base or acid) agents that are used. Afterwards, extraction of the released proteins and its denaturation is carried out. Denaturation is a crucial step especially for ancient proteins in order to unfold the quaternary, tertiary structure, and secondary protein structures. There are different mechanisms that explain the protein unfolding but the process is mostly based on ionic or polar interactions between the denaturing agent and the charged part of the protein. Denaturation therefore allows the unfolded primary amino acid protein to be accessible for subsequent enzymatic proteolysis. This is usually performed with the use of trypsin since it is a specific enzyme that digests the C-terminal of lysine and arginine. Finally, before enzymatic hydrolysis, a pre-digestion step is sometimes performed, for example by using rLysC [9,10], to enhance trypsin digestion. The last step of each proteomic protocol is peptide clean up where desalting and protein purification are performed before mass spectrometry analysis. In this step peptide cleanup pipette tips, normally packed with C18 silica, are used for peptide purification and concentration [11]. Table 10.1 summarizes some of the most common bone extraction protocols.

Table 1: Summary of some proteomic bone collagen extraction protocols . All these examples use trypsin as protease.

Step	Treatment	Duration	Temperature
Demineralization	EDTA	2-10 days	4-28 °C
	HCl	1-24 hours	25 °C
Reduction & alkylation	DTT & IAM	2-4 hours	>60 °C
	TCEP & CAA	2-4 hours	>60 °C
Denaturation	GuHCl	30 minutes	25 °C
	Urea	30 minutes	25 °C
	Buffer (mix of Ura, Thiourea, CHAPS, GuHCl)	30 minutes	25 °C
Pre-digestion	rLysC	1-4 hours	37 °C
Digestion	Enzymatic solution	14-16 hours	37 °C

Chemical abbreviations: Ethylenediaminetetraacetic acid (EDTA), Dithiothreitol (DTT), Iodoacetamide (IAM), tris(2-carboxyethyl)phosphine (TCEP), Chloroacetaldehyde (CAA), Guanidine Hydrochloride (GuHCL), 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS).References: [12–17].

Protein survival in ancient enamel

Bone, dentine and enamel are the main mineralized components of the mammalian skeleton. Of these, enamel is the densest tissue. Dentine, bone and enamel have many differences that are present from the beginning of their biological formation. Enamel originates from the ectoderm of the embryo whereas the dentin originates from the ectomesenchyme (Richard Scott and Turner 2000). The ameloblasts (enamel-forming cells) are differentiated from their precursor cells before the odontoblasts (dentin-forming cells) do. The ameloblasts die after the formation of enamel and before eruption of the tooth is completed. Enamel is therefore an acellular tissue for most of its existence, and cannot regenerate or remodel. Both of them also differ in their proteome composition. Enamel is nearly 96% inorganic with the remaining 4% being organic compounds (enamel-specific proteins), and 1% water. Dentin, on the other hand, is 70% inorganic with the remaining 30% being organic matter (20%) and water (10%). The organic portion in dentin is mostly composed of collagen fibers (~18% of the 20%) while the remainder contains non-collagenous proteins, DNA, and lipids[18,19].

Recent research has demonstrated that enamel proteins survive diagenesis over long time scales. Ancient enamel proteomes are relatively similar to their modern counterparts, albeit with higher diagenetic modifications[20–22]. Furthermore amelogenin, one of the major proteins in enamel, allows for proteomic sex identification. The amelogenin gene is located on the X-chromosome (AMELX) and on the Y-chromosome (AMELY). Both copies have acquired a different amino acid sequence over time due to non-recombination of the Y chromosome. Hence, the proteomic detection of AMELY can reliably identify if an individual is male (XY) and its absence might indicate that an individual is a female (XX). This advantage creates a potential for animal and human sex identification studies in ancient samples in the future [23,24].

Palaeoproteomics and phylogeny

The evidence for the evolutionary relationships between extinct and extant species is nearly always incomplete, as the vast majority of species that have ever lived are extinct and relatively few of their remains have been preserved in the fossil record. In addition, evolutionary processes shaping the skeletal skeleton, such as convergence and size-induced allometry, have the potential to mask true phylogenetic relationships. Therefore, in the past decades genetic analysis of extant and extinct species through DNA analysis has rewritten significant parts of the animal tree of life.

Ancient DNA analysis of ancient remains is currently able to sequence complete genomes – allowing the vast, unbiased recovery of evolutionary information present in such genomes. This has led the

reconstruction of the evolutionary relationship of extinct species such as mammoths[25], flightless birds [26], or extinct lemurs [27] in comparison to their modern counterparts. They have even led to the identification of “Denisovans”, an otherwise unknown group of hominins more closely related to Neanderthals than to modern humans. In addition, as these analyses are not restricted to the maternally-inherited mitochondrial genome only, they are able to detect subtle but significant signatures of ancient introgression between closely related species. For example, all humans outside of Africa contain a small percentage (usually 2-3%) of DNA deriving from Neanderthals [28]. In other words, somewhere during the migration of humans out of Africa, they interacted with Neanderthals, interbred, and this genetic exchange is now identifiable in all non-African humans.

However, due to the intrinsic instability of ancient DNA, the preservation of ancient DNA is both temporally and geographically biased. Ancient DNA studies generally focus on specimens younger than 100,000 years, and in most cases a lot younger than that. In addition, ancient DNA preservation is generally good in temperate or permafrost samples, but rarely survives over significant amounts of time in subtropical or tropical regions. The evolutionary processes happening in the Pleistocene or in hot, humid environments are therefore not accessible in the majority of cases through ancient DNA analysis. Recent studies have indicated that ancient proteins survive for time periods longer than DNA, however [29,30] . Ancient proteomes can be phylogenetically informative even in cases where there is no ancient DNA preservation, implying that ancient protein analysis is able to recover evolutionary information beyond the preservation horizon of ancient DNA [31,32].

This has led a number of studies to explore the phylogenetic analysis of ancient proteins derived from now-extinct species [33] collagen type I sequences from *Plesioycteropus*. *Plesioycteropus* is an extinct mammal from Madagascar, a highly unfavourable environment for ancient DNA preservation. The phylogeny presented resolves the position of the “Malagasy aardvark” as more closely related to extant tenrecs than extant aardvarks (*Orycteropus*), demonstrating that the skeletal similarity between the *Plesioycteropus* and *Orycteropus* is due to convergence. Similar analyses focused on the recovery of collagen type I sequences, the dominant bone protein, have subsequently analyzed extinct ground sloths[34], South American ungulates (Figure 10.2; [35]), and an extinct giant beaver [36]. In each case, ancient DNA was either not present, unlikely to be present, or highly contaminated with modern DNA. Subsequently, a number of other studies focused on the recovery of entire ancient proteomes for similar purposes. Such larger proteomes are more informative, and therefore able to analyze more complex speciation patterns or species divergences that occurred close in time (i.e., there was little time available to accumulate sufficient, observable, changes in the protein sequence). They have been applied to the rhino genus *Stephanorhinus* [22,37], helped revise

the taxonomic identity of a syntype of the Asian elephant as an African elephant [38], and even elucidated the evolutionary placement of a now-extinct great ape [39].

Similarly, the ability to recover phylogenetically informative protein sequences from ancient skeletal tissues can also be applied to human and hominin tissues. For example, ancient genomes are available for Neanderthals, Denisovans, modern humans, and all great ape species. Together, these provide a catalogue of protein amino acid differences between homologous protein sequences. These single amino acid polymorphisms (SAPs) can be observed through protein mass spectrometry, something that recently has been demonstrated for Neanderthals and Denisovans [40,41]. Combined with the longevity of ancient protein survival compared to ancient DNA survival, this suggests that proteins represent a biomolecule able to provide molecular insights into the evolutionary processes shaping human evolution during the Early and Middle Pleistocene – time periods without regular ancient DNA survival, but a key period in the formation of the human lineage.

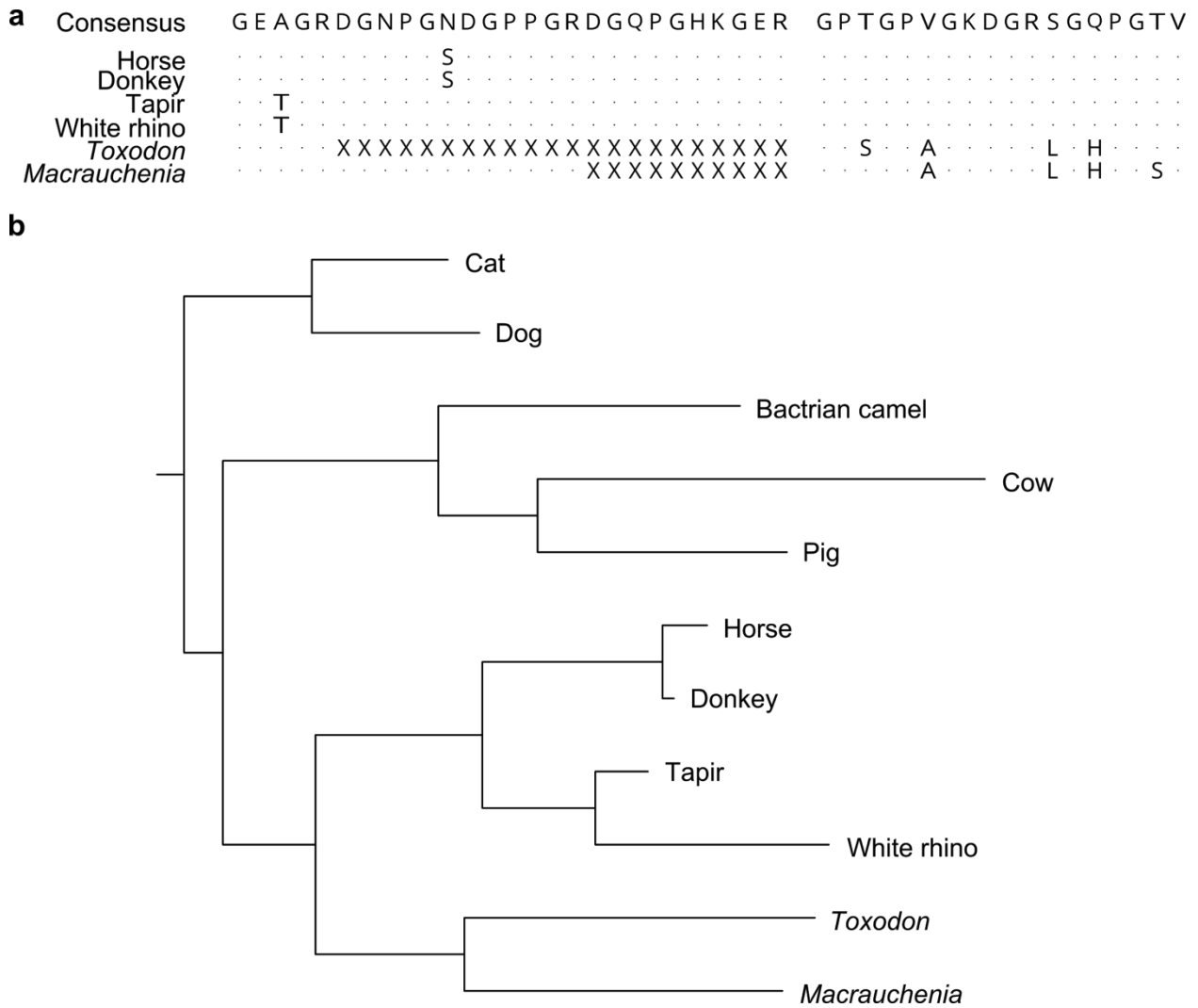


Figure 2: The phylogenetic analysis of ancient proteomes. **a** Partial protein sequence alignment of *Macrauchenia* and *Toxodon COL1a1*. Although the ancient protein sequences are incomplete, they contain phylogenetically-relevant positions (SAPs). Amino acid positions identical to the consensus are indicated by the “.”. **b** Phylogenetic analysis of a larger alignment of mammalian collagen type I sequences, indicating that *Macrauchenia* and *Toxodon* are distant relatives of extant odd-toed ungulates (Perissodactyla). Data from [35].

Contribution of palaeoproteomics to study ancient pathogens

The research field of paleopathology, or the study of the presence and prevalence of diseases in the past, has a history of more than 150 years. The scientific interest has been concentrated on the analysis of bones, teeth, and dental calculus since they are potential reservoirs of ancient biomolecules. Especially for the latter, teeth are coated with a bio-film known as dental plaque, that is made up out of microbes, proteins,

DNA, lipids, and other micro-residues that at some point entered the oral cavity. Over time, minerals deposited in plaque hardens to form tartar, or dental calculus, trapping its organic and non-organic components. The formed dental calculus can therefore provide valuable information regarding past diet and oral hygiene.

On the other hand, metagenomic and metaproteomic MS analysis of the oral microbiome of human dental calculus that is dated to the medieval time can indicate the diet, the environment, hygiene and diseases of people's past. Traditionally the microbiome is studied with techniques such as DNA sequencing. Warinner and colleagues demonstrated that protein mass spectrometry characterization of human dental plaque is possible. Subsequently, Jersie-Christensen and colleagues illustrated its potential to investigate the interactions between the host and the pathogenic species in a quantitative manner for ancient samples based on ancient proteins alone [42].

Additionally, apart from diseases in the past, dental plaque has been studied for the investigation of dietary habits. For instance, animal milk as a food resource has been identified to samples dating back to at least the Bronze Age (ca. 3000 BCE), directly from individuals and their dairy livestock and not from the survival of milk residues or milk bacteria in pottery fragments [43]. The identification of the whey protein β -lactoglobulin (*BLG*) with the use of tandem mass spectrometry has demonstrated its use as a milk biomarker of dairy consumption. In addition, the *BLG* protein has a different amino acid sequence between cattle, sheep, and goat species. This allows investigation of cultural, social, and environmental factors affecting past dairy economies at both a population and an individual level. Other studies have subsequently given insights into the origins of dairy consumption within different European populations. For instance, the work of Charlton et al. reveals the earliest direct evidence for the consumption of dairy products in the British Neolithic period, by applying shotgun proteomics to human dental calculus. The identification of *BLG* peptides not only illustrates the animal origin of the consumed milk but also creates new research avenues exploring how Neolithic populations may have been consuming and processing raw milk, including the potential variability which may have existed within such processes. Moreover, other proteomic studies illustrate the presence of other dietary proteins, such as those deriving from cereal grains, legumes, and vegetative crops, alongside milk-derived proteins [44]. The potential of dental calculus to provide insights into food consumption are therefore not limited to a single dietary class. However, even though dental calculus consists of a long-term reservoir of proteins, its contents are influenced by a combination of *in vivo* mechanisms and post-depositional degradation processes that can be expected to act differently between individuals [45]. High inter-individual variability of dental calculus proteomes might therefore complicate the synthesis of many datasets across time periods and geographic regions. A better understanding of the

aforementioned factors will lead to a better interpretation of the archaeological findings. Such knowledge should then allow to shed light on past human dietary choices in previously unforeseen ways.

Conclusions

This chapter reveals that bone, dentine and enamel proteomes are dynamic tissues with various potentials in the palaeoproteomic field. Rich of remaining proteins, skeletal proteomes can address phylogenetic questions and assist in species identification. Depending on the tissue of origin, there are marked differences in proteome composition between bone and dentine proteomes on the one hand and enamel proteomes on the other hand. This provides opportunities, but also challenges on data integration for phylogenetic studies. The presence of pathogenic microorganisms can indicate dietary habits, hygiene and diseases affecting past populations. In addition, the characterization of diagenetic and *in vivo* protein modifications further highlights the potential of bone, dentine, and enamel in providing access to past molecular processes occurring within the skeletal tissues of living individuals.

Author Contributions

Equal for both authors.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area).

References

- [1] B. Clarke, Normal bone anatomy and physiology., *Clin. J. Am. Soc. Nephrol.* 3 Suppl 3 (2008) S131. <https://doi.org/10.2215/CJN.04151206>.
- [2] E. Vuorio, B. De Crombrugge, The family of collagen genes, *Annu. Rev. Biochem.* 59 (1990) 837–872. <https://doi.org/10.1146/annurev.bi.59.070190.004201>.
- [3] D.J.S. Hulmes, On molecular packing in pN-collagen sheets, *Int. J. Biol. Macromol.* 14 (1992) 173–174. [https://doi.org/10.1016/S0141-8130\(05\)80009-1](https://doi.org/10.1016/S0141-8130(05)80009-1).
- [4] N. Procopio, A.T. Chamberlain, M. Buckley, Exploring Biological and Geological Age-related Changes through Variations in Intra- and Intertooth Proteomes of Ancient Dentine, *J. Proteome Res.* 17 (2018) 1000–1013. <https://doi.org/10.1021/acs.jproteome.7b00648>.
- [5] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J. V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.J. Hublin, L. Orlando, I. Barnes, R.D.E. Macphee, Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, *Nature.* 522 (2015) 81–84.

<https://doi.org/10.1038/nature14249>.

[6] M. Buckley, R.A. Fariña, C. Lawless, P.S. Tambusso, L. Varela, A.A. Carlini, J.E. Powell, J.G. Martinez, Collagen Sequence Analysis of the Extinct Giant Ground Sloths *Lestodon* and *Megatherium*, *PLoS One*. 10 (2015) e0139611. <https://doi.org/10.1371/journal.pone.0139611>.

[7] T.P. Cleland, E.R. Schroeter, R.S. Feranec, D. Vashishth, Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics, *Proc. R. Soc. B Biol. Sci.* 283 (2016). <https://doi.org/10.1098/rspb.2016.0593>.

[8] P. Petrone, P. Pucci, A. Vergara, A. Amoresano, L. Birolo, F. Pane, F. Sirano, M. Niola, C. Buccelli, V. Graziano, A hypothesis of sudden body fluid vaporization in the 79 AD victims of Vesuvius, *PLoS One*. 13 (2018) 1–27. <https://doi.org/10.1371/journal.pone.0203210>.

[9] T. Glatter, C. Ludwig, E. Ahrné, R. Aebersold, A.J.R. Heck, A. Schmidt, Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion, *J. Proteome Res.* 11 (2012) 5145–5156. <https://doi.org/10.1021/pr300273g>.

[10] Z. Wu, J. Huang, J. Huang, Q. Li, X. Zhang, Lys-C/Arg-C, a More Specific and Efficient Digestion Approach for Proteomics Studies, *Anal. Chem.* 90 (2018) 9700–9707. <https://doi.org/10.1021/acs.analchem.8b02448>.

[11] J. Rappsilber, Y. Ishihama, M. Mann, Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics, *Anal. Chem.* 75 (2003) 663–670. <https://doi.org/10.1021/ac026117i>.

[12] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: From bones to paintings, *Microchem. J.* 126 (2016) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.

[13] R.R. Jersie-Christensen, L.T. Lanigan, D. Lyon, M. Mackie, D. Belstrøm, C.D. Kelstrup, A.K. Fotakis, E. Willerslev, N. Lynnerup, L.J. Jensen, E. Cappellini, J. V. Olsen, Quantitative metaproteomics of medieval dental calculus reveals individual oral health status, *Nat. Commun.* 9 (2018). <https://doi.org/10.1038/s41467-018-07148-3>.

[14] E. Cappellini, A. Gentry, E. Palkopoulou, Y. Ishida, D. Cram, A.-M. Roos, M. Watson, U.S. Johansson, B. Fernholm, P. Agnelli, F. Barbagli, D.T.J. Littlewood, C.D. Kelstrup, J. V. Olsen, A.M. Lister, A.L. Roca, L. Dalén, M.T.P. Gilbert, Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae), *Zool. J. Linn. Soc.* 170 (2014) 222–232. <https://doi.org/10.1111/zoj.12084>.

[15] T.P. Cleland, E.R. Schroeter, M.H. Schweitzer, Biologically and diagenetically derived peptide modifications in moa collagens, *Proc. R. Soc. B Biol. Sci.* 282 (2015) 20150015. <https://doi.org/10.1098/rspb.2015.0015>.

[16] F. Welker, M.A. Soressi, M. Roussel, I. van Riemsdijk, J.J. Hublin, M.J. Collins, Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen, *Sci. Technol. Archaeol. Res.* 3 (2017) 15–27. <https://doi.org/10.1080/20548923.2016.1258825>.

[17] R. Sawafuji, E. Cappellini, T. Nagaoka, A.K. Fotakis, R. Rakownikow Jersie-Christensen, J. V Olsen, K. Hirata, S. Ueda, Proteomic profiling of archaeological human bone, (n.d.). <https://doi.org/10.1098/rsos.161004>.

[18] Identification of Pathological Conditions in Human Skeletal Remains - 2nd Edition, (n.d.). <https://www.elsevier.com/books/identification-of-pathological-conditions-in-human-skeletal-remains/ortner/978-0-12-528628-2> (accessed October 12, 2020).

[19] Demineralization–remineralization dynamics in teeth and bone, (n.d.). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5034904/> (accessed October 12, 2020).

[20] I.M. Porto, H.J. Laure, R.H. Tykot, F.B. de Sousa, J.C. Rosa, R.F. Gerlach, Recovery and identification of mature enamel proteins in ancient teeth, *Eur. J. Oral Sci.* 119 (2011) 83–87. <https://doi.org/10.1111/j.1600-0722.2011.00885.x>.

[21] G.A. Castiblanco, D. Rutishauser, L.L. Ilag, S. Martignon, J.E. Castellanos, W. Mejía, Identification of

proteins from human permanent erupted enamel, *Eur. J. Oral Sci.* 123 (2015) 390–395. <https://doi.org/10.1111/eos.12214>.

[22] E. Cappellini, F. Welker, L. Pandolfi, J.R. Madrigal, A. Fotakis, D. Lyon, V. Moreno Mayar, M. Bukhsianidze, R.R. Jersie-Christensen, M. Mackie, A. Ginolhac, R. Ferring, M. Tappen, E. Palkopoulou, D. Samodova, P. R  ther, M. Dickinson, T. Stafford, Y. Chan, A. G  therstr  m, S.K. Nathan, P. Heintzman, J. Kapp, I. Kirillova, Y. Moodley, J. Agusti, R.-D. Kahlke, G. Kiladze, B. Mart  nez-Navarro, S. Liu, M.S. Velasco, M.-H. Sinding, C. Kelstrup, M. Allentoft, A. Krogh, L. Orlando, K. Penkman, B. Shapiro, L. Rook, L. Dal  n, M.T.P. Gilbert, J. Olsen, D. Lordkipanidze, E. Willerslev, Early Pleistocene enamel proteome sequences from Dmanisi resolve *Stephanorhinus* phylogeny, *Bienvenido* 10 Mart  nez-Navarro. 17 (2018) 407692. <https://doi.org/10.1101/407692>.

[23] N.A. Stewart, G.F. Molina, J.P. Mardegan Issa, N.A. Yates, M. Sosovicka, A.R. Vieira, S.R.P. Line, J. Montgomery, R.F. Gerlach, The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etc extraction, *RSC Adv.* 6 (2016) 61673–61679. <https://doi.org/10.1039/c6ra05120k>.

[24] N.A. Stewart, R.F. Gerlach, R.L. Gowland, K.J. Gron, J. Montgomery, Sex determination of human remains from peptides in tooth enamel, *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017) 13649–13654. <https://doi.org/10.1073/pnas.1714926115>.

[25] E. Palkopoulou, M. Lipson, S. Mallick, S. Nielsen, N. Rohland, S. Baleka, E. Karpinski, A.M. Ivancevic, T.H. To, R. Daniel Kortschak, J.M. Raison, Z. Qu, T.J. Chin, K.W. Alt, S. Claesson, L. Dal  n, R.D.E. MacPhee, H. Meller, A.L. Roca, O.A. Ryder, D. Heiman, S. Young, M. Breen, C. Williams, B.L. Aken, M. Ruffier, E. Karlsson, J. Johnson, F. Di Palma, J. Alfoldi, D.L. Adelson, T. Mailund, K. Munch, K. Lindblad-Toh, M. Hofreiter, H. Poinar, D. Reich, A comprehensive genomic history of extinct and living elephants, *Proc. Natl. Acad. Sci. U. S. A.* 115 (2018) E2566–E2574. <https://doi.org/10.1073/pnas.1720554115>.

[26] M. Meyer, M. Kircher, M.T. Gansauge, H. Li, F. Racimo, S. Mallick, J.G. Schraiber, F. Jay, K. Pr  fer, C. De Filippo, P.H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R.E. Green, K. Bryc, A.W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M.F. Hammer, M. V. Shunkov, A.P. Derevianko, N. Patterson, A.M. Andr  s, E.E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. P  abo, A high-coverage genome sequence from an archaic Denisovan individual, *Science* (80-.). 338 (2012) 222–226. <https://doi.org/10.1126/science.1224344>.

[27] L. Kistler, A. Ratan, L.R. Godfrey, B.E. Crowley, C.E. Hughes, R. Lei, Y. Cui, M.L. Wood, K.M. Muldoon, H. Andriamialison, J.J. McGraw, L.P. Tomsho, S.C. Schuster, W. Miller, E.E. Louis, A.D. Yoder, R.S. Malhi, G.H. Perry, Comparative and population mitogenomic analyses of Madagascar’s extinct, giant “subfossil” lemurs, *J. Hum. Evol.* 79 (2015) 45–54. <https://doi.org/10.1016/j.jhevol.2014.06.016>.

[28] R.E. Green, J. Krause, A.W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M.H.Y. Fritz, N.F. Hansen, E.Y. Durand, A.S. Malaspinas, J.D. Jensen, T. Marques-Bonet, C. Alkan, K. Pr  fer, M. Meyer, H.A. Burbano, J.M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. H  ber, B. H  ffner, M. Siegemund, A. Weihmann, C. Nusbaum, E.S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic,  . Kucan, I. Gu  ic, V.B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. De La Rasilla, J. Fortea, A. Rosas, R.W. Schmitz, P.L.F. Johnson, E.E. Eichler, D. Falush, E. Birney, J.C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. P  abo, A draft sequence of the neandertal genome, *Science* (80-.). 328 (2010) 710–722. <https://doi.org/10.1126/science.1188021>.

[29] B. Demarchi, S. Hall, T. Roncal-Herrero, C.L. Freeman, J. Woolley, M.K. Crisp, J. Wilson, A. Fotakis, R. Fischer, B.M. Kessler, R.R. Jersie-Christensen, J. V. Olsen, J. Haile, J. Thomas, C.W. Marean, J. Parkington, S. Presslee, J. Lee-Thorp, P. Ditchfield, J.F. Hamilton, M.W. Ward, C.M. Wang, M.D. Shaw, T. Harrison, M. Dom  nguez-Rodrigo, R.D.E. Macphee, A. Kwekason, M. Ecker, L.K. Horwitz, M. Chazan, R. Kroger, J. Thomas-Oates, J.H. Harding, E. Cappellini, K. Penkman, M.J. Collins, Protein sequences bound to mineral surfaces persist into deep time, *Elife.* 5 (2016). <https://doi.org/10.7554/eLife.17092>.

[30] N. Rybczynski, J.C. Gosse, C. Richard Harington, R.A. Wogelius, A.J. Hidy, M. Buckley, Mid-Pliocene

warm-period deposits in the High Arctic yield insight into camel evolution, *Nat. Commun.* 4 (2013) 1–9. <https://doi.org/10.1038/ncomms2516>.

[31] E. Cappellini, A. Prohaska, F. Racimo, F. Welker, M.W. Pedersen, M.E. Allentoft, P. de Barros Damgaard, P. Gutenbrunner, J. Dunne, S. Hammann, M. Roffet-Salque, M. Ilardo, J.V. Moreno-Mayar, Y. Wang, M. Sikora, L. Vinner, J. Cox, R.P. Evershed, E. Willerslev, Ancient Biomolecules and Evolutionary Inference, *Annu. Rev. Biochem.* 87 (2018) 1029–1060. <https://doi.org/10.1146/annurev-biochem-062917-012002>.

[32] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J. V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.J. Hublin, L. Orlando, I. Barnes, R.D.E. Macphee, Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, *Nature.* 522 (2015) 81–84. <https://doi.org/10.1038/nature14249>.

[33] M. Buckley, A Molecular Phylogeny of Plesiorycteropus Reassigns the Extinct Mammalian Order 'Bibymalagasia,' *PLoS One.* 8 (2013) e59614. <https://doi.org/10.1371/journal.pone.0059614>.

[34] M. Buckley, R.A. Fariña, C. Lawless, P.S. Tambusso, L. Varela, A.A. Carlini, J.E. Powell, J.G. Martinez, Collagen Sequence Analysis of the Extinct Giant Ground Sloths *Lestodon* and *Megatherium*, *PLoS One.* 10 (2015) e0139611. <https://doi.org/10.1371/journal.pone.0139611>.

[35] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J. V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.J. Hublin, L. Orlando, I. Barnes, R.D.E. Macphee, Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, *Nature.* 522 (2015) 81–84. <https://doi.org/10.1038/nature14249>.

[36] T.P. Cleland, E.R. Schroeter, M.H. Schweitzer, Biologically and diagenetically derived peptide modifications in moa collagens, *Proc. R. Soc. B Biol. Sci.* 282 (2015) 20150015. <https://doi.org/10.1098/rspb.2015.0015>.

[37] F. Welker, G.M. Smith, J.M. Hutson, L. Kindler, A. Garcia-Moreno, A. Villaluenga, E. Turner, S. Gaudzinski-Windheuser, Middle Pleistocene protein sequences from the rhinoceros genus *Stephanorhinus* and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae, *PeerJ.* 2017 (2017). <https://doi.org/10.7717/peerj.3033>.

[38] E. Cappellini, A. Gentry, E. Palkopoulou, Y. Ishida, D. Cram, A.M. Roos, M. Watson, U.S. Johansson, B. Fernholm, P. Agnelli, F. Barbagli, D.T.J. Littlewood, C.D. Kelstrup, J. V. Olsen, A.M. Lister, A.L. Roca, L. Dalén, M.T.P. Gilbert, Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae), *Zool. J. Linn. Soc.* 170 (2014) 222–232. <https://doi.org/10.1111/zoj.12084>.

[39] F. Welker, Palaeoproteomic analysis of Early Pleistocene *Gigantopithecus blacki*, *Am. J. Phys. Anthropol.* 168 (2019).

[40] F. Welker, M. Hajdinjak, S. Talamo, K. Jaouen, M. Dannemann, F. David, M. Julien, M. Meyer, J. Kelso, I. Barnes, S. Brace, P. Kamminga, R. Fischer, B.M. Kessler, J.R. Stewart, S. Pääbo, M.J. Collins, J.J. Hublin, Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 11162–11167. <https://doi.org/10.1073/pnas.1605834113>.

[41] F. Chen, F. Welker, C.C. Shen, S.E. Bailey, I. Bergmann, S. Davis, H. Xia, H. Wang, R. Fischer, S.E. Freidline, T.L. Yu, M.M. Skinner, S. Stelzer, G. Dong, Q. Fu, G. Dong, J. Wang, D. Zhang, J.J. Hublin, A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau, *Nature.* 569 (2019) 409–412. <https://doi.org/10.1038/s41586-019-1139-x>.

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

<https://doi.org/10.1038/s41467-018-07148-3>.

[43] C. Warinner, J.F.M. Rodrigues, R. Vyas, C. Trachsel, N. Shved, J. Grossmann, A. Radini, Y. Hancock, R.Y. Tito, S. Fiddyment, C. Speller, J. Hendy, S. Charlton, H.U. Luder, D.C. Salazar-García, E. Eppler, R. Seiler, L.H. Hansen, J.A.S. Castruita, S. Barkow-Oesterreicher, K.Y. Teoh, C.D. Kelstrup, J. V. Olsen, P. Nanni, T. Kawai, E. Willerslev, C. Von Mering, C.M. Lewis, M.J. Collins, M.T.P. Gilbert, F. Rühli, E. Cappellini, Pathogens and host immunity in the ancient human oral cavity, *Nat. Genet.* 46 (2014) 336–344. <https://doi.org/10.1038/ng.2906>.

[44] J. Hendy, C. Warinner, A. Bouwman, M.J. Collins, S. Fiddyment, R. Fischer, R. Hagan, C.A. Hofman, M. Holst, E. Chaves, L. Klaus, G. Larson, M. Mackie, K. McGrath, A.Z. Mundorff, A. Radini, H. Rao, C. Trachsel, I.M. Velsko, C.F. Speller, Proteomic evidence of dietary sources in ancient dental calculus, *Proc. R. Soc. B Biol. Sci.* 285 (2018) 20180977. <https://doi.org/10.1098/rspb.2018.0977>.

[45] M. Mackie, J. Hendy, A.D. Lowe, A. Sperduti, M. Holst, M.J. Collins, C.F. Speller, Preservation of the metaproteome: variability of protein preservation in ancient dental calculus, *Sci. Technol. Archaeol. Res.* 3 (2017) 58–70. <https://doi.org/10.1080/20548923.2017.1361629>.

3.2 Manuscript 2: Written in bone proteins: molecular signatures in the 79 AD skeletal remains from Herculaneum and Pompeii.

Georgia Ntasi, Ismael Rodriguez Palomo, Gennaro Marino, Francesco Sirano, Enrico Cappellini, Leila Birolo, and Pierpaolo Petrone.

Abstract

In search of markers for protein aging, we used an extraordinary test case of human bones from the excavations sites of Pompeii and Herculaneum (79 AD), analyzed by a shotgun proteomic approach using LC-MS/MS. Human bones from Scaldrone Bay (II sec AD, another volcanic area, yet not related to Vesuvius eruption) were also analysed for the comparison of the PTMs provoked from a volcanic ground to an almost coeval standard burial condition. In addition, already published archaeological bones, approximately 17 centuries younger than the test case, were used as control samples [1]. Deamidation of asparagine and glutamine was examined as a now widely considered aging molecular marker [2]. Deamidation profile was quite robust for samples with a similar age whereas the fewer proteins were detected, the more degraded the sample were found to be. Positional investigation of deamidation in the polypeptide sequence of collagen type I revealed that some zones are more susceptible to deamidation than others. Backbone cleavage analysis confirmed that some zones apart from deamidation are more susceptible to spontaneous hydrolysis too. Analysis of post-translational modifications showed higher advanced glycation products of lysine and arginines in the samples of Pompeii and Herculaneum. Furthermore, oxidative products of methionine, histidine and the conversion of ST to glycine were also detected. Interestingly, multiple oxidation products of proline and dehydration of hydroxyproline were detected for the first time, turning the scientific interest to this amino acid (P), which is usually ignored due to its natural conversion to hydroxyproline. As a matter of fact, we can observe that collagen degradation in bones from Pompeii and Herculaneum underwent different degradation processes.

Introduction

Ancient protein analysis provides clues to human life and diseases from ancient times. Proteins represents promising means for identifying and characterizing archaeological and fossil samples, either human and animal as well as food and goods.

Among archaeological remains, bones are most interesting. Bones, indeed, can be an abundant reservoir of ancient biomolecules due to natural resistance to post-mortem decay arising from a unique combination of mechanical, structural, and chemical properties. The survival of proteins in fossil bones has been of great

interest for their potential in recovering phylogenetic information for over 30 years [1]. The application of proteomics to the study of ancient bones holds great potential for increasing our understanding of many aspects of “paleobiology” such as the evolution, the habits and the ecology of ancient organisms [2-6].

Diagenesis of bones, however, the process by which skeletal tissues are transformed post-mortem, is quite complex and depends upon several factors [7,8]. Central to the extreme preservation is the close interplay between the organic and the inorganic constituents, while the diagenetic destiny is closely linked to the nature of the depositional environment [7]. Bone is a composite material made up by inorganic material, namely bioapatite, carbonate-substituted hydroxyapatite, and organic portion, of which about 90% is Type I collagen, the remaining 10% comprising the so-called non-collagenous proteins (NCPs), lipoproteins and mucopolysaccharides [7]. Collagens are unique and the dominant proteins in most eukaryotes, well preserved and highly conserved throughout the evolution of species due to their essential biological function as well as their peculiar sequence, made of one third of glycine, and characterized by the repetitious sequence Gly-Xaa-Yaa, where the Xaa and Yaa positions are typically found to be proline and hydroxyproline, respectively, the latter being a peculiar and abundant post-translation modification contributing to collagen stability via hydrogen bridges. Despite the quite simple amino acid sequence of collagen chains, collagen experiences a wide and heterogeneous array of post-translational modifications (PTMs), ranging from hydroxylation (on P and L), glycosylation, and crosslinking. Mass spectrometry-based techniques have been widely applied to characterize collagen and its PTMs [9] (and references therein), and are expected to provide interesting clues also in chemical modifications induced by post-mortem protein degradation. So far, analysis of ancient bone proteins highlighted expected modifications such as deamidation of asparagine and glutamine Gln [10-13] and methionine oxidation [14]. Actually some extent of deamidation is now considered so unavoidable for aged collagen [15] to be taken into account as forensic marker in the evaluation of post-mortem decay of bones [16]. A few further diagenetically induced modifications were observed: some aminoadipic acid formation from lysine and tryptophan oxidation products [14], advanced glycation end-product (AGEs) [14,17,18], backbone cleavage [17,18], and some reactions occurring on prolines have been only hypothesized [17,14].

After death, bone tissues undergo different stages of modification during which microcrystalline biogenic apatite and the fibrous collagen matrix stabilise each other, with chemical degradation of collagen influenced by several factors. Definitively, temperature is an obviously suggested major factor in affecting collagen loss in bone, with burial soil conditions such as extreme acidity or alkalinity exacerbating the hydrolysis of proteins [7].

In this respect, the skeletal remains from the archaeological sites of Herculaneum and Pompeii represent a unique and remarkable case of study, because of the peculiar destiny, state of preservation [19] and, above all, because all individuals experienced analogous environmental burial conditions. Herculaneum, Pompeii and other Roman settlements up to 20 kilometers away from the volcano were suddenly hit and buried by successive hot pyroclastic flows produced by the eruption of Vesuvius in AD 79, that killed everyone who had not been evacuated or managed to flee [20].

Here we searched for molecular signatures imprinted in the survived protein component performed on bones collected from a significant group of victims from the ancient beach of Herculaneum [21], and also from Pompeii [19]. The skeletons were in an extraordinarily good state of preservation as a result of the unusual death and burial conditions involved: instant death caused at the emplacement of hot pyroclastic surge at about 500 °C, [13]. In the case of Herculaneum, death was followed by rapid vaporization of soft tissues and replacement by the volcanic ash [19].

At Herculaneum the ash bed deposit was permanently saturated by groundwater [15], therefore the anoxic burial environment of the skeletons was most likely able to inhibit microbial attack and related diagenetic processes [22]. This is testified by the microscopic examination of ground cross sections of the juvenile and adult long bones from both sexes, which did not show evidence of structural diagenetic change by microorganism activity [21]. A sole osteoalteration consisted of widespread micro cracking, due to exposure of the victims' corpses to the pyroclastic surge high temperature [23,24]. Being a unique cross section of the entire living population, the Herculaneum skeletons are particularly suited to paleobiological investigation. Several bone samples were also taken from the Pompeii plaster casts [25].

In Pompeii, the temperature of 250-300 °C experienced by the victims was sufficient to kill people instantly, but it was not high enough to vaporize the soft tissues as was the case in Herculaneum, then the corpses were preserved intact inside the volcanic ash deposit. The cavity formed around the victim's body after the slow disappearance of the flesh would then be filled with plaster of Paris in order to obtain plaster casts, technique adopted for the first time on human victims in 1863 by Giuseppe Fiorelli [19]. For comparison, the bones from a coeval skeletal population from the Campanian region were considered (Baia, Scalandrone site, II sec. AD, Roman imperial age, Puteoli, Naples, Italy). It's interesting to notice that some family groups belonging to *gentes* from Pompeii and Herculaneum migrated to Puteoli after the eruption of Vesuvius [19].

Bones can be considered time capsules, like the 79 AD victims' corpses as well [19], since individual destiny can be imprinted on their molecules [12]. The observed chemical modifications are all signatures of the

oxidative environment to which the bodies were exposed, and data herein reported testify the exceptional preservation of proteins. In particular, bone collagen survives even the harshest conditions of temperature imposed by the volcano eruption, bearing encoded in the chemical modifications a piece of history.

Results

Proteome of human bones.

A shotgun proteomics approach by LC-MS/MS (Figure S2) was applied to skeletal samples excavated from Pompeii (7 samples) and Herculaneum (5 samples) archaeological sites (79 AD victims of Vesuvius eruption). Samples collected from three individuals from Roman Imperial Age (II sec. AD) cemetery in Baia Scalandrone locality were also analyzed to compare results from the victims of the AD 79 eruption with a coeval skeletal population. All specimens are illustrated in the table S1 and figure S1. Moreover, raw data from proteomic analyses of archaeological bones from a completely independent excavation site were also used for comparative evaluation of the results: samples H-162, H-142 from [15] were selected because proteins have been extracted with almost the same protocol and analysed on the same instrument as herein samples; Very stringent criteria for protein identification were used: only peptides with score higher than 70 were considered and proteins were considered as identified only when 2 or more peptides have been detected (Table S2).

The number of identified proteins in the samples from Herculaneum, Pompeii and Baia Scalandrone varied among individuals from 2 to 27 and included collagenous and non-collagenous proteins (NCPs) (see Table S3). The expected dominance of collagen in bone tissue is reflected by the result that the two chains of type 1 collagen, namely collagen alpha-1 (I) and collagen alpha-2 (I), were confidently identified in all the samples. Of the 15 samples, 5 contained only type 1 collagen chains (four samples out of the seven from Pompeii, 2.8 proteins on average, ± 1.12 , and one sample from Baia Scalandrone) while samples from Herculaneum exhibited a quite wider protein content (13 proteins on average, ± 7). Bone proteome complexity is affected by several factors, including burial age [26], but, in this case, since bones from Pompeii and Herculaneum are exactly coeval, death/burial environmental conditions have done the difference in protein survival in the two different sets of samples [16]. Raw data of Control samples were processed with the same constraints as the samples herein prepared, and a large number of proteins, (38 ± 0) were identified, as already reported [15].

Despite a general big variability among the individuals within the different groups (Figure S3), samples can be grouped in function of the excavation site in respect to NCP content. We definitely observed that samples from Herculaneum conserved more NCPs than samples from Pompeii. Venn diagram using the ensembles of

the proteins identified in each sample group shows that the proteins identified in Pompei are common to all the groups, and that Herculaneum and Scalandrone Bay shared several proteins (Figure S4, table S4).

The types of non-collagenous proteins identified in this study agree with those expected for an archaeological bone [27, 28]. Although the abundant collagen might be masking the presence of some other low-abundance NCPs, most of the identified NCPs proteins are small leucine-rich proteoglycans (SLRPs) from the extracellular matrix (namely chondroadherin, biglycan, decorin, lumican and osteomodulin), all involved in biomineralisation or interacting with fibrillar collagen, (such as vitronectin and pigment epithelium-derived factor). Moreover, alpha-2-HS-glycoprotein, also known as fetuin-A is a bone matrix protein, known to have a high affinity for apatite. It is worth mentioning that several of the NCPs are related to the coagulation pathway (namely, prothrombin and antithrombin III, and Protein Z-dependent protease inhibitor), and can be functionally connected in the STRING network (Figure S5).

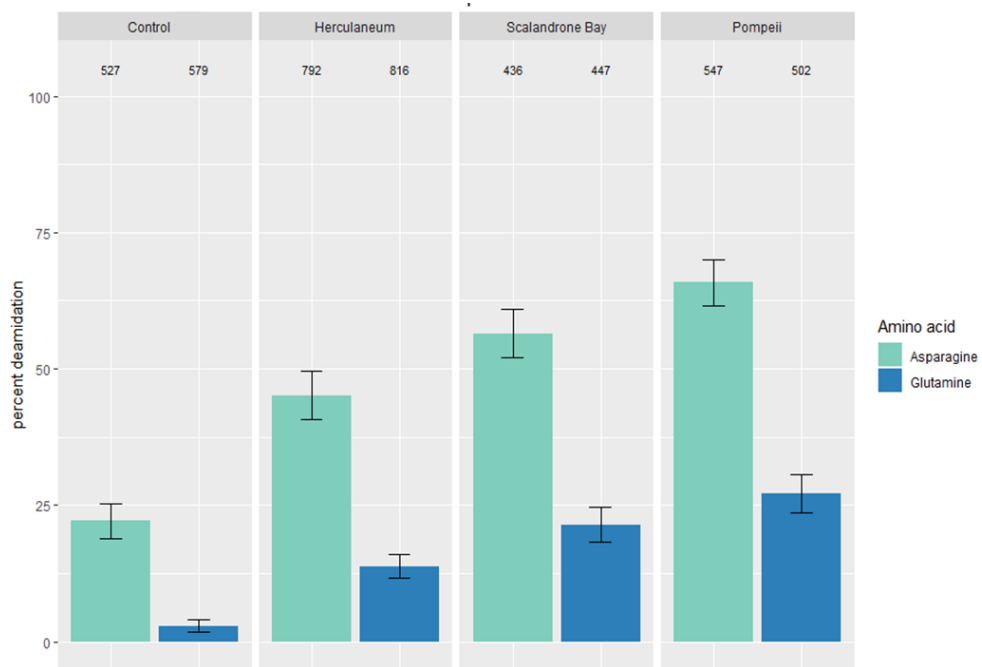
Diagenetically induced modifications in bone proteins

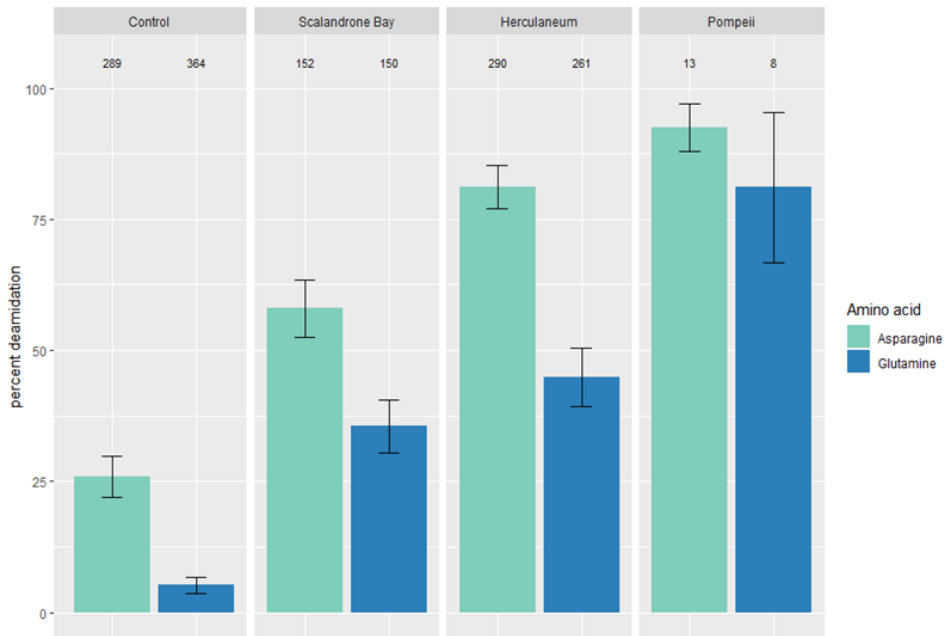
Modifications of amino acids, such as oxidation of methionine, deamidation of asparagine and glutamine, as well as the backbone cleavage, are all degradation phenomena commonly observed and routinely searched for in ancient/aged proteins [14, 29].

Deamidation. Firstly, deamidation of asparagine (N) and glutamine (Q) residues, among the most common and most informative diagenetically derived modifications in proteins [30, 13, 12], was examined. While the occurrence of extensive deamidation technically complicates the analysis of proteins, it also is a general and relevant glance on the myriad changes that archaeological bone proteins undergo and potentially relates them to the age and/or preservation of the sample under consideration [31]. Deamidation (N,Q) is indeed considered an indication of “authentic age”, and it has recently been systematically measured in an extensive set of archaeological and paleontological bones as well as other archaeological remains [30]. We evaluated deamidation occurrence in this set of samples, keeping in mind the suggestion by [30] that deamidation should be rather viewed as a global indicator of sample preservational quality, since rates and levels of deamidation are affected by several chemical and environmental factors. As expected, proteins in our samples are extensively deamidated and asparagine sites are much more deamidated than glutamines [6,30], (Figure 1). Moreover, we split the evaluation of the deamidation levels for collagenous and non-collagenous proteins and, in agreement to what already reported by [18], peptides from non-collagenous proteins showed very high to complete deamidation in comparison to peptides derived from collagenous

proteins (Figure 1). On average, peptides from Pompeii samples are the most deamidated (see bulk deamidation per site, Figure 1).

It is worth mentioning that Control samples are significantly younger than samples from Vesuvius area (Control samples date 1657–1683 AD [15]), and are definitively less deamidated than Pompeii and Herculaneum samples. Variations were observed from individual to individual in the three groups from the Neapolitan volcanic area (Figure S6-S7). As a general trend, we can confidently assess that the lower the number of surviving proteins the higher the deamidation level, and, it can be worth observing that the few NCPs identified in samples from Pompeii are almost completely deamidated. Apparently, other factors rather than temperature might have played the biggest role in deamidation, since we can see that samples from Scalandrone bay, that were not exposed to the high temperatures of the volcanic eruption, but had been also buried in a soil of volcanic origin (standard burial in agricultural soil, yet still in the rather similar volcanic area of Campi Flegrei), exhibited almost comparable deamidation level as samples from Herculaneum and Pompeii.





B

Figure 1: Overall percentage of deamidation for asparagine and glutamine residues of collagenous (A) and non-collagenous (B) bone proteins from Pompeii, Herculaneum, Scalandrone bay and control (H-162, H-142) from [15]). Error bars represent standard deviation and numbers above each bar represent the number of deamidation sites the data is based on.

We analysed the distribution of deamidation level along the sequence of collagen type I chains, to explore the possibility of hot spots for deamidation rather than an average distribution. Figure S8 illustrates the deamidation values at single deamidation sites along collagen alpha-1 (I) and alpha-2 (I) chains. The label size indicates the relative intensity of each position in each sample. The values for Control are always well below those calculated for samples of the volcanic area, in agreement with the global deamidation level calculated in figure 1. This difference is even more evident in glutamines, conceivably because glutamine deamidates more slowly. There is a trend in the deamidation; there are some zones where deamidation is more pronounced than others. This trend is almost reproducible in the samples of Pompeii, Herculaneum and Scalandrone Bay giving the sense that deamidation profile is quite robust for samples with a similar age and similar burial soil and suggesting that three-dimensional arrangement might affect the local deamidation level.

Oxidation of methionines: With the same approach the oxidation of methionines (M) was evaluated. Figures S9&S10 illustrate the global oxidation levels of collagenous and non-collagenous proteins in all groups. Apart from the zero values of control samples, all the proteins in all groups are almost totally oxidized (100%),

demonstrating that methionine oxidation follows another pattern than deamidation. Furthermore, we investigate the oxidation values at single oxidation sites along collagen alpha-1 (I) and alpha-2 (I) chains. Almost all oxidation values are either 1 or 0, meaning that methionines are either fully oxidized or not oxidized at all.(Fig S11). However, it is worth saying that several Met were not detected, despite the generally good protein sequence coverage.

Backbone cleavage. Backbone cleavage of the polypeptide chain is also expected as a degradation feature in ancient proteins [32, 18, 33], and can be evaluated since, upon trypsin hydrolysis, semi-tryptic peptides will be generated, with trypsin cleavage site only at one end. Search for semi-tryptic peptides was carried out on collagen type I chains only for comparative purposes, since they are the only polypeptide chains shared among all the samples. The frequency of semitryptic peptides was evaluated as the percentage of semitryptic peptides over the total number of identified peptides for each chain, on the basis of spectrum matches (PSMs).

Fig. 2 reports the relative abundance of peptide-spectrum matches (PSMs) of semitryptic peptides over the total number of peptides of collagen alpha-1 (I) and alpha-2 (I) chains in all the samples considered. The frequency of backbone cleavages is generally high. However, no clear-cut difference was observed among the samples from the volcanic area or with the control sample.

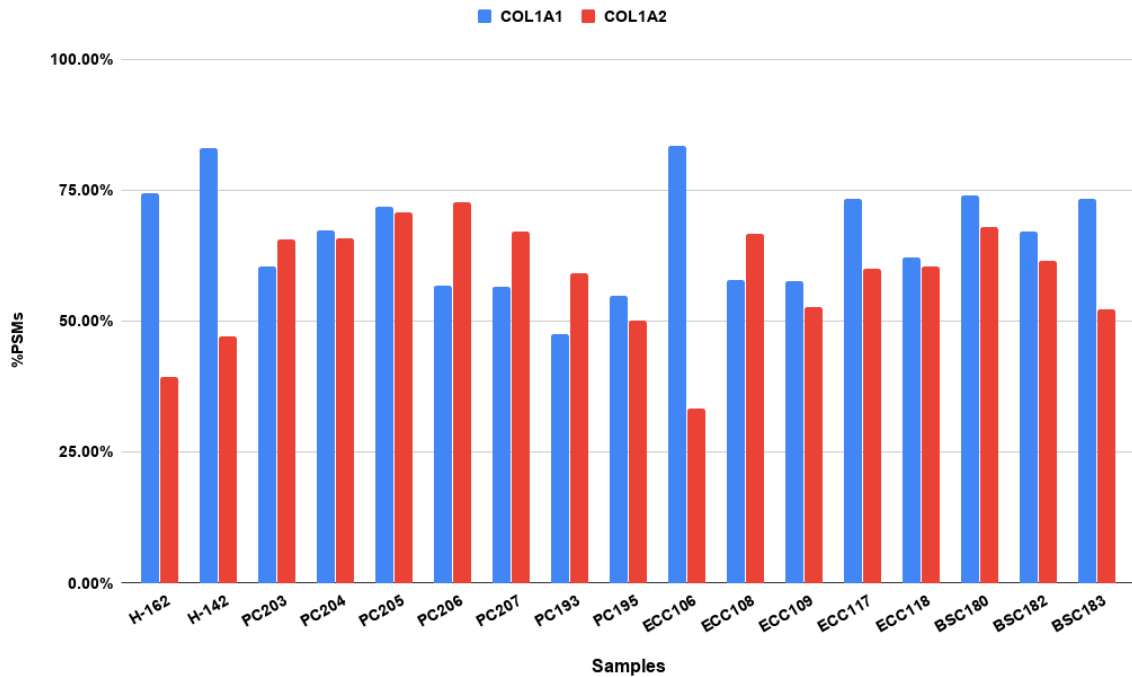


Figure 2: Occurrence of backbone cleavages evaluated as percentage of PSMs of semitryptic peptides in the single samples. The number of peptide-spectrum matches (PSMs) of semitryptic peptides over the total number of peptide-spectrum matches (tryptic plus semitryptic peptides) is reported for collagen alpha 1(I) and collagen alpha 2 (I) for each sample.

The peptides showed a very degraded pattern similar to a “pacman” pattern meaning that the observed peptides are the result of cleavages occurring in regions of the collagen chains rather than in specific peptide bonds (Figure S12), differing in one amino acid. A manual alignment of all the semitryptic peptides in the five different groups (Pompeii, Herculaneum, Scalandone bay Control) to COL1A1 and COL1A2 sequences, however, reveals that while in case of controls the number cleavages are spread along the sequences while in the samples of Pompeii, Herculaneum and Scalandone Bay are localized in some regions of the protein sequences. These hot spots are between 266-286, 481-511, 772-796 and 1051-1118 sites of COL1A1 with a window of ± 2 amino acids, and in COL1A2 between 154-167, 232-250, 320-340, 425-438, 499-517, 681-708, 965-1006 and 1042-1066 with a window of ± 2 amino acids (Figure S13).

The cleavage frequency was then re-evaluated considering the regions rather than the single peptide bonds, by calculating the number of PSMs with semitryptic cleavages identified in a region divided by the total PSMs in the same region, including both tryptic and semitryptic matches. As figure S13 reports, the regions listed above are more hydrolyzed in samples of Pompeii, Herculaneum and Scalandrone Bay. The higher occurrence of backbone cleavages suggests a different state of preservation in a volcanic soil.

Other non-enzymatic modifications. The data-dependent peptide algorithm of MaxQuant [34][35] was used for an unbiased discovery of chemical modifications (CMs) in the samples. The CMs were ranked by their occurrence within the dataset. The modifications were chosen after filtering with localization probabilities of $\geq 80\%$ for modified peptides and occurrence of detection of DP Cluster Mass ≥ 5 times for each sample (see figure 3). As expected, hydroxylation of prolines is fairly abundant, actually overwhelming most of the other modifications, as well as deamidation at asparagines and glutamines.

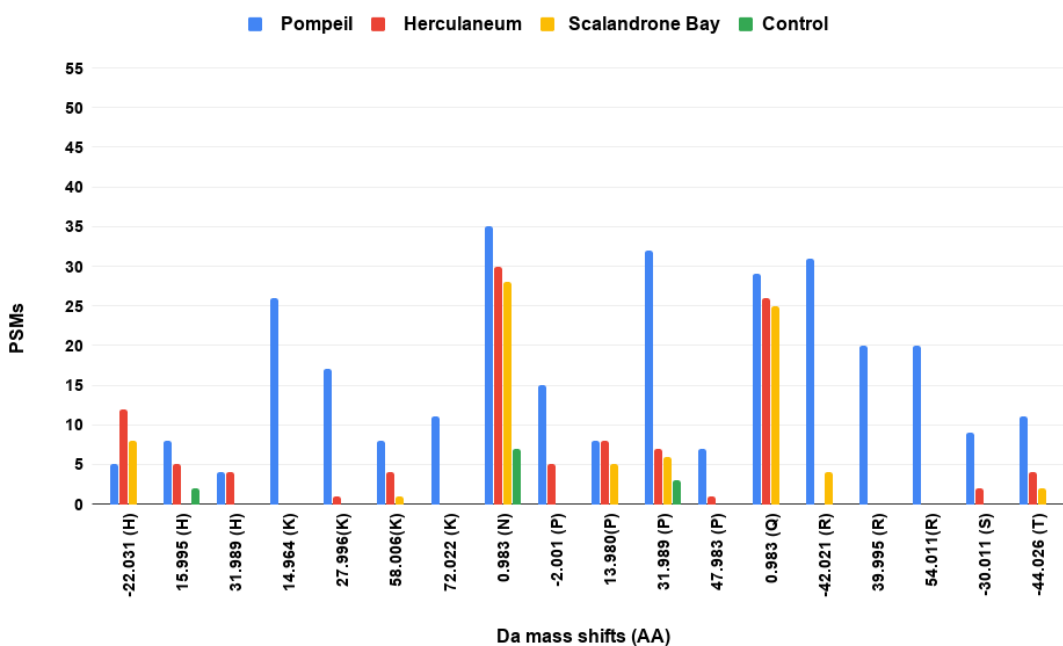


Figure 3: Number of identified PSMs of “dependent peptides” with mass shifts for collagen chains in the groups of Pompeii, Herculaneum, Scalandrone bay and the control samples (H-142, H-162) [15]. Mass shifts were selected after filtering with localization probabilities of $\geq 80\%$ for modified peptides and occurrence of detection of DP Cluster Mass ≥ 5 times for each sample. Reported data only include mass shifts corresponding to known oxidative modifications with matching amino acid targets (Unimod, <http://www.unimod.org/>).

As a second step, the selected CMs were inserted as variable modifications in standard MaxQuant searches, by setting the modifications as variable in separate runs, for each group separately, as detailed in table S2. To confirm peptide assignment, we manually inspected MS/MS spectra (and some examples are reported in the supplementary information) thus also allowing us to confidently assess the site localization of the selected chemical modifications.

The frequency of modified residues in respect to the amino acid detection is reported in the tables S4. It is worth mentioning that each position was considered only once in this calculation, even when the position was present in overlapping peptides.

Furthermore, the frequency of chemical modifications at a specific primary structure position was semi quantitatively evaluated using the MaxQuant calculation of mod/base ratio as reported in [36].

Interestingly, a high occurrence of mass shifts on lysine (K) and arginine (R) (Figure 4 and 5) was observed, that were interpreted as glycation products, with a high incidence in the samples group of Pompeii. Protein glycation involves the binding of reducing sugar carbonyl groups to protein amino groups, or the reaction of α -dicarbonyls such as glyoxal or methylglyoxal, that are continuously formed during oxidative degradation of sugars, with lysine and arginine residues, leading to a series of molecular reactions collectively called Maillard reaction that generate a variety of complex compounds called advanced glycation end products (AGEs) [37-40]. Among lysine-derived AGEs, N^ε-(carboxymethyl)lysine (CML) and N^ε-(carboxyethyl)lysine (CEL) are the most studied representatives and were significantly observed in the samples from the eruptive area (Figure 4). Formylation at lysine side chains, oxidative deamination of lysine to amino adipic acid, another marker of protein carbonyl oxidation [41] that can be associated to decomposition after death [14], and carbamylation, that has been reported as a hallmark of protein aging [42], were all also observed in collagen from samples from Herculaneum and Pompeii. Among arginine-related AGEs we detected the hydroimidazolones MG-H1 and G-H1 formed by reaction of arginine side chain with the oxoaldehydes methylglyoxal and glyoxal [43], respectively, and a substantial formation of ornithine (Fig S16 and fragmentation spectra at fig S26) [44], that was also recently identified in ancient dental enamel proteins [45].

These modifications are less frequent in the control sample and, within samples from the eruptive area, are significantly more represented in collagen from bones from Pompeii (Figure 4 and 5).

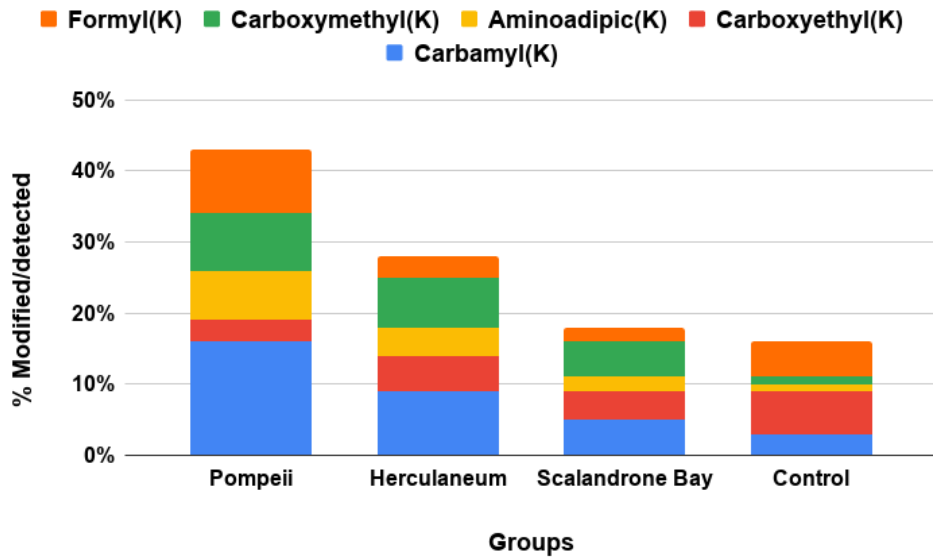


Figure 4: Extent of modified lysine residues, reported as percentage of modified over detected (modified plus unmodified) ones.

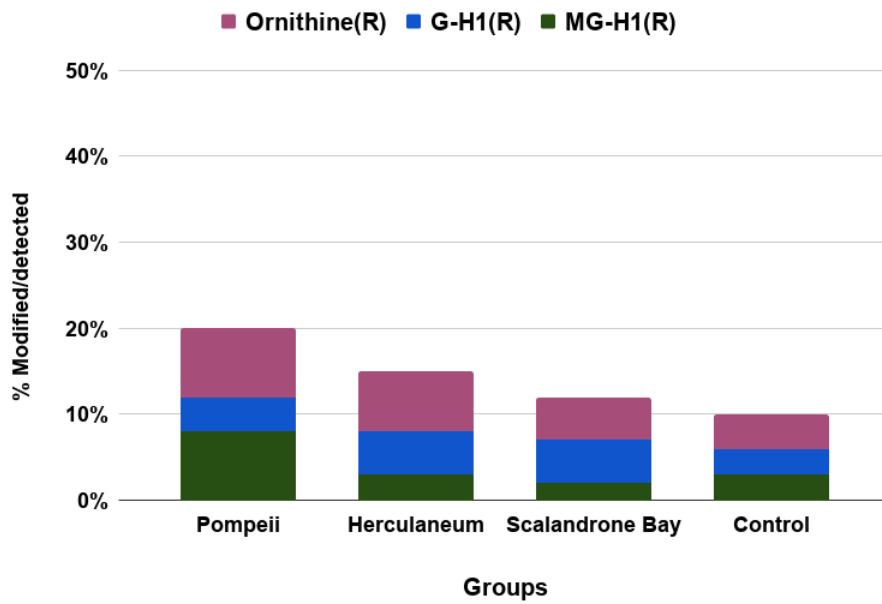


Figure 5: Extent of modified arginine residues, reported as percentage of modified over detected (modified plus unmodified) ones.

Histidine is one of the targets of oxidative modifications [29], generating 2-oxohistidine and dioxohistidine that can evolve further to break down to aspartic acid. An extensive oxidation of histidine residues in collagen chains from the bones from the eruptive area was observed (Fig. 6). In fact, more than 65% of collagenous histidine residues in Pompeii and Herculaneum have been found modified (tableS3d).

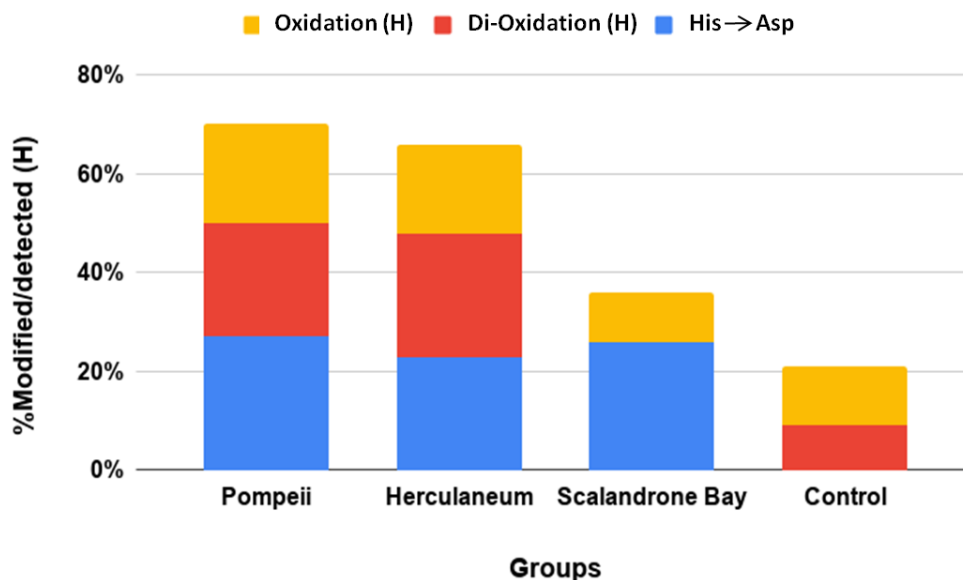


Figure 6: Extent of modified histidine residues, reported as percentage of modified over detected (modified plus unmodified) ones.

Mass shifts that are consistent with the C α -C β bond cleavage of the side chains of serine and threonine, which result in the formation of glycine (G) (-30.011 Da and -44.026 Da, respectively) were observed (Fig. 7, Fig S17, Fig S27). This modification resembles what recently reported on histidine residues [29] and generally postulated as a result of radical transfer to backbone following oxidation reactions [46-48], although it has never been reported so far for serine and threonine residues. However, this modification is not a prerogative of samples from volcanic areas, since it has been consistently observed also in the ancient bone control samples.

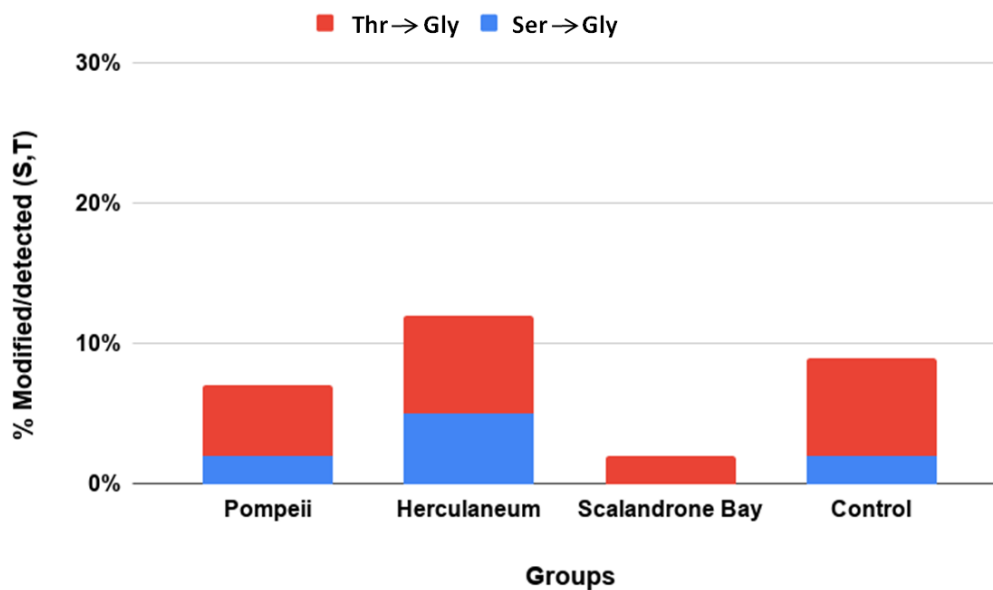


Figure 7: Extent of $\text{C}\alpha\text{-C}\beta$ bond cleavage at serine and threonine reported as percentage of modified over detected (modified plus unmodified) residues.

Proline is a rather complex and often neglected target of chemical modification. The abundance of this residue in collagen, exceeding 20% of the total amino acids in human type I collagen, however, increases the rate of detection of modifications on this peculiar residue, although the abundant and variable incidence of hydroxylation makes detection of any other modification quite challenging (see figure S14 for the occupancy of hydroxylation of proline along the sequences of COL1A1 and COL1A2). It has already been suggested that an increased level of hydroxylated prolines might result from a non-enzymatic oxidation[49]. The peculiar cyclic structure of proline results in an oxidative fate different from that of other aliphatic side chain [50]. Unfortunately, some oxidation products, such as glutamic semialdehyde are isobaric with hydroxylation [50, 51], impairing their unequivocal identification. Nevertheless, consistent formation of pyroglutamic acid from proline ($\Delta M +13.980$ Da) and di and tri-oxidation products ($\Delta M +31.989$ Da and $+47.983$ Da respectively), with di-oxidation that also matches formation of glutamic acid, are eventually suggestive of oxidative diagenetic modification (double hydroxylation is not reported as a physiological post-translational modification) (Fig S21-S24). Most interestingly, a mass shift of $\Delta M -2.001$ Da, consistent with the loss of 2 hydrogens, was repeatedly detected and only in the samples of Pompeii and Herculaneum (Fig. 8). We suggest (fig S15 and fig S25) that this mass shift is attributed to 3,4 dehydro-proline, which is the only stable form of the five possible isomers of olefinic proline [52], and could arise from dehydration of 4-

hydroxyproline or 3-hydroxyproline. From now it will be called Dhp, with a mass shift of -18.001 Da from hydroxyproline and ΔM -2.001 Da from proline.

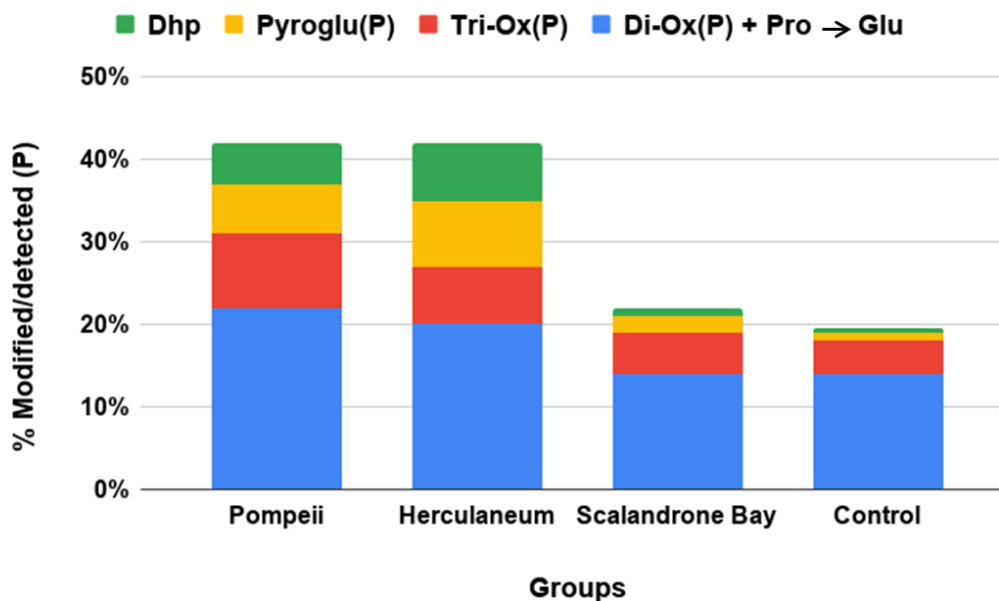


Figure 8: Extent of modified proline residues, reported as percentage of modified over detected (modified plus unmodified) ones.

We also explored the occupancy of the non-enzymatic identified modifications along the sequence of COL1A1 and COL1A2. In general, the distribution of modifications is uneven, with residues with high modification occupancy and sites with low occupancy (Figures S18-S20). However, as far as the glycation products, that are the most striking peculiarity of Pompeii samples, rather interestingly, the G-H1 and MG-H1 modifications seem to be localized in some specific arginine positions, namely positions 564, 574, 1014, 1026 and 1034 of COL1A1, and positions 448, 474, 673 and 691 in COL1A2 (Figure S18). Conversely, in agreement with the observation of a higher average modification of lysines (according to Table S4a), glycation products on lysines seem more spread along the polypeptide chain in Pompeii samples (Figure S19). It is worth mentioning that almost all of the lysine and arginine were actually covered.

More than 65% of the detected histidines in the samples from Pompeii and Herculaneum have been found modified (tableS4d). Figure S20 reports the occupancy of the identified modifications along the sequence of COL1A1 and COL1A2 in all the sample groups. Pompeii and Herculaneum samples behave quite similarly, and histidine 267 in COL1A1 seems to be a rather hot spot for oxidation.

Proline oxidation products are quite spread along the sequences and follow the general trend of samples from Pompeii more modified than Herculaneum ones (in agreement also to the observation that collectively 8% of prolines have been found to be (differently from hydroxylation) modified in samples of Pompeii and Herculaneum (Table S4C), in turn more modified than Baia Scalandrone and Control samples. Position 592 seems to be a hotspot in all the cases.

Discussion

The diagenesis of organic matter in archaeological bone is a complex phenomenon [8]. Lack of cellular proteins, extensive deamidation, backbone cleavage, oxidative chemical modifications, are all markers of taphonomy and diagenesis of organic matter. All these signatures are present in the proteins extracted from the sample bones from human skeletons collected in Pompeii and Herculaneum, molecularly imprinting the lethal effects of pyroclastic surges.

However, bodies in Pompeii experienced a different fate than in Herculaneum and this is also imprinted molecularly. The body flesh of victims in Pompeii disappeared quite slowly if compared with the Herculaneum ones, whose soft tissues were rapidly consumed, so that plaster casts could be obtained by filling the cavities identified in the ash, while in Herculaneum only skeletons were found in the ash deposits, although exceptionally well preserved [20]. The differences detected comparing Herculaneum with Pompeii samples may arise from the fact that the Pompeii victims were affected by the third and fourth surges, while at Herculaneum people were hit by the first surge. As recently showed for the Herculaneum victims [13], local environmental conditions during the eruption such as the peak of maximum temperature of the ash cloud and the time needed for the ash deposit to cool would have likely produced different effects on the bones.

Extensive investigations at the macroscopic level have been carried out in recent years to outline the processes of exposure of bones to heat and fire, but few attempts to characterize the effects at the molecular level. We analysed by a proteomic approach a set of rib bones from skeletons collected in Herculaneum and Pompeii. As a first glance observation, we observed big differences in color change and bone fragility (see fig. S1). Bones from Pompeii are generally white and grey, to be compared to the generally dark color of the Herculaneum bones.

Interestingly, a striking feature of the exactly coeval sets of samples of Herculaneum and Pompeii is the almost complete absence of NCPs in the bones from Pompeii in comparison to those from Herculaneum, thus suggesting an incomplete consumption of the organic matter in the samples from Herculaneum.

The pathway of chemical reactions that break down the proteins within the bioapatite cage appears still fairly mysterious, with proteins normally degrading principally via a combination of two parallel as well as interplaying mechanisms: digestion by microbes and chemical modification/degradation [53], with time, temperature and burial environment all contributing to influence the kinetics of protein decay. For instance, the high acidity of the volcanic soil due to the presence of many copper minerals such as sulfates, oxides, carbonates, and phosphates increases the solubility of hydroxyapatite thus leading more quickly to bone loss [54]. Proteome complexity is generally considered an hallmark of bone degradation, inversely proportional to age, with most of the samples older than 20,000 years containing predominantly and almost exclusively the collagen proteins that benefit from the interaction with the bioapatite cage that protects them from degradation [26]. We can observe that in five out of the seven Pompeii “cooked” bone samples, only collagen was detected, and in the other two samples, beside collagen, only chondroadherin and biglycan were identified. Moreover, the lower NCPs content, the higher deamidation level and, in general, the higher extent of modification of collagen in the bones from Pompeii, all point towards a more degraded state despite the slower consumption of the body flesh.

Despite the higher temperature that the bodies in Herculaneum experienced in comparison to Pompeii, a good number of NCPs was identified in most of the bones. The most common NCPs identified in the Herculaneum samples include Alpha-2-HS-glycoprotein, biglycan, chondroadherin, pigment epithelium derived factor (PEDF), lumican, and prothrombin, in good agreement to proteins identified in ancient bones [53, 55], all proteins that are known to bind collagen or calcium ions. Moreover, it was recently observed that fetuin-A (herein reported as Alpha-2-HS-glycoprotein), a glycoprotein present in the serum, is relatively stable after death [16]. Here we observe that this protein, that prevents mineral precipitation during mineralization process by stabilizing supersaturated mineral solutions upon formation of soluble colloidal nanospheres [55], is among the NCP survivors to the cooking conditions of Herculaneum.

Interestingly, in our samples, also Vitronectin survived quite well, (it was identified in six out of the seven samples from Herculaneum, as frequently as Biglycan). Vitronectin is an abundant multifunctional glycoprotein found in serum, the extracellular matrix, and bone, and is involved in various physiological processes such as cell attachment, spreading, and migration, interacting, among others, also with collagen type I [56].

Only proteins stabilized by the binding to collagen or to the inorganic component of bones survived in Herculaneum, while all other proteins probably decayed rapidly due to the intense heat of the ash deposit. It is worth mentioning that none of the NCPs recently detected by immunological methods in calcined bone

tissue [57] has been identified herein by proteomic approach, while the set of proteins identified is in agreement with those recently identified by similar proteomic approach in rat model bones [58].

However, NCPs were absent in samples from Pompeii. It might be hypothesized that, in the case of bones from Pompeii, where the flesh survived longer than in Herculaneum, proteins underwent a more “normal” degradation process, initially speeded up by the temperature, so that it eventually results as the set of samples with the fewest proteins and most modified.

Oxidative modifications in the sets of volcanic samples are extensive, very close to what expected to occur in a cooking process, which is still a debated question on a molecular basis [46]. Diagenetic increase of AGEs correlates with oxidative conditions [59] and extensive glycation products were observed in the samples from the volcanic area, always more pronounced in samples from Pompeii, likely originating from reactions with the sugars originally present in the extracellular matrix.

Several oxidative processes have been postulated to occur on prolines and hydroxyprolines upon heating [60], according to chemical pathway depicted by Hellwig [46], who predicts hydroperoxides formation from addition of oxygen to radical at the aliphatic side-chain of prolines, as stable intermediates in protein oxidation.

The high incidence of prolines in collagen allowed to highlight the occurrence of oxidative modifications on this peculiar side chain, some of which possibly explained as modifications originating from hydroxyproline (such as that corresponding to a ΔM -2.001 Da when considering proline as the unmodified amino acid). Histidine was herein confirmed as oxidative target among the amino acids and formation of radicals at C- α backbone can also eventually lead to backbone fragmentation [58], thus suggesting an oxidative origin at the basis of the extensive backbone cleavage observed rather than hydrolysis in an environment where water evaporation is expected.

It is also interesting to observe that chemical modifications (although identified throughout the sequence), appear to be more pronounced in specific regions. See Fig. 9, where diagenetic modifications were collectively reported along the collagen sequences, highlighting a different behavior of the samples from the eruptive area in respect to Baia Scalandrone and control samples, that appear clearly less modified and modifications are spread along the sequence. This suggest a strong three-dimensional effect in directing chemical modification events, that will deserve further investigation.

These data do not claim to be conclusive of differences between diagenetic processes in bones from victims in Herculaneum and Pompeii but rather demonstrate that molecular differences exist and can be seen as a perspective on the chemist's approach to read through the processes that alter proteins in bone during burial, the history written in the molecules, a kind of "chemical life history tracer". Despite the intra-samples variability observed, paleoproteomic analyses revealed that diagenetic processes generated by different environmental conditions are significantly reflected in the protein survival and modification. Many puzzles remain. Why proteins survived better in Herculaneum bones, where flesh was immediately consumed and why modifications were more evident in Pompeii bones are only two of the several open questions. Post-mortem microbial attack, from both fungi and bacteria, usually taking place within the first few months post-mortem and considered as a major actor in biological deterioration of bones [8], might well have been altered as well, since also microorganisms experienced the high temperatures of the eruptive disaster and might have been working only later on.

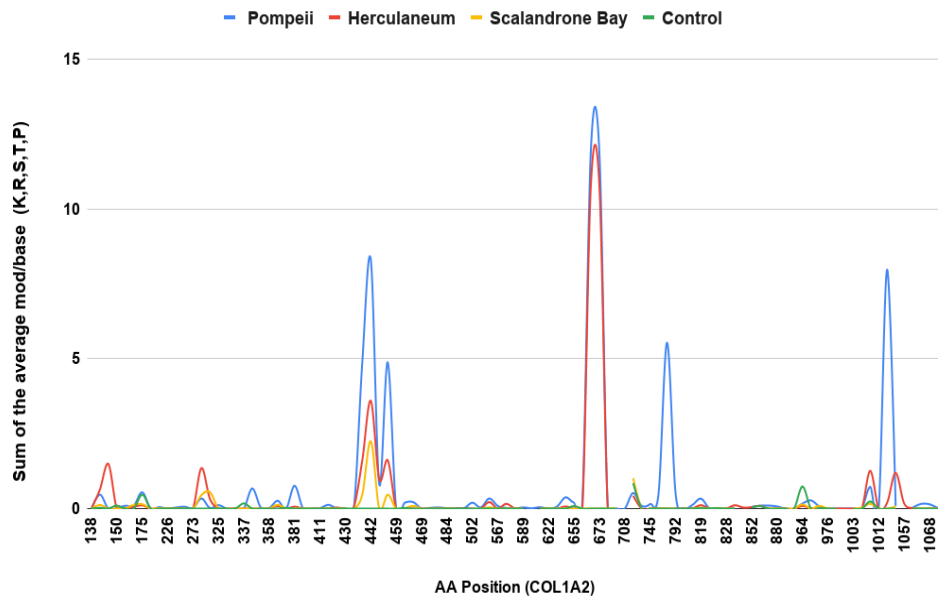
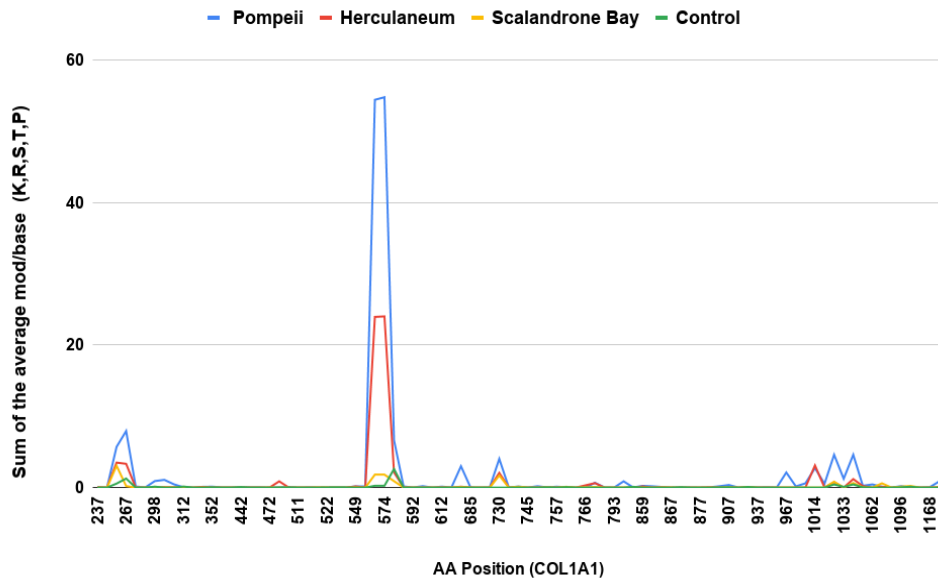


Figure 9: Comparative study of the total “damage signatures” in COL1A1 and COL1A2 of all groups. Control : H162-142 [9] . The figure represents the summary of all average modified/unmodified values regarding K, R S,T and proline modifications(except hydroxyproline) at the specific primary structure positions of COL1A1 and COL1A2.

Author Contributions

G.N, L.B., G.M, E.C. designed research; G.N. I.R.P carried out the experimental research; G.N., I.P.R.,P.P.P, E.C L.B. analysed data; G.N., I.P.R.,P.P.P, E.C L.B. wrote the paper.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska- Curie Grant Agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area).

References

- [1] R. E. M. Hedges, *Archaeometry* **2002**, *44*, 319–328.
- [2] N. L. Martisius, F. Welker, T. Dogandžić, M. N. Grote, W. Rendu, V. Sinet-Mathiot, A. Wilcke, S. J. P. McPherron, M. Soressi, T. E. Steele, *Sci. Rep.* **2020**, *10*, 1–12.
- [3] S. Brown, T. Higham, V. Slon, S. Paabo, M. Meyer, K. Douka, F. Brock, D. Comeskey, N. Procopio, M. Shunkov, A. Derevianko, M. Buckley, *Sci. Rep.* **2016**, *6*, 23559.
- [4] M. Buckley, *PLoS One* **2013**, *8*, DOI 10.1371/journal.pone.0059614.
- [5] I. Mikšík, P. Sedláková, S. Pataridis, F. Bortolotti, R. Gottardo, F. Tagliaro, *Chromatographia* **2014**, *77*, 1503–1510.
- [6] S. Presslee, G. J. Slater, F. Pujos, A. M. Forasiepi, R. Fischer, K. Molloy, M. Mackie, J. V. Olsen, A. Kramarz, M. Taglioretti, F. Scaglia, M. Lezcano, J. L. Lanata, J. Southon, R. Feranec, J. Bloch, A. Hajduk, F. M. Martin, R. Salas Gismondi, M. Reguero, C. de Muizon, A. Greenwood, B. T. Chait, K. Penkman, M. Collins, R. D. E. MacPhee, *Nat. Ecol. Evol.* **2019**, *3*, 1121–1130.
- [7] C. Kendall, A. M. H. Eriksen, I. Kontopoulos, M. J. Collins, G. Turner-Walker, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2018**, *491*, 21–37.
- [8] M. J. Collins, J. Hiller, C. I. Smith, J. P. Roberts, R. V Prigodich, T. J. Wess, A. R. Millard, *Archaeometry* **2002**, *3*, 383–394.
- [9] N. A. Van Huizen, J. N. M. Ijzermans, P. C. Burgers, T. M. Luider, *Mass Spectrom. Rev.* **2019**, *1*, 1–27.
- [10] P. Perez Hurtado, P. B. O'Connor, *Anal. Chem.* **2012**, *84*, 3017–3025.
- [11] L. Orlando, A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen, M. Stiller, M. Schubert, E. Cappellini, B. Petersen, I. Moltke, P. L. F. Johnson, M. Fumagalli, J. T. Vilstrup, M. Raghavan, T. Korneliusen, A. S. Malaspinas, J. Vogt, D. Szklarczyk, C. D. Kelstrup, J. Vinther, A. Dolocan, J. Stenderup, A. M. V. Velazquez, J. Cahill, M. Rasmussen, X. Wang, J. Min, G. D. Zazula, A. Seguin-Orlando, C. Mortensen, K. Magnussen, J. F. Thompson, J. Weinstock, K. Gregersen, K. H. Røed, V. Eisenmann, C. J. Rubin, D. C. Miller, D. F. Antczak, M. F. Bertelsen, S. Brunak, K. A. S. Al-Rasheid, O. Ryder, L. Andersson, J. Mundy, A. Krogh, M. T. P. Gilbert, K. Kjær, T. Sicheritz-Ponten, L. J. Jensen, J. V. Olsen, M. Hofreiter, R. Nielsen, B. Shapiro, J. Wang, E. Willerslev, *Nature* **2013**, *499*, 74–78.
- [12] N. L. Van Doorn, J. Wilson, H. Hollund, M. Soressi, M. J. Collins, *Rapid Commun. Mass Spectrom.* **2012**, *26*, 2319–2327.
- [13] J. Wilson, N. L. Van Doorn, M. J. Collins, *Anal. Chem.* **2012**, *84*, 9041–9048.
- [14] E. Cappellini, L. J. Jensen, D. Szklarczyk, A. Ginolhac, R. A. R. Da Fonseca, T. W. Stafford, S. R. Holen, M. J. Collins, L. Orlando, E. Willerslev, M. T. P. Gilbert, J. V. Olsen, *J. Proteome Res.* **2012**, *11*, 917–926.
- [15] R. Sawafuji, E. Cappellini, T. Nagaoka, A. K. Fotakis, R. R. Jersie-Christensen, J. V Olsen, K. Hirata, S. Ueda, *R. Soc. open Sci.* **2017**, *4*, 161004.

- [16] N. Procopio, A. Williams, A. T. Chamberlain, M. Buckley, *J. Proteomics* **2018**, *177*, 21–30.
- [17] T. P. Cleland, E. R. Schroeter, M. H. Schweitzer, *Proc. R. Soc. B Biol. Sci.* **2015**, *282*, 1–6.
- [18] R. C. Hill, M. J. Wither, T. Nemkov, A. Barrett, A. D'Alessandro, M. Dzieciatkowska, K. C. Hansen, *Mol. Cell. Proteomics* **2015**, *14*, 1946–1958.
- [19] P. Petrone, P. Pucci, A. Vergara, A. Amoresano, L. Birolo, F. Pane, F. Sirano, M. Niola, C. Buccelli, V. Graziano, *PLoS One* **2018**, *13*, 1–27.
- [20] P. Petrone, *J. Anthropol. Sci.* **2019**, *97*, DOI 10.4436/jass.97008.
- [21] P. Petrone, M. Giordano, S. Giustino, F. M. Guarino, *PLoS One* **2011**, *6*, 1–14.
- [22] A. Tressaud, *Fluorine and the Environment*, Elsevier, Amsterdam, Oxford, **2006**.
- [23] G. Mastrolorenzo, P. P. Petrone, M. Pagano, A. Incoronato, P. J. Baxter, A. Canzanella, L. Fattore, *Nature* **2001**, *410*, 769–770.
- [24] G. Mastrolorenzo, P. Petrone, L. Pappalardo, F. M. Guarino, *PLoS One* **2010**, *5*, DOI 10.1371/journal.pone.0011127.
- [25] P. P. Petrone, *Cronache Ercolanesi* **2019**, *49*, 287–299.
- [26] C. Wadsworth, M. Buckley, *Rapid Commun. Mass Spectrom.* **2014**, *28*, 605–615.
- [27] X. Jiang, M. Ye, X. Jiang, G. Liu, S. Feng, L. Cui, H. Zou, *J. Proteome Res.* **2007**, *6*, 2287–2294.
- [28] M. Buckley, C. Wadsworth, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2014**, *416*, 69–79.
- [29] M. Mackie, P. Rütther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D. A. Peggie, H. Howard, L. Harrison, L. J. Jensen, J. V. Olsen, E. Cappellini, *Angew. Chemie - Int. Ed.* **2018**, *57*, 7369–7374.
- [30] E. R. Schroeter, T. P. Cleland, *Rapid Commun. Mass Spectrom.* **2016**, *30*, 251–255.
- [31] M. Pal Chowdhury, R. Wogelius, P. L. Manning, L. Metz, L. Slimak, M. Buckley, *Archaeometry* **2019**, *61*, 1382–1398.
- [32] S. Orsini, A. Yadav, M. Dilillo, L. A. McDonnell, I. Bonaduce, *Anal. Chem.* **2018**, *90*, 6403–6408.
- [33] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, *Microchem. J.* **2016**, *126*, 341–348.
- [34] M. M. Savitski, M. L. Nielsen, R. A. Zubarev, *Mol. Cell. Proteomics* **2006**, *5*, 935–948.
- [35] S. Tyanova, T. Temu, J. Cox, *Nat. Protoc.* **2016**, *11*, 2301–2319.
- [36] D. R. Martin, P. Dutta, S. Mahajan, S. Varma, S. M. Stevens, *Sci. Rep.* **2016**, *6*, 1–12.
- [37] A. Perrone, A. Giovino, J. Benny, F. Martinelli, *Oxid. Med. Cell. Longev.* **2020**, *2020*, DOI 10.1155/2020/3818196.
- [38] N. Rabbani, P. J. Thornalley, *Amino Acids* **2012**, *42*, 1087–1096.
- [39] A. Soboleva, R. Schmidt, M. Vikhnina, T. Grishina, A. Frolov, *Int. J. Mol. Sci.* **2017**, *18*, 2677.
- [40] A. Bachi, I. Dalle-Donne, A. Scaloni, *Chem. Rev.* **2013**, *113*, 596–698.
- [41] D. R. Sell, C. M. Strauch, W. Shen, V. M. Monnier, *Biochem. J.* **2007**, *404*, 269–277.
- [42] L. Gorisse, C. Pietrement, V. Vuiblelet, C. E. H. Schmelzer, M. Köhler, L. Duca, L. Debelle, P. Fornès, S. Jaisson, P. Gillery, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 1191–6.
- [43] S. Chetyrkin, M. Mathis, V. Pedchenko, O. A. Sanchez, W. H. McDonald, D. L. Hachey, H. Madu, D. Stec, B. Hudson, P. Voziyan, *Biochemistry* **2011**, *50*, 6102–6112.
- [44] D. R. Sell, V. M. Monnier, *J. Biol. Chem.* **2004**, *279*, 54173–54184.
- [45] E. Cappellini, F. Welker, L. Pandolfi, J. Ramos-Madrigal, D. Samodova, P. L. Rütther, A. K. Fotakis, D. Lyon, J. V. Moreno-Mayar, M. Bukhsianidze, R. Rakownikow Jersie-Christensen, M. Mackie, A. Ginolhac, R. Ferring, M. Tappen, E. Palkopoulou, M. R. Dickinson, T. W. Stafford, Y. L. Chan, A. Götherström, S. K. S. S. Nathan, P. D. Heintzman, J. D. Kapp, I. Kirillova, Y. Moodley, J. Agusti, R. D. Kahlke, G. Kiladze, B. Martínez-Navarro, S. Liu, M. Sandoval Velasco, M. H. S. Sinding, C. D. Kelstrup, M. E. Allentoft, L. Orlando, K. Penkman, B. Shapiro, L. Rook, L. Dalén, M. T. P. Gilbert, J. V. Olsen, D. Lordkipanidze, E. Willerslev, *Nature* **2019**, *574*, 103–107.
- [46] M. Hellwig, *Angew. Chemie - Int. Ed.* **2019**, 16742–16763.
- [47] H. A. Headlam, A. Mortimer, C. J. Easton, M. J. Davies, *Chem. Res. Toxicol.* **2000**, *13*, 1087–1095.
- [48] H. A. Headlam, M. J. Davies, *Free Radic. Biol. Med.* **2004**, *36*, 1175–1184.

- [49] I. Mikšík, P. Sedláková, S. Pataridis, F. Bortolotti, R. Gottardo, *Protein Sci.* **2016**, *25*, 2037–2044.
- [50] G. Xu, M. R. Chance, *Chem. Rev.* **2007**, *107*, 3514–3543.
- [51] A. Amici, R. L. Levine, L. Tsai, E. R. Stadtman, *J. Biol. Chem.* **1989**, *264*, 3341–3346.
- [52] A. B. Mauger, B. Witkop, *Chem. Rev.* **1966**, *66*, 47–86.
- [53] M. Buckley, C. Wadsworth, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2014**, *416*, 69–79.
- [54] V. Memoli, E. Eymar, C. García-Delgado, F. Esposito, L. Santorufo, A. De Marco, R. Barile, G. Maisto, *Sci. Total Environ.* **2018**, *625*, 16–26.
- [55] W. Jahnen-Dechent, A. Heiss, C. Schäfer, M. Ketteler, *Circ. Res.* **2011**, *108*, 1494–1509.
- [56] D. I. Leavesley, A. S. Kashyap, T. Croll, M. Sivaramakrishnan, A. Shokoohmand, B. G. Hollier, Z. Upton, *IUBMB Life* **2013**, *65*, 807–818.
- [57] R. D. Díaz-Martín, J. R. Ambrosio, R. M. Flores, S. González-Pozos, L. Valencia-Caballero, *Forensic Sci. Int.* **2019**, *305*, DOI 10.1016/j.forsciint.2019.110027.
- [58] E. Johnston, M. Buckley, *J. Proteome Res.* **2020**, DOI 10.1021/acs.jproteome.0c00555.
- [59] J. Wiemann, M. Fabbri, T. R. Yang, K. Stein, P. M. Sander, M. A. Norell, D. E. G. Briggs, *Nat. Commun.* **2018**, *9*, DOI 10.1038/s41467-018-07013-3.
- [60] G. Vistoli, D. De Maddis, A. Cipak, N. Zarkovic, M. Carini, G. Aldini, *Free Radic. Res.* **2013**, *47*, 3–27.

Supporting Information

Samples description


Bone samples collected from 15 individuals unearthed in Pompeii (7), Herculaneum (5) and Baia Scalandrone (3) have been analyzed. Table 1 reports each specimen and related information.

Table S1: Specimens analysed.

Sample	Bone element	Age	Site
PC203	Ribs	79 AD	Pompeii
PC204	Femoral	79 AD	Pompeii
PC205	Skull	79 AD	Pompeii
PC206	Tibia	79 AD	Pompeii
PC207	Scapula	79 AD	Pompeii
PC193	Ribs	79 AD	Pompeii
PC195	Femoral	79 AD	Pompeii
ECC106	Ribs	79 AD	Herculaneum
ECC108	Ribs	79 AD	Herculaneum
ECC109	Ribs	79 AD	Herculaneum
ECC117	Ribs	79 AD	Herculaneum
ECC118	Ribs	79 AD	Herculaneum
BSC180	Ribs	II sec AD	Scalandrone Bay
BSC182	Ribs	II sec AD	Scalandrone Bay
BSC183	Ribs	II sec AD	Scalandrone Bay

Figure S1: Schematic representation of the bones fragments.

Sample Name	Bone	Description on bone integrity	EDTA Fraction
ECC106		Compact in the outer/Crumbly inside	
ECC108		Compact in the outer/Crumbly inside	
ECC109		Compact in the outer/Crumbly inside	
ECC117		Crumbly & flaky in both parts	
ECC118		Crumbly & flaky in both parts	
PC193		Friable, crumbly	
PC195		Compact in the outer/ friable inside	
PC203		Friable, crumbly	
PC204		Porous & Crumbly	

PC205		Friable, crumbly	
PC206		Friable, crumbly	
PC207		Friable, crumbly	
BSC180		Crumbly, friable	
BSC182		Compact in the outer , crumbly inside	
BSC183		Compact	

Experimental procedures

Protein extraction and digestion

Samples were prepared as reported in [1] with slight modifications. Figure S2 represents the whole procedure.

Briefly, in order to prevent contamination, surfaces were removed from bone samples with sandpaper. Samples (30–50 mg) were wrapped in clean aluminum foil and fragmented into powder using a conventional hammer. The hammer was cleaned with bleach and ethanol, repeatedly. 300 μ l of 0.5 M EDTA (pH 8.0) was added to the powdered bones, followed by incubation at 4 °C for 24–48 h with agitation. Every 24h fresh EDTA was added after the samples had been spun down and the supernatants were separated from the pellet (EDTA fraction). Subsequently, pellets were washed three times with 100 μ l 0.1 M Tris (pH 8.0) and the supernatants (Tris fraction) were combined to the previous fraction (EDTA fraction). The EDTA and Tris fraction were concentrated on an Amicon 3000 Da MWCO spin column, the flow through was discarded, and the retained sample was washed with 3ml of guanidinium hydrochloride (GuHCl) 2M extraction solution, 0.5M Tris (2-carboxyethyl) phosphine, 0.5M chloroacetamide, 100 mM Tris (pH 8.0) and water. Finally, the membrane was washed with 300 μ l of GuHCl 2M extraction solution that was added to the protein pellet. The pH was adjusted to 8.0 and the mixture of proteins was heated at 80°C for 3 hours under agitation and cooled to room temperature. Without separating supernatants and the residual bone powder, samples were heated at 80 °C for 2 h, and cooled to room temperature LysC-Trypsin mix, 1/100 by amount of protein, was added to the samples for protein digestion. Mixed samples were incubated at 25 °C for 30 min then diluted to 2 M GuHCl with 25 mM Tris (pH 8.0), followed by incubation at 37 °C overnight under agitation. Digestion was terminated with 10% trifluoroacetic acid to a final concentration of 1%. After centrifugation at 14 000 g for 10 min, the tryptic peptides in the supernatant were immobilized on C18 stage tips as previously described [2].

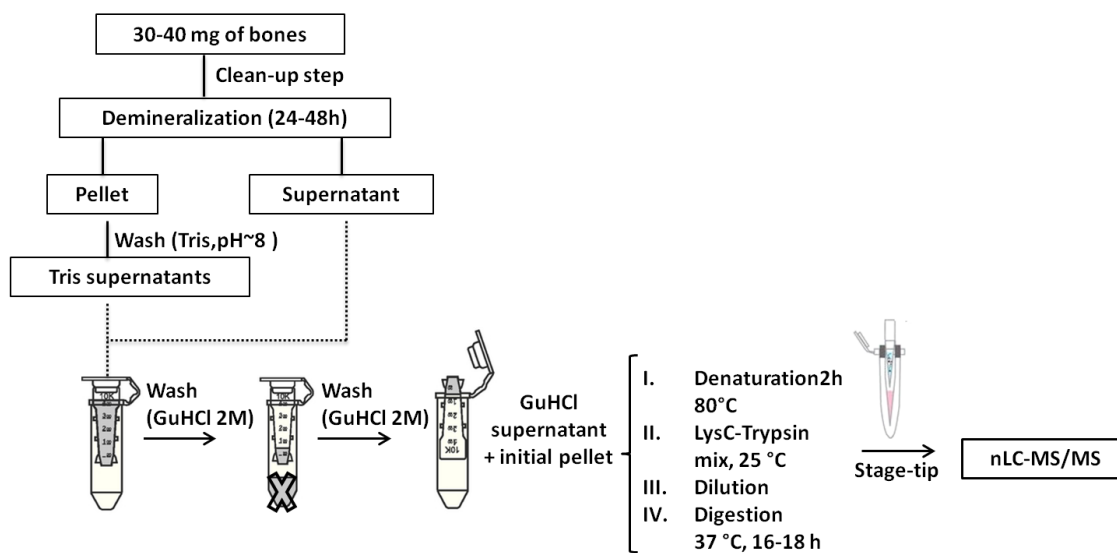


Figure S2: Schematic representation of the workflow for protein extraction and digestion from the bone samples.

nLC-MSMS

Samples were eluted from the stage tips using 20 μ L 40% ACN and then 10 μ L 60% ACN into a 96 well MS plate. Samples were concentrated under vacuum at 40°C up to approximately 3 μ L and then 5 μ L of 0.1% TFA, 5% ACN were added.

Samples were then separated on a 15 cm column (75 μ m inner diameter) in-house laser pulled and packed with 1.9 μ m C18 beads (Dr. Maisch, Germany) on an EASY-Nlc 1000 (Proxeon, Odense, Denmark) connected to a Q-Exactive HF (Thermo Scientific, Bremen, Germany) on a 77 min gradient. Buffer A was milliQ water. The peptides were separated with increasing buffer B (80% ACN and 0.1% formic acid), going from 5% to 30% in 50 min, 30% to 45% in 10 min, 45% to 80% in 2 min, held at 80% for 5 min before dropping back down to 5% in 5 min and held for 5 min. Flow rate was 250 nL/min. The column temperature was maintained at 40°C using an integrated column oven. A wash-blank method using 0.1% TFA, 5% ACN was run in between each sample to hinder cross contamination.

The Q-Exactive HF was operated in data dependent top 10 mode. Spray voltage was 2 kV, S-lens RF level at 50, and heated capillary at 275°C. Full scan mass spectra were recorded at a resolution of 120,000 at m/z 200 over the m/z range 350–1400 with a target value of 3e6 and a maximum injection time of 25 ms. HCD-

generated product ions were recorded with a maximum ion injection time set to 108 ms and a target value set to $2e5$ and recorded at a resolution of 60,000. Normalized collision energy was set at 28% and the isolation window was 1.2 m/z with the dynamic exclusion set to 20 s.

Data analysis

The resulting raw files (EvoG_sample name, in total 15 files) were searched and analysed using the MaxQuant (MQ) software [3] against a UniProt database (759,512 sequences, 37,179,137 residues) with *Homo sapiens* as taxonomic restriction (20199 sequences, 928,813 residues). Initially standard searches (1-3, table S2) were performed, tolerances were those preset for Orbitrap, using a tryptic search with up to two missed cleavages. Minimum peptide length was set to 7. Carbamidomethylation was set as a fixed modification, while methionine oxidation, hydroxylation of proline, Gln-pyro Glu-pyro-Glu- were set as variable modifications, up to a maximum of 5 modifications per peptide. Protein identifications were supported by a false discovery rate (FDR) of 0.01 applied (same FDR for dependent peptides when applied) and manually filtered by at least 2 different non overlapping peptides above 70 ion score threshold. Contaminant proteins were assessed using the contamination.fasta provided by MQ which includes common laboratory contaminants [3] (see MaxQuant Downloads -contaminants.fasta,) can be found under http://www.coxdocs.org/doku.php?id=maxquant:start_downloads.htm, n.d. These protein hits were excluded from further analysis. MQ searches were also carried out on raw data from two archaeological Japanese bones samples that had been processed as herein samples from Pompeii and Herculaneum [1] (original work reference: samples H-142 and H-162, in the work by Sawafuji *et al.* [1]), but completely unrelated to the bones from Vesuvius area and much younger (Hitotsubashi site AD 1657-1683) to compare and assess diagenetically induced chemical modifications in our samples.

In search of non-enzymatic chemical modifications we utilized the concept of “dependent peptides” for unbiased, comprehensive analysis of diagenetically induced chemical modifications in collagen proteins. The “dependent peptides” are modified peptides identified in an unrestricted database search against a library of previously identified non-modified “base peptides”, a peptide search feature integrated in MaxQuant with statistical analysis for validation of modified peptides [4][5]. This strategy was already demonstrated successful with painting samples [6] and *in vitro* metal ion induced oxidation [7]. In a separate run (run 4, table S2) all samples including the control samples (original work reference: samples H-142 and H-162, in the work by Sawafuji *et al.* [1]) were re-analyzed and mass shifts were ranked by their occurrence within the dataset and grouped on the amino acid on which occurred. The modifications were firstly chosen after filtering with localization probabilities of $\geq 80\%$ for modified peptides and occurrence of detection of DP

Cluster Mass ≥ 5 times for each sample. Afterwards, alleged statistically relevant modifications for which a plausible chemical interpretation was given (“UNIMOD, protein modifications for mass spectrometry” http://www.unimod.org/modifications_list.php) were subsequently searched in separate runs , as group datasets (runs 5-24, table S2) as variable modifications in order to map selected PTMs on all identified as further confirmation. To reduce the search space, we restricted the database to the most relevant identified protein sequences, i.e. type I collagen chains, namely Collagen alpha-1(I) chain and Collagen alpha-2(I), which are, as expected, largely overrepresented, with more than 58% of the total assigned spectra on average. Moreover, type I collagen chains $\alpha 1$ and $\alpha 2$ are the two polypeptide chains identified in all the samples. Backbone cleavage was searched against all groups in separate runs (25-28), inserting the standard parameters of the runs 1-3 and semitrypsin as the enzyme in the MQ searches.

Table S2: Details of the Different MaxQuant (MQ) search runs and data analysis performed in this study.

MQ search	Details on specific Parameters	Aim
1-3	Standard search in the UniProt human protein database (759,512 sequences, 37,179,137 residues) with Homo sapiens as taxonomic restriction (20199 sequences, 928,813 residues) to identify proteins in Pompeii, Herculaneum and Scalandrone bay samples in separate runs (run 1-3). Hydroxylation of proline and lysine, deamidation of glutamine and asparagine, oxidation of methionine, glutamine and glutamic acid to pyroglutamate were considered as variable modifications, carbamidomethylation of cysteine was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides, dependent peptides search	1- Initial discovery of any common proteins found in human bones 2- Untargeted discovery of AA substitutions and chemical modifications 3- Evaluation of the PTMs frequency of detection
4	Addition of the two control archaeological bone samples (Control) in a separate run (run 4) in the UniProt human protein database (759,512 sequences, 37,179,137 residues). with Homo sapiens as taxonomic restriction (20199 sequences, 928,813 residues). Hydroxylation of proline and lysine, deamidation of glutamine and asparagine, oxidation of methionine glutamine and glutamic acid to pyroglutamate were considered as variable modifications, carbamidomethylation of cysteine was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides. dependent peptides search	Comparison of PTMs among Pompeii, Herculaneum and Scalandrone Bay with the control samples
5-8	Analysis of all archaeological bone samples, including controls with a database of the confidently identified <i>Homo sapiens</i> collagen proteins(163 sequences 176,326	Discovery of (K) AGEs

	residues) in separate runs (run 6-10). Hydroxylation of proline, deamidation of glutamine and asparagine, carbamylation, formylation, carboxyethylation, carboxymethylation of lysine, and conversion of lysine to aminoadipic acid were considered as variable modifications, carbamidomethylation of cysteine was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides. Maximum 4 modifications per peptide.	
9-12	Analysis of all archaeological bone samples, including controls with a database of the confidently identified <i>Homo sapiens</i> collagen proteins (163 sequences 176,326 residues) in separate runs (run 11-15). Hydroxylation of proline, deamidation of glutamine and asparagine, formation of glyoxal, methylglyoxal at arginine, and conversion of arginine to ornithine were considered as variable modifications, carbamidomethylation of cysteine was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides. Maximum 4 modifications per peptide	Discovery of (R) AGEs
13-16	Analysis of all archaeological bone samples, including controls with a database of the confidently identified <i>Homo sapiens</i> collagen proteins (163 sequences 176,326 residues) in separate runs (run 11-15). Hydroxylation of proline, deamidation of glutamine and asparagine, mono-oxidation and bi-oxidation of histidine, and conversion of histidine to aspartic acid were considered as variable modifications, carbamidomethylation of cysteine was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides. Maximum 4 modifications per peptide	Discovery of (H) modifications
17-20	Analysis of all archaeological bone samples, including controls with a database of the confidently identified <i>Homo sapiens</i> collagen proteins (163 sequences 176,326 residues) in separate runs (run 16-20). Hydroxylation of proline, deamidation of glutamine and asparagine, proline oxidation to pyroglutamic acid, proline di-oxidation, tri-oxidation and dehydration of hydroxyproline (Dhp) were considered as variable modifications, carbamidomethylation was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides. Maximum 4 modifications per peptide.	Discovery of (P) modifications

21-24	Analysis of all archaeological bone samples, including controls with a database of the confidently identified <i>Homo sapiens</i> collagen proteins (163 sequences 176,326 residues) in separate runs (run 11-15). Hydroxylation of proline, deamidation of glutamine and asparagine, oxidation of methionine and conversion of serine and threonine to glycine were considered as variable modifications, carbamidomethylation was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides. Maximum 4 modifications per peptide	Discovery of (ST) modifications
25-27	Back bone cleavage (Semitryptic analysis) search in the UniProt human protein database (759,512 sequences, 37,179,137 residues) with <i>Homo sapiens</i> as taxonomic restriction (20199 sequences, 928,813 residues) to identify proteins in Pompeii, Herculaneum and Scaladrone bay samples in separate runs (run 25-27). Hydroxylation of proline and lysine, deamidation of glutamine and asparagine, oxidation of methionine, glutamine and glutamic acid to pyroglutamate were considered as variable modifications, carbamidomethylation was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides,	Back bone cleavage search
28	Back bone cleavage (semitryptic analysis) search of the two control archaeological bone samples (Control) in a separate run (run 28) in the UniProt human protein database (759,512 sequences, 37,179,137 residues) with <i>Homo sapiens</i> as taxonomic restriction (20199 sequences, 928,813 residues). Hydroxylation of proline and lysine, deamidation of glutamine and asparagine, oxidation of methionine glutamine and glutamic acid to pyroglutamate were considered as variable modifications, carbamidomethylation was set as fixed modification; ion score cut off ≥ 70 for unmodified and modified peptides.	Back bone cleavage search

Evaluation of diagenetically induced chemical modifications.

Deamidation of asparagine (N) and glutamine (Q), oxidation of methionine (M), and hydroxylation of proline (P) and lysine (K) were evaluated as variable modifications in standard MaxQuant searches. Additionally, MaxQuant's "evidence.txt" file was used to calculate separate deamidation rates for N and Q for individual samples with a code that is freely available on GitHub (<https://github.com/dblyon/deamidation>). Also using both "evidence.txt" and "peptides.txt", positional evaluation of the relative deamidation (N, Q) and oxidation (M) was performed using code available on <https://github.com/ismaRP/MSMSdeamidation>. For

each N or Q position, the values are calculated dividing the sum of intensities of the peptides containing the modification by the total intensity of the peptides containing that position.

Backbone cleavage was assessed in all samples (runs 25-28, table S2) searching for semitryptic peptides allowing for one non specific end by setting semitrypsin as enzyme in the MQ searches. Search for semitryptic peptides was carried out on collagen type I chains only, since they are the only polypeptide chains shared among all the samples. In terms of accuracy, only the peptides with the MSMS count more than 1 and one unspecific cleavage site were considered. The frequency of semitryptic peptides was evaluated by calculating the percentage of semitryptic peptides over the total number of identified peptides for each chain, counting peptide to spectrum matches (PSMs) from the peptides.txt files of MQ [2, 7]. Subsequently, manual alignment of all the semitryptic peptides in the five different groups of collagen type I sequences was performed in search of collagen regions that are most frequently hydrolyzed. The cleavage frequency was then re-evaluated considering the regions rather than the single peptide bonds (window of ± 2 AA) by calculating the number of PSMs with semitryptic cleavages identified in a region divided by the total PSMs in the same region, including both tryptic and semitryptic matches.

Non-enzymatic chemical modifications were initially evaluated by manual inspection of all MS/MS spectra, thus also allowing to confidently assess the site localization of the selected chemical modifications. The relative abundance of a mass shift associated with a type of amino acid (i.e. +31.989(P)) was calculated for each group, as an average of the modified amino acids normalized for the sum of all the corresponding detected amino acids in the primary structures of COL1A1 and COL1A2. The number of the modified positions was extracted from the modificationSites.txt files of MQ excluding potential false positives without an unmodified counterpart and excluding overlap (each primary structure position was considered once) although the number of the unmodified positions from the evidence.txt files. On the other hand, the modification occupancy at a specific primary structure position for each modification was calculated from the mod/base values extracted from the modification(X)site.txt files of MaxQuant [8]. Ratios of the modified: non-modified peptides were determined for each sample and then averaged for each group. These average ratios were then used to determine the significance of each modification in the volcanic samples in comparison to the average ratios of the control samples. Finally, in order to have a total vision of collagen “damage” among the different groups the summary of the average ratios modified/unmodified (K,R,P,S,T) was calculated along the sequence of COL1A1 & COL1A2 .

Results

Protein Identification

Table S3: Proteins identified in the bone samples from Pompeii, Herculaneum and Scalandrone Bay archeological sites. Proteins were identified by searching by MAXQuant the UniProt protein database (759,512 sequences, 37,179,137 residues) with *Homo sapiens* as taxonomic restriction (20199 sequences, 928,813 residues). Carbamidomethylation of Cys was set as fixed modification; hydroxylation of proline and lysine, deamidation of glutamine and asparagine, oxidation of methionine, glutamine and glutamic acid to pyroglutamate were considered as variable modifications; ion score cut off ≥ 70 for unmodified and modified peptides. Proteins were considered as identified only when 2 or more peptides have been detected.

Sample	Protein	UniProt entry	Gene name	Razor + Unique peptides	Unique peptides	Sequence coverage (%)	Sequence length
PC203	Collagen alpha-1(I) chain	P02452	COL1A1	58	57	47.6	1464
	Collagen alpha-2(I) chain	P08123	COL1A2	57	57	52.2	1364
PC204	Collagen alpha-1(I) chain	P02452	COL1A1	80	78	53.7	1464
	Collagen alpha-2(I) chain	P08123	COL1A2	65	65	57.5	1364
PC205	Collagen alpha-1(I) chain	P02452	COL1A1	61	59	43.2	1464
	Collagen alpha-2(I) chain	P08123	COL1A2	58	58	53.6	1364
PC206	Collagen alpha-2(I) chain	P08123	COL1A2	37	37	41.3	1364
	Collagen alpha-1(I) chain	P02452	COL1A1	39	38	37.4	1464
	Collagen alpha-1(II) chain	P02458	COL2A1	2	2	3.5	1418

PC207	Collagen alpha-2(I) chain	P08123	COL1A2	62	62	53.4	1364
	Chondroadherin	O15335	CHAD	3	3	11.4	359
	Collagen alpha-1(I) chain	P02452	COL1A1	89	88	55.5	1464
	Biglycan	P21810	PGS1	3	3	10.9	368
PC193	Collagen alpha-2(I) chain	P08123	COL1A2	93	93	65.5	1364
	Collagen alpha-1(I) chain	P02452	COL1A1	114	112	59.9	1464
	Chondroadherin	O15335	CHAD	5	5	17.8	359
	Collagen alpha-2(V) chain	P05997	COL5A2	6	6	7.4	1499
	Collagen alpha-1(II) chain	P02458	COL2A1	3	3	4.2	1418
PC195	Collagen alpha-2(I) chain	P08123	COL1A2	30	30	36.7	1364
	Collagen alpha-1(I) chain	P02452	COL1A1	31	30	30.5	1464
ECC106	Collagen alpha-1(I) chain	P08123	COL1A2	16	16	19.4	1464
	Collagen alpha-2(I) chain	P02452	COL1A1	11	11	17.1	1364
	Hemoglobin subunit beta	P68871	HBB	2	1	25.6	90
	Pigment epithelium-derived factor	P36995	SERPINF1	2	2	5.3	418

	Prothrombin	P00734	F2	2	2	1.9	583
ECC108	Collagen alpha-2(I) chain	P08123	COL1A2	82	82	58.7	1364
	Collagen alpha-1(I) chain	P02452	COL1A1	106	103	58	1464
	Pigment epithelium-derived factor	P36995	SERPINF1	19	19	53.1	418
	Chondroadherin	O15335	CHAD	16	16	44.6	359
	Biglycan	P21810	BGN	15	15	41.6	368
	Alpha-2-HS-glycoprotein	P02765	AHSG	5	5	25.8	368
	Prothrombin	P00734	F2	11	11	20.8	583
	Vitronectin	P04004	VTNC	7	7	16.3	478
	Lumican	P51884	LUM	4	4	15.1	338
	Antithrombin-III	P01008	SERPINC1	3	3	14.7	464
	Decorin	P07585	DCN	5	5	14.2	359
	Periostin	Q15063	POSTN	4	4	11.4	808
	Osteomodulin	Q99983	OMD	3	3	9.7	421
	Protein Z-dependent protease	Q9UK55	SERPINA10	2	2	9.5	484

	inhibitor						
	Kininogen-1	P01042	KNG1	3	3	7.3	427
	Collagen alpha-2(V) chain	P05997	COL5A2	3	3	6.5	1112
	Olfactomedin-like protein 3	Q9NRN5	OLFML3	2	2	6.2	406
	Collagen alpha-1(II) chain	P02458	COL2A1	2	2	4.5	1418
	Coagulation factor IX	P00740	F9	2	2	3.5	423
	Collagen alpha-1(XXII) chain	Q8NFW1	COL22A1	3	3	3	1319
	Collagen alpha-1(XII) chain	Q99715	COL12A1	6	6	2.5	3062
	Collagen alpha-1(III) chain	P02461	COL3A1	2	2	2.3	1466
	NHS-like protein 2	Q5HYW2	NHSL2	2	2	0.8	950
	Basement membrane-specific heparan sulfate proteoglycan core protein	P98160	HSPG2	2	2	0.8	4391
ECC109	Collagen alpha-2(I) chain	P08123	COL1A2	77	77	61.8	1364
	Collagen alpha-1(I) chain	P02452	COL1A1	98	96	57.2	1464
	Chondroadherin	O15335	CHAD	5	5	22.3	359

	Biglycan	P21810	BGN	4	4	13	368
	Collagen alpha-2(V) chain	P05997	COL5A2	3	3	5.8	1112
	Vitronectin	P04004	VTNC	2	2	4.8	478
ECC117	Collagen alpha-1(I) chain	P02452	COL1A1	94	92	58	1464
	Collagen alpha-2(I) chain	P08123	COL1A2	63	63	51.5	1364
	Chondroadherin	O15335	CHAD	15	15	42.6	359
	Pigment epithelium-derived factor	P36995	SERPINF1	10	10	33.5	418
	Biglycan	P21810	PGS1	13	13	28.5	368
	Alpha-2-HS-glycoprotein	P02765	AHSG	3	3	21.2	368
	Osteocalcin	P02818	BGLAP	2	2	20	100
	Protein Z-dependent protease inhibitor	Q9UK55	SERPINA10	6	6	16.9	484
	Prothrombin	P00734	F2	9	9	16.3	583
	Vitronectin	P04004	VTNC	7	7	16.3	478
	Lumican	P51884	LUM	4	4	15.1	338

	Antithrombin-III	P01008	SERPINC1	3	3	14.7	464
	Osteomodulin	Q99983	OMD	3	3	8.1	421
	Olfactomedin-like protein 3	Q9NRN5	OLFL3	2	2	6.2	406
	Kininogen-1	P01042	KNG1	2	2	4.9	427
	Coagulation factor IX	P00740	F9	3	3	4.7	423
	Collagen alpha-1(XII) chain	Q8NFW1	COL22A1	2	2	2	1319
ECC118	Collagen alpha-2(I) chain	P08123	COL1A2	78	78	61.4	1364
	Collagen alpha-1(I) chain	P02452	COL1A1	98	97	56	1464
	Chondroadherin	O15335	CHAD	14	14	41.5	359
	Pigment epithelium-derived factor	P36995	SERPINF1	9	9	29.2	418
	Biglycan	P21810	PGS1	9	9	25.5	368
	Alpha-2-HS-glycoprotein	P02765	AHSG	4	4	23.9	368
	Lumican	P51884	LUM	4	4	15.1	338
	Antithrombin-III	P01008	SERPINC1	3	3	14.7	464
	Vitronectin	P04004	VTNC	6	6	12.6	478

	Decorin	P07585	DCN	4	4	10.9	359
	Prothrombin	P00734	F2	4	4	7.9	583
	Collagen alpha-2(V) chain	P05997	COL5A2	4	4	7.7	1112
	Osteomodulin	Q99983	OMD	2	2	7.4	421
BSC180	Collagen alpha-2(I) chain	P08123	COL1A2	47	47	51.5	1364
	Collagen alpha-1(I) chain	P02452	COL1A1	56	56	45.4	1464
BSC182	Collagen alpha-2(I) chain	P08123	COL1A2	67	67	56.7	1364
	Collagen alpha-1(I) chain	P02452	COL1A1	88	87	52.7	1464
	Vitronectin	P04004	VTNC	2	2	3.8	478
BSC183	Collagen alpha-1(I) chain	P02452	COL1A1	77	76	55.7	1464
	Collagen alpha-2(I) chain	P08123	COL1A2	67	67	53.1	1364
	Chondroadherin	O15335	CHAD	17	17	48.2	359
	Biglycan	P21810	PGS1	14	14	42.7	368
	Alpha-2-HS-glycoprotein	P02765	AHSG	6	6	28.8	368
	Pigment epithelium-derived factor	P36995	SERPINF1	10	10	26.6	418

	Protein Z-dependent protease inhibitor	Q9UK55	SERPINA10	6	6	19.6	484
	Vitronectin	P04004	VTNC	12	12	19.5	478
	Coagulation factor VII	P08709	F7	4	4	16.8	382
	Antithrombin-III	P01008	SERPINC1	4	4	16.8	464
	Prothrombin	P00734	F2	7	7	16.1	583
	Lumican	P51884	LUM	4	4	15.1	338
	Decorin	P07585	DCN	4	4	13.6	359
	Collagen alpha-1(II) chain	P02458	COL2A1	11	11	12.6	1418
	Kininogen-1	P01044	PNG1	3	3	10.5	427
	Collagen alpha-1(IX) chain	P20849	COL9A1	2	2	9.5	328
	Osteomodulin	Q99983	OMD	2	2	7.4	421
	Collagen alpha-2(XI) chain	P13942	COL11A2	9	9	6.7	1623
	Fibronectin	P02751	FN1	8	8	6.6	2031
	Cartilage oligomeric matrix protein	P49747	COMP	4	4	6.1	724

	Matrilin-3	O15232	MATN3	3	3	5.4	444
	Coagulation factor IX	P00740	F9	3	3	4.7	423
	Collagen alpha-1(X) chain	Q03692	COL10A1	3	3	3.7	680
	Collagen alpha-1(XI) chain	P12107	COL11A1	3	3	2.8	1690
	Collagen alpha-1(XII) chain	Q99715	COL12A1	4	4	2.1	3062
	Collagen alpha-1(XXII) chain	Q8NFW1	COL22A1	2	2	2	1319
	Basement membrane-specific heparan sulfate proteoglycan core protein	P98160	HSPG2	6	6	1.9	4391

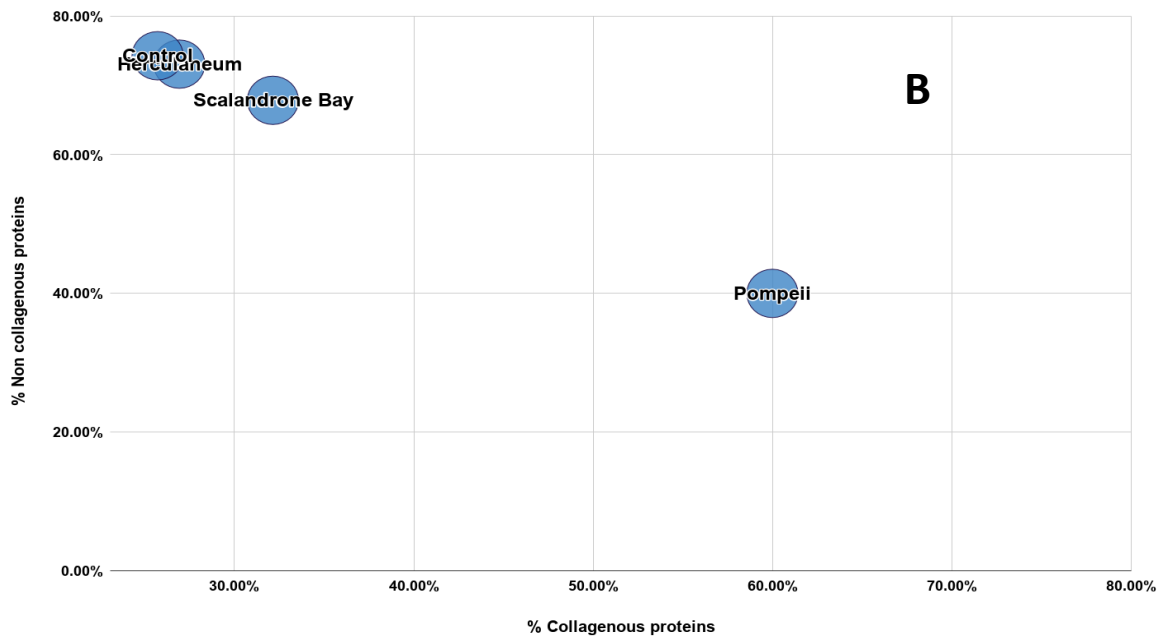
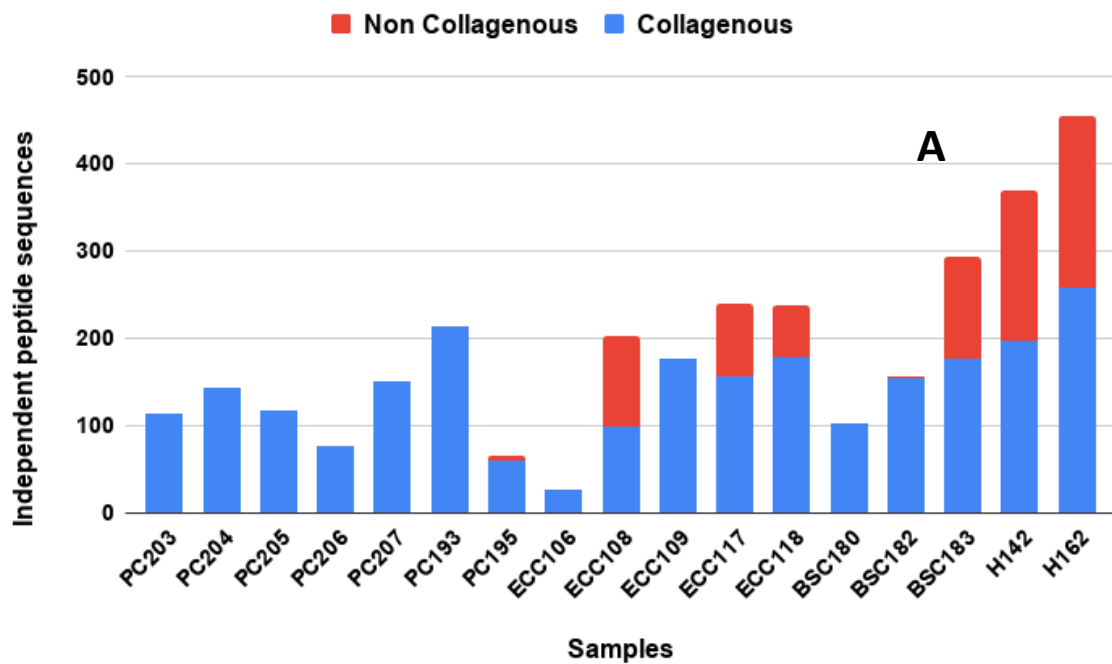


Figure S3: A) Number of independent peptides of collagenous (blue) and non-collagenous (red) proteins in each single samples. B) Grouping samples with respect to collagenous and non collagenous protein content for each group.

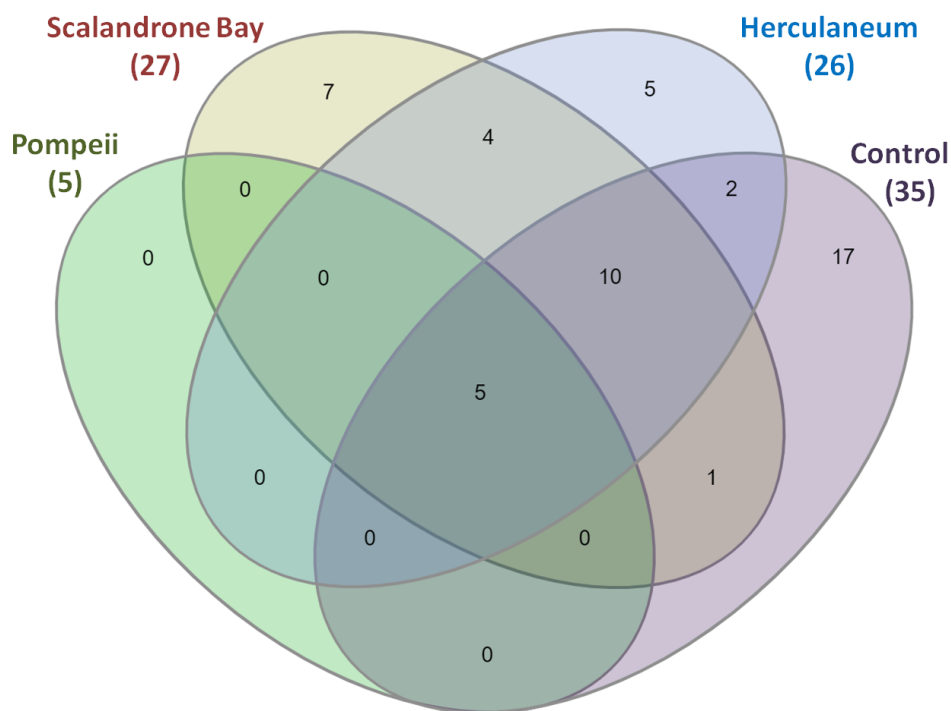
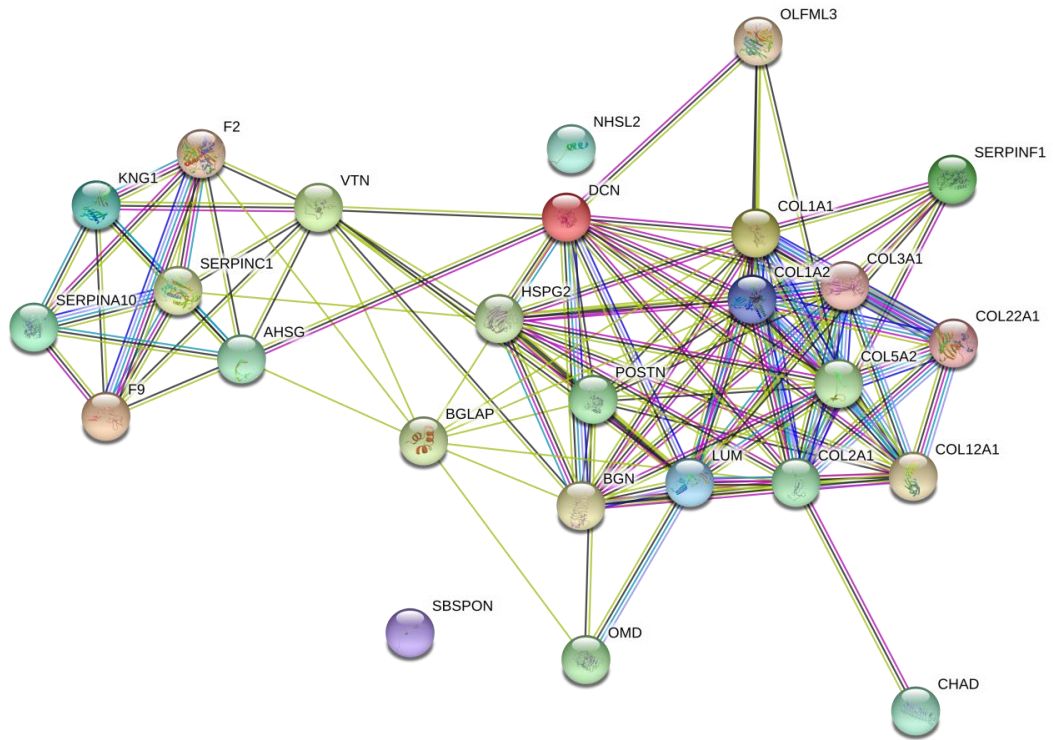


Figure S4: Venn diagram of proteins that are shared among the groups of Pompeii, Herculaneum , Scalandrone Bay and control (H-162, H-142 from [1]) (<http://www.interactivenn.net/>).

Table S4: Unique and shared proteins between the different sample groups.

Group	Number of Proteins	Unique proteins
Herculaneum	5	Hemoglobin subunit beta, Periostin, Collagen alpha1(III) chain, NHSlike protein 2, Osteocalcin
Scalandrone Bay	7	Collagen alpha-1(IX) chain, Collagen alpha-2(XI) chain, Fibronectin, Cartilage oligomeric matrix protein, Matrilin-3, Collagen alpha-1(X) chain, Collagen alpha-1(XI) chain
Control	17	Collagen alpha-3(VI) chain, Apolipoprotein E, Coagulation factor X, Insulin-like growth factor II, Cathepsin G, Matrix Gla protein, Eosinophil peroxidase, Bone marrow proteoglycan, Collagen alpha-1(V) chain, Insulin-like growth factor-binding protein 5, Neutrophil defensin 3, Sushi repeat-containing protein SRPX, Collagen alpha-1(VII) chain, Dermatotontin
Groups	Number of Proteins	Proteins in common
Herculaneum Pompeii Scalandrone Bay Control	5	Collagen alpha-2(I) chain, Chondroadherin, Collagen alpha-1(I) chain, Collagen alpha-1(II) chain, Biglycan

Herculaneum Scalandrone Bay	20	Alpha-2-HS-glycoprotein, Antithrombin-III, Basement membrane-specific heparan sulfate proteoglycan core protein, Biglycan, Chondroadherin, Coagulation factor IX, Collagen alpha-1(I), Collagen alpha-1(II), Collagen alpha-1(XII), Collagen alpha-1(XXII), Collagen alpha-2(I), Collagen alpha-2(V), Decorin, Kininogen-1, Lumican, Osteomodulin, Pigment epithelium-derived factor, Protein Z-dependent protease inhibitor, Prothrombin, Vitronectin
Herculaneum Pompeii	6	Collagen alpha-1(I), Collagen alpha-2(I), Collagen alpha-1(II), Chondroadherin, Biglycan, Collagen alpha-2(V)
Herculaneum Control	20	Alpha-2-HS-glycoprotein, Basement membrane-specific heparan sulfate proteoglycan core protein, Biglycan, Chondroadherin, Coagulation factor IX, Collagen alpha1(I) chain, Collagen alpha1(II) chain, Collagen alpha1(XII) chain, Collagen alpha1(XXII) chain, Collagen alpha2(I) chain, Collagen alpha2(V) chain, Decorin, Kininogen, Lumican, Olfactomedin like protein 3, Osteomodulin, Pigment epithelium derived factor, Protein Z dependent protease inhibitor, Prothrombin, Vitronectin
Scalandrone Bay Control	20	Alpha-2-HS-glycoprotein, Basement membrane-specific heparan sulfate proteoglycan core protein, Biglycan, Chondroadherin, Coagulation factor IX, Coagulation factor VII, Collagen alpha-1(I), Collagen alpha-1(II), Collagen alpha-1(XII), Collagen alpha-1(XXII), Collagen alpha-2(I), Decorin, Kininogen-1, Lumican, Osteomodulin, Pigment epithelium-derived factor, Protein Z-dependent protease inhibitor, Prothrombin, Vitronectin, Coagulation factor-VII



Abbreviation	Protein Name
POSTN	Periostin
VTN	Vitronectin
HBB	Hemoglobin subunit beta
BGLAP	Osteocalcin
BGN	Biglycan
SERPINC1	Antithrombin-III
DCN	Decorin
SERPINF1	Pigment epithelium-derived factor
COL1A1	Collagen alpha-1(I) chain
COL1A2	Collagen alpha-2(I) chain
COL3A1	Collagen alpha-1(III) chain
COL2A1	Collagen alpha-1(II) chain

CHAD	Chondroadherin
COL22A1	Collagen alpha-1(XXII) chain
COL5A2	Collagen alpha-2(V) chain
OMD	Osteomodulin
COL12A1	Collagen alpha-1(XII) chain
KNG1	Kininogen-1
LUM	Lumican
F2	Prothrombin
HSPG2	Basement membrane-specific heparan sulfate proteoglycan core protein
AHSG	Alpha-2-HS-glycoprotein
F9	Coagulation factor IX
ITGB1	Integrin beta-1
OLFML3	Olfactomedin-like protein 3
NHSL2	NHS-like protein 2
SERPINA10	Protein Z-dependent protease inhibitor

Figure S5: STRING interaction pathway of bone proteins. Functional association network of identified proteins in the archaeological bones of Pompeii, Herculaneum and Scaladrome Bay (<https://string-db.org/>).

Deamidation (N,Q)

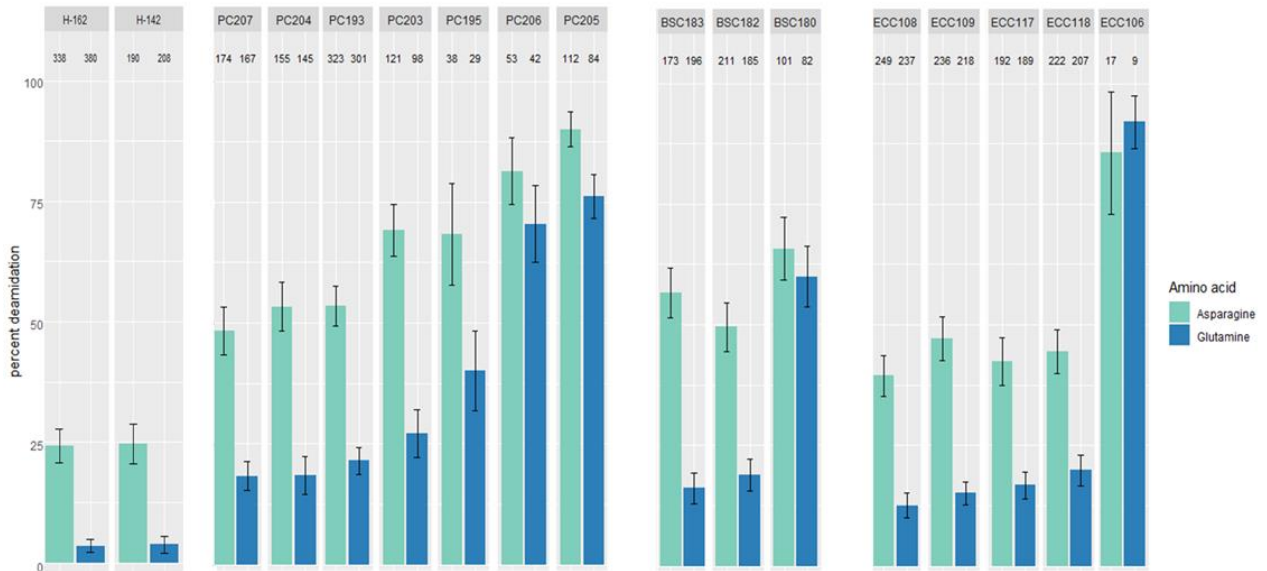


Figure S6: Percentage of deamidation (N,Q) of collagenous proteins in each single sample of Pompeii, Herculaneum and Scalandrone bay versus control (H-162-H-142) single samples[1]. Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

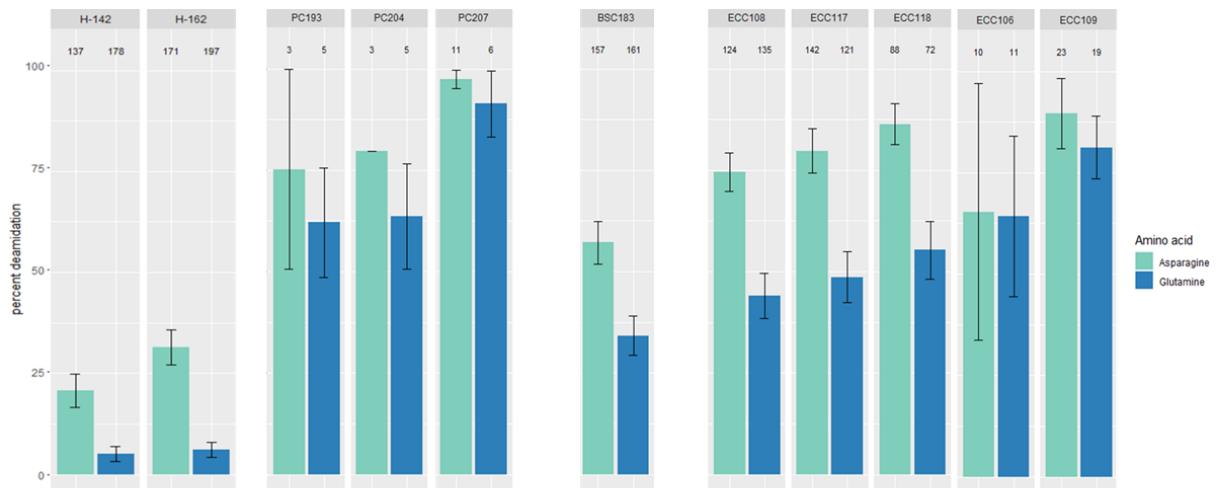


Figure S7: Percentage of deamidation (N,Q) of non-collagenous proteins in the single samples of Pompeii, Herculaneum and Scalandrone bay versus control (H-162-H-142) single samples[1]. Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

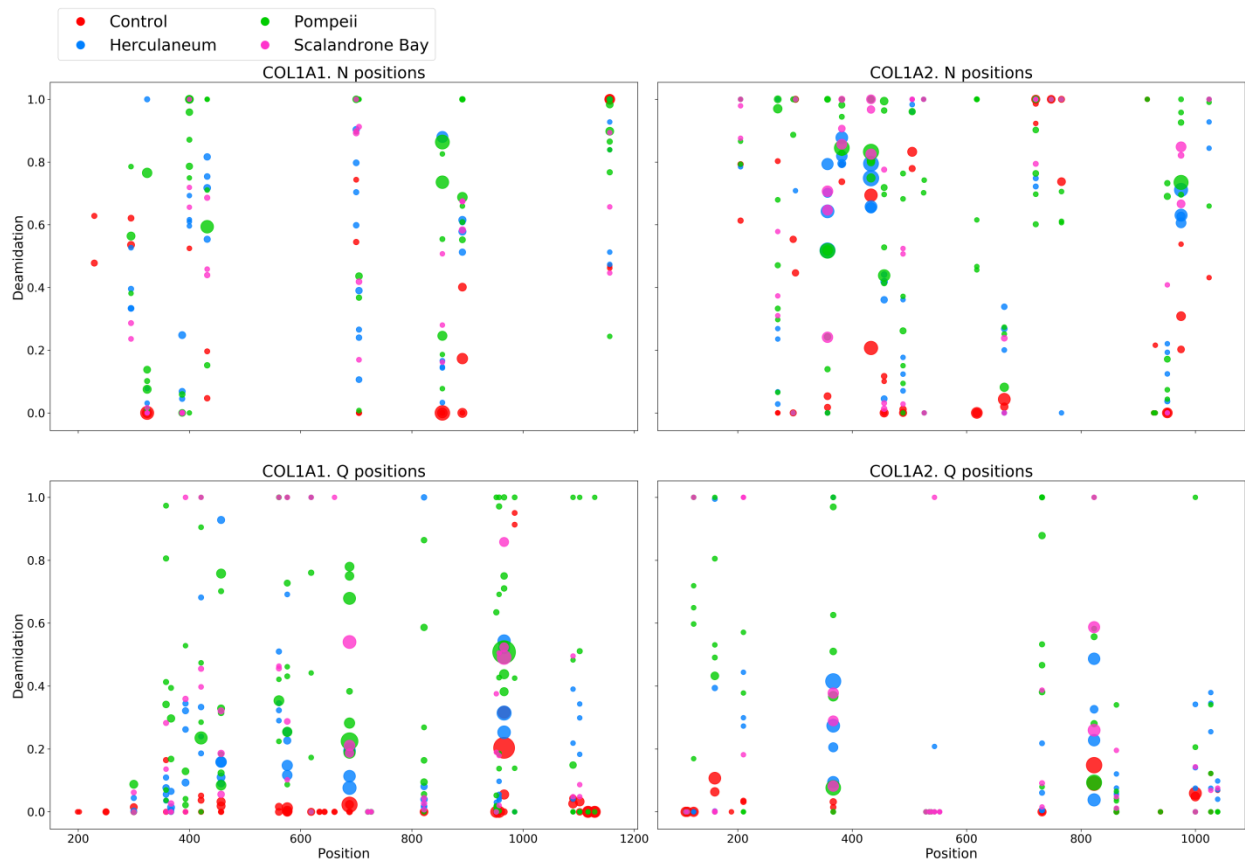


Figure S8: Visualization of deamidation level (N,Q) along the sequence of collagen alpha-1 (I) (upper panel) and alpha-2 (I) chains (lower panel), in the groups of Pompeii, Herculaneum, Scalandrone bay, Control (H-162, H-142) from [1].

Oxidation (M)

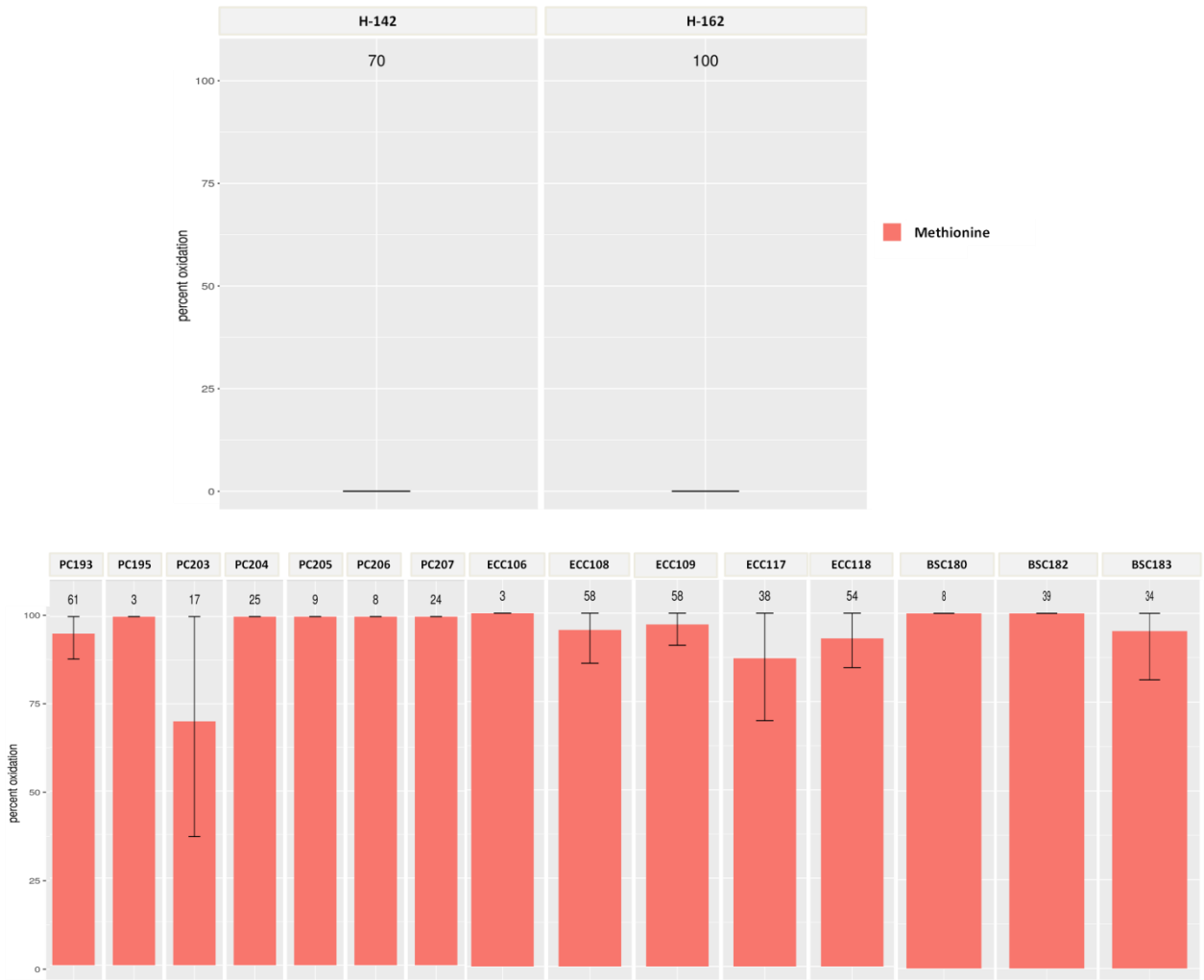


Figure S9: Percentage of oxidation (M) of collagenous proteins in the single samples of Pompeii, Herculaneum and Scaldrone bay versus control (H-162-H-142) single samples [1]. Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

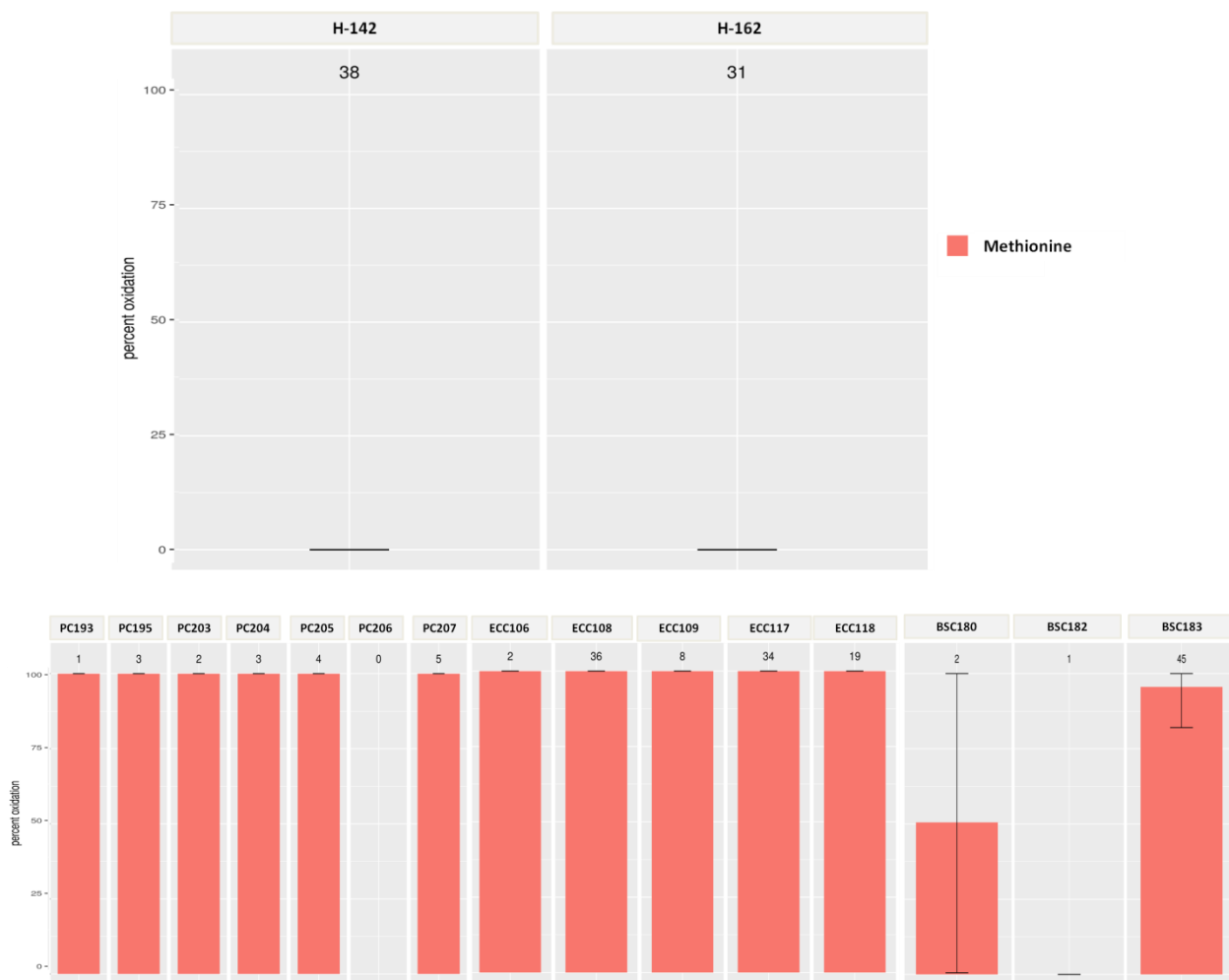


Figure S10: Percentage of oxidation (M) of non-collagenous proteins in the single samples of Pompeii, Herculaneum and Scalandrone bay versus control (H-162-H-142) single samples [1]. Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

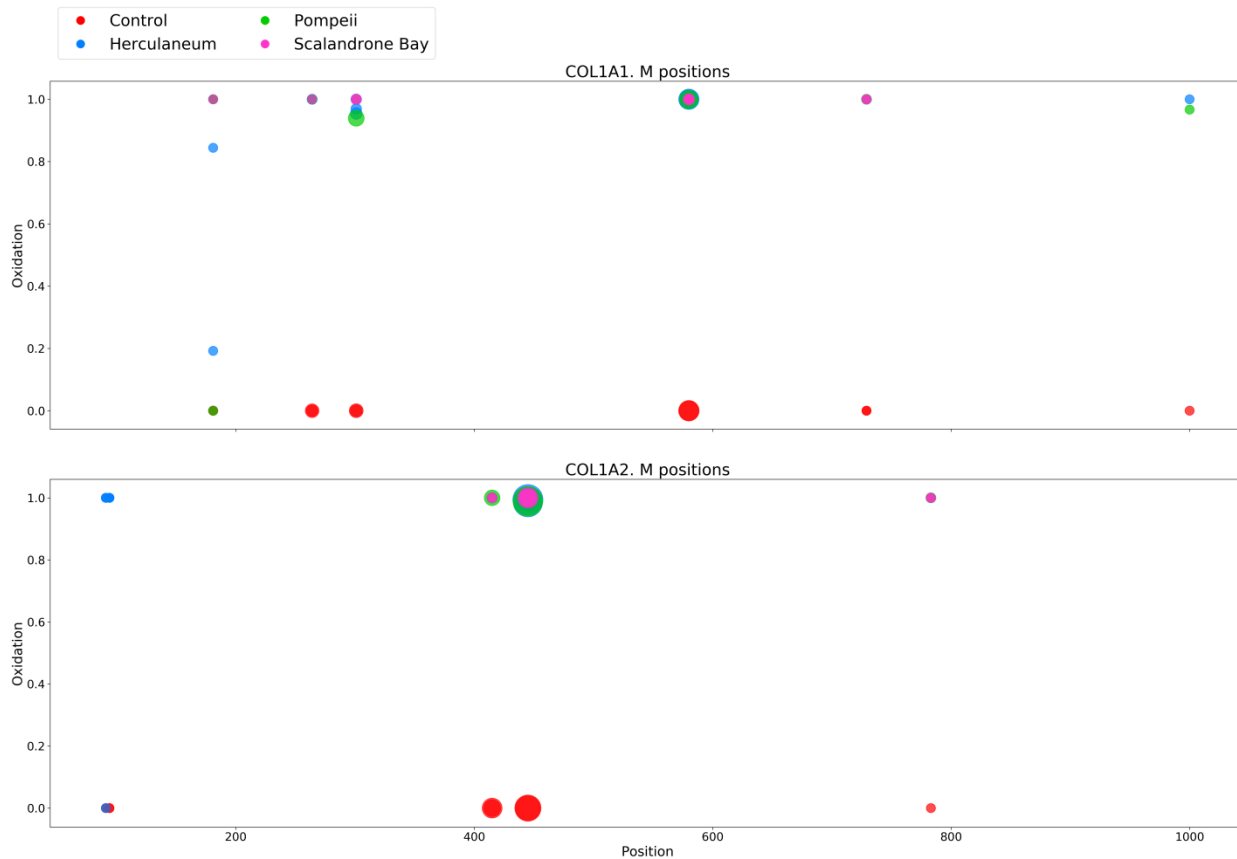


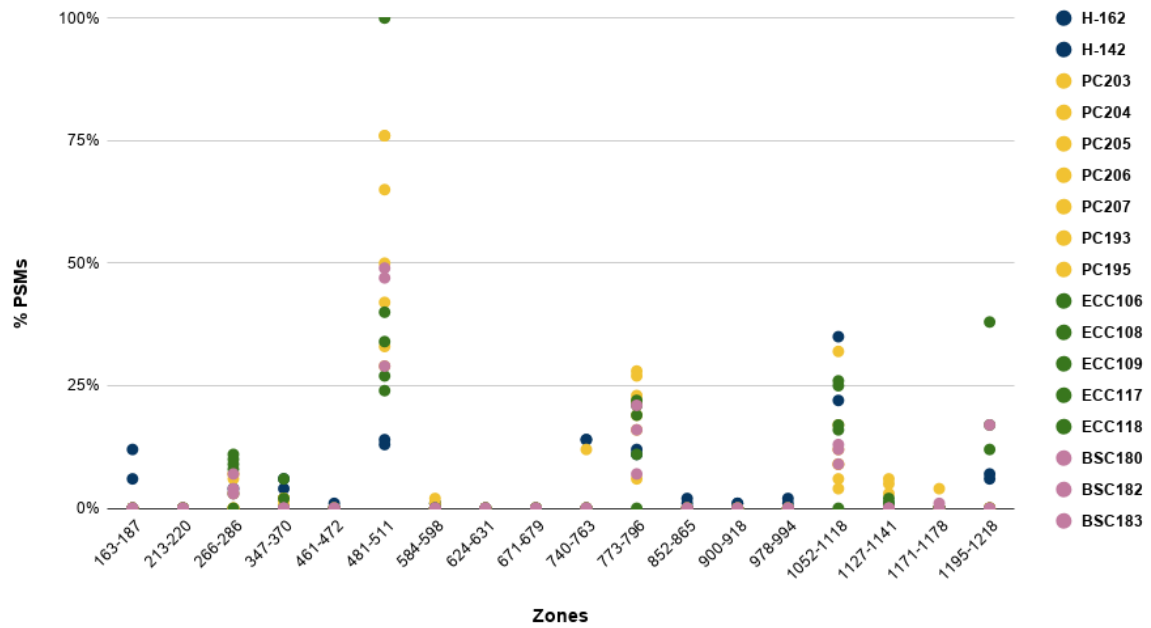
Figure S11: Visualization of oxidation level (M) along the sequence of collagen alpha-1 (I) (upper panel) and alpha-2 (I) chains (lower panel), in the groups of Pompeii, Herculaneum, Scalandrone bay, Control (H-162, H-142 from [1]).

Backbone cleavage

GSDGSVGPVGP
 GSDGSVGPVGPAG
 GSDGSVGPVGPAGPIG
 GSDGSVGPVGPAGPIGSA
 GSDGSVGPVGPAGPIGSAG
 GSDGSVGPVGPAGPIGSAGP
 GSDGSVGPVGPAGPIGSAGPP
 GSDGSVGPVGPAGPIGSAGPPG

Figure S12: Example of the “pacman” effect observed in the search for the semitryptic peptides in order to visualize spontaneous backbone cleavages.

Collagen I-alpha 1



Collagen I alpha-2

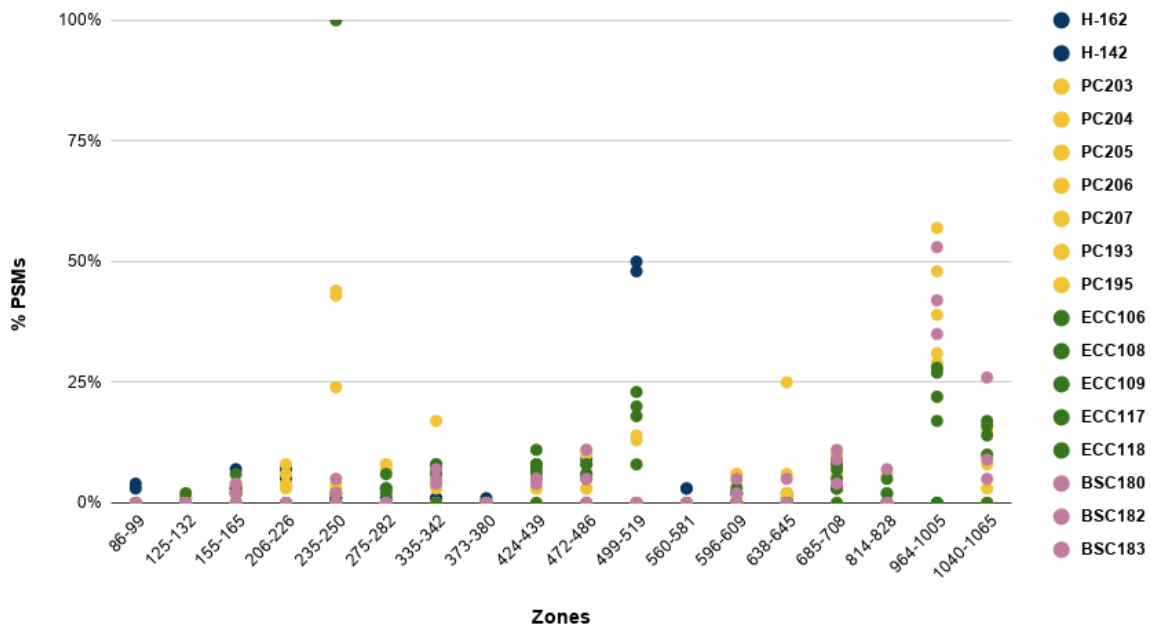


Figure S13: Semi-quantitative evaluation of the backbone cleavage along the sequence of COL1A1 and COL1A2 chains detected as semitryptic peptides.

Non-enzymatic chemical modifications

Table s4a-e: Frequency of chemical modification occurrence within the dataset of the sample groups of Pompeii, Herculaneum, Scalandrone Bay and Control (H162-142) from [1] .

A- Lysine Modification	ΔM	Pompeii	Herculaneum	Scalandrone Bay	Control
		K positions (mod/detected/theoretical number in the sequence of COL1A1 and COL1A2)			
Carboxyethyl	72.022	2/51/69	6/54/69	3/50/69	12/64/69
Amino adipic	14.964	6/51/69	4/54/69	1/50/69	1//64/69
Carboxymethyl	58.006	6/51/69	7/54/69	3/50/69	1//64/69
Formyl	27.996	6/51/69	3/54/69	1/50/69	6//64/69

B- Arginine Modification	ΔM	Pompeii	Herculaneum	Scalandrone Bay	Control
		R positions (mod/detected/theoretical number in the sequence of COL1A1 and COL1A2)			
MG-H1	54.011	9/87/108	8/88/108	3/91/108	4/103/108
G-H1	39.995	7/87/108	5/88/108	6/91/108	4//103/108
Ornithine	- 42.021	8/87/108	14/88/108	6/91/108	5//103/108

C- Proline Modification	ΔM	Pompeii	Herculaneum	Scalandrone Bay	Control
		P positions (mod/detected/theoretical number in the sequence of COL1A1 and COL1A2)			
Di-Oxidation	31.989	30/344/449	42/306/449	28/328/449	21/372/449
Tri-Oxidation	47.983	10/344/449	19/306/449	8/328/449	10/372/449
PyroGlu	13.980	10/344/449	23/306/449	8/328/449	10/372/449
Dhp	-2.001	17/344/449	14/306/449	3/328/449	3/372/449

D- Histidine Modification	ΔM	Pompeii	Herculaneum	Scalandrone Bay	Control
		H positions (mod/detected/theoretical number in the sequence of COL1A1 and COL1A2)			
His->Asp	-22.031	2/7/15	3/8/15	2/7/15	0/13/15
Di-Oxidation	31.989	2/7/15	3/8/15	0/7/15	1//13/15
Oxidation	15.995	2/7/15	2/8/15	1/7/15	2//13/15

E- Serine and Threonine Modification	ΔM	Pompeii	Herculaneum	Scalandrone Bay	Control
		S and T positions (mod/detected/theoretical number in the sequence of COL1A1 and COL1A2)			
S>G	-30.010	3/54/71	3/53/71	0/51/71	2/62/71
T->G	-44.026	3/29/37	3/30/37	1/32/37	2/32/37

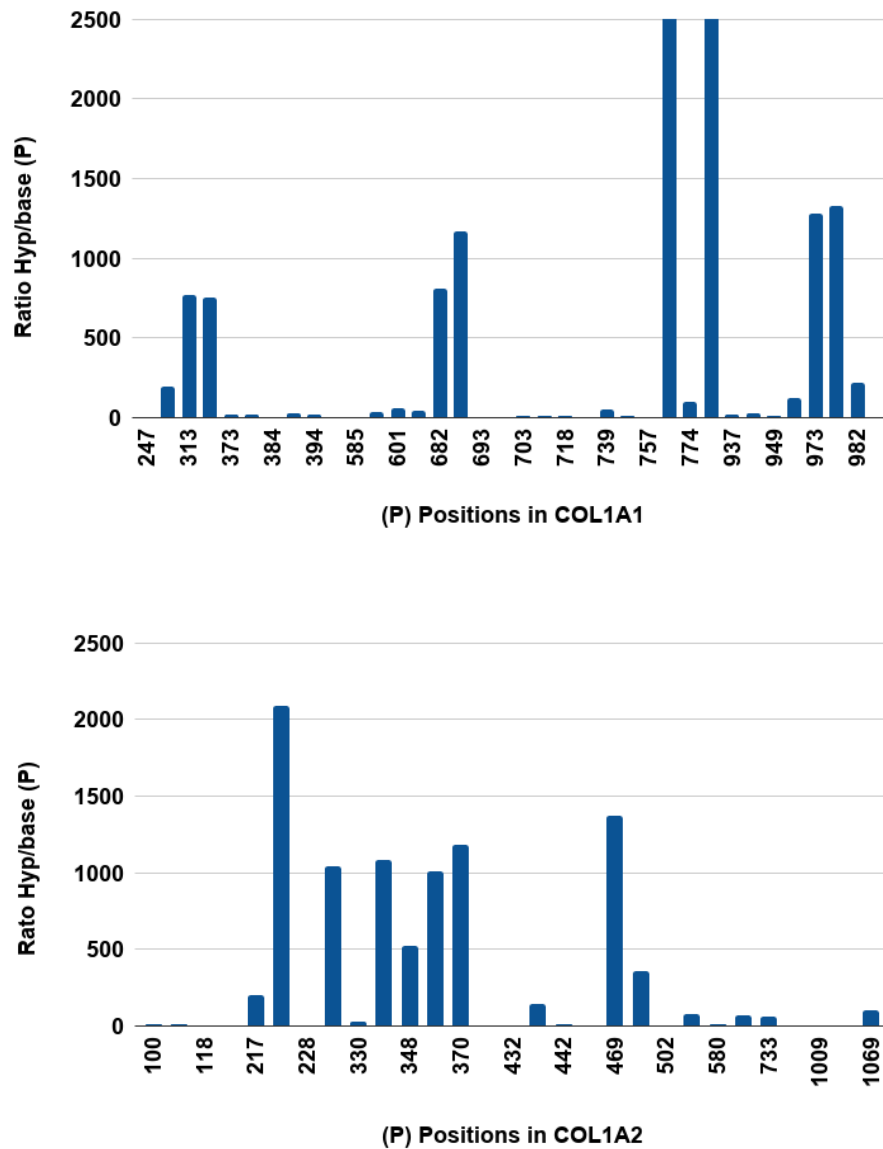


Figure S14: Example of primary structure hydroxylation occupancy on proline residues along the sequence of COL1A1 (left) and COL1A2 (right) of control samples [1]. The modification occupancy at a specific primary structure position arises from the average ratio modified/non modified values extracted from the HydroxyprolineSites.txt files of MQ. [8].

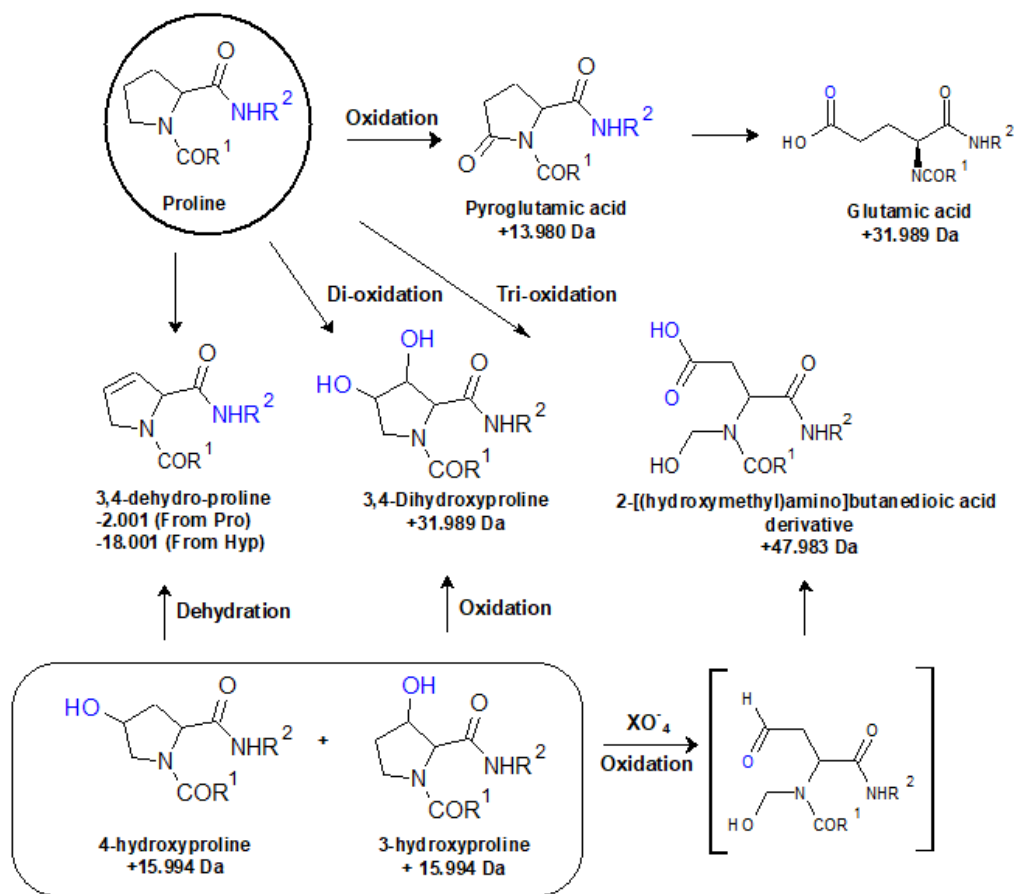


Figure S15: Possible proline oxidation products.

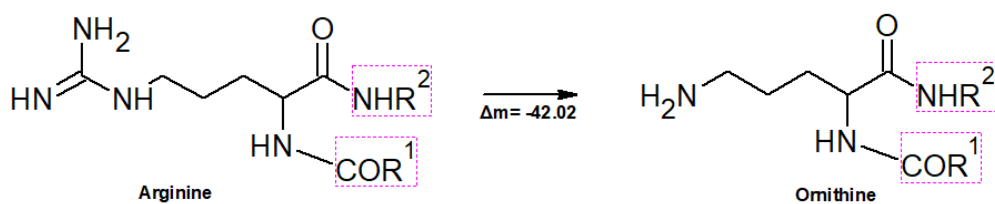


Figure S16: Conversion of Arginine to Ornithine.

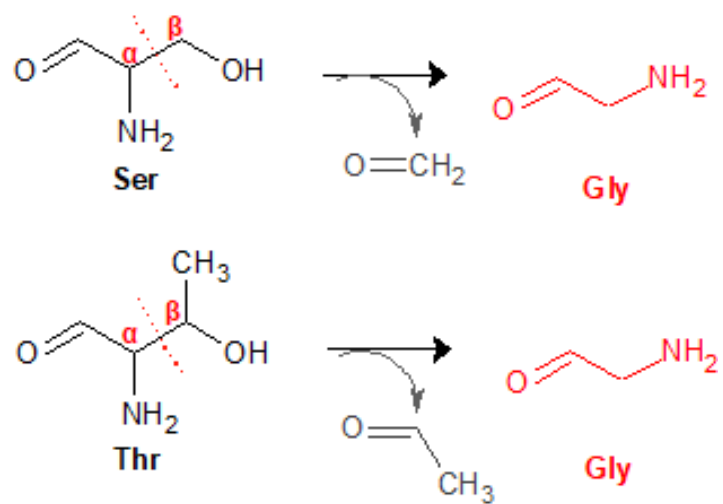


Figure S17: Oxidation reaction at the side chain of threonine (T) and serine (S), leading to C α -C β bond cleavage to Glycine (G).

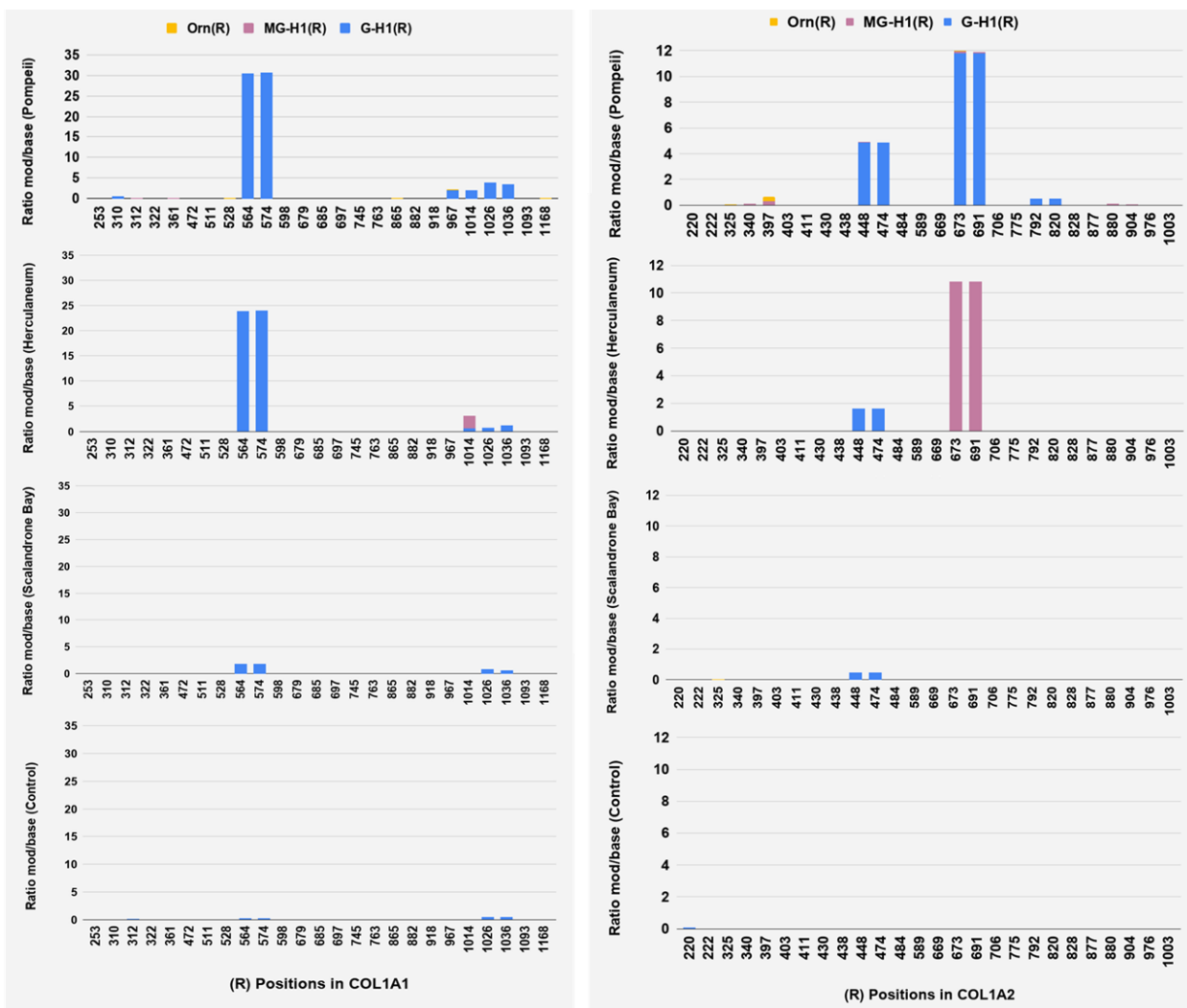


Figure S18: Primary structure modification occupancy on arginine residues along the sequence of COL1A1 (left) and COL1A2 (right). The modification occupancy at a specific primary structure position was semi-quantitatively calculated from the average modified/non modified values, extracted from the modificationSites.txt files of MQ[8]. Methylglyoxal (MGH-1) and glyoxal (GH-1) derived hydroimidazolones, and conversion of arginine to ornithine (Orn).

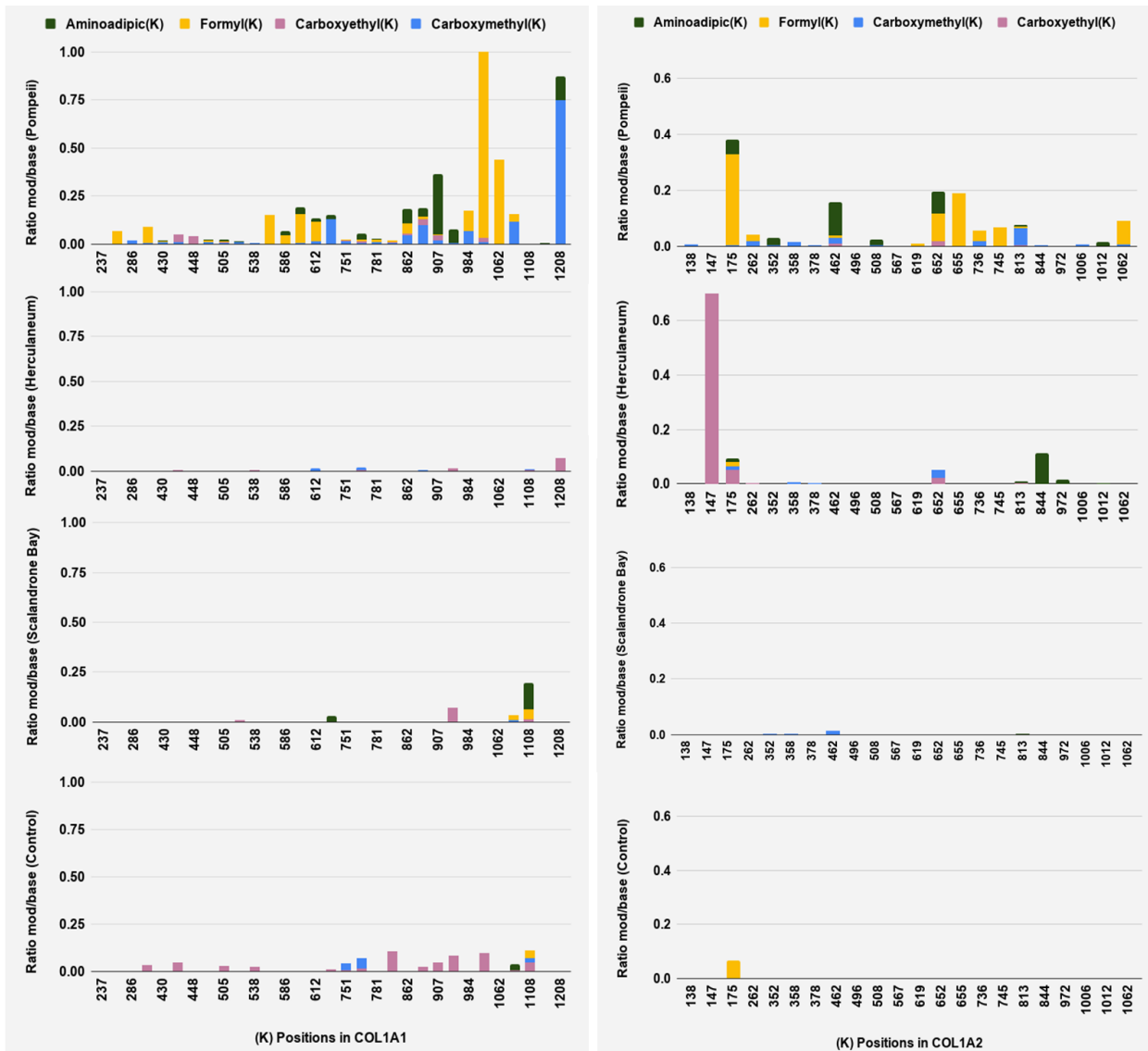


Figure S19: Primary structure modification occupancy on lysine residues along the sequence of COL1A1 (left) and COL1A2 (right). The modification occupancy at a specific primary structure position was semi-quantitatively calculated from the average modified/non modified values, extracted from the modificationSites.txt files of MQ[8].

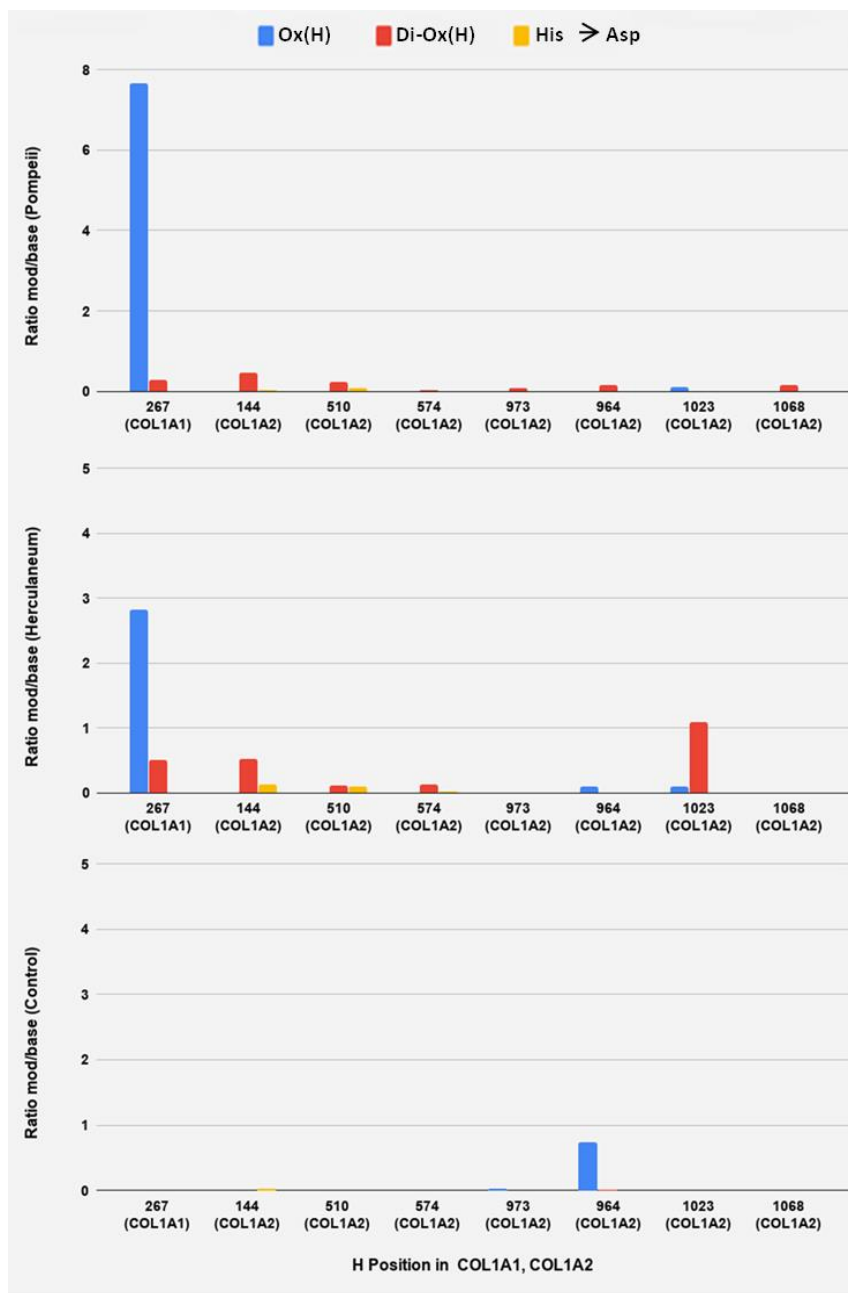


Figure S20: Primary structure modification occupancy on histidine residues along the sequence of COL1A1 (a single His was detected) and COL1A2. The modification occupancy at a specific primary structure position was semi-quantitatively calculated from the average modified/non modified values, extracted from the modificationSites.txt files of MQ[8].

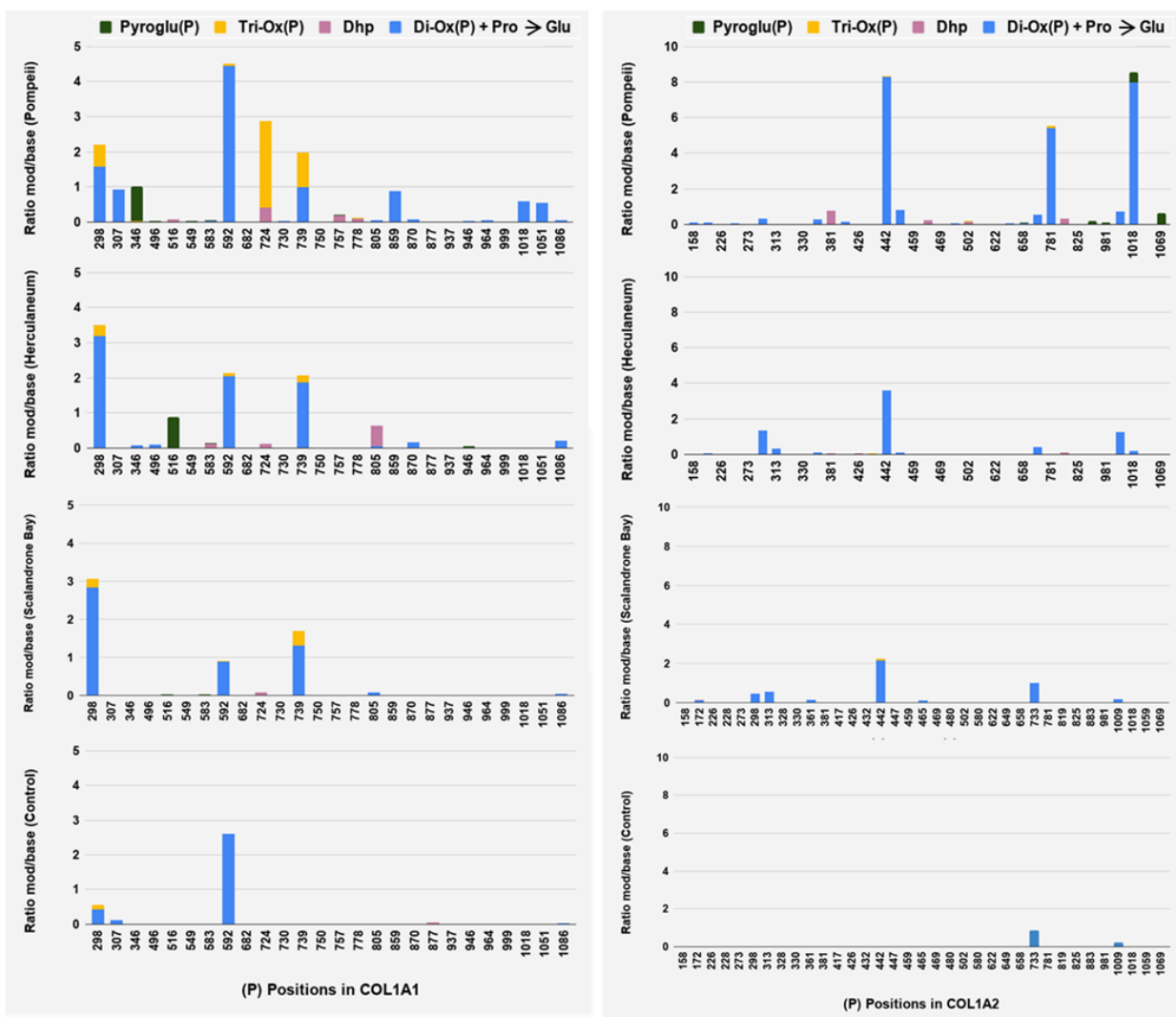


Figure S21: Primary structure modification occupancy on proline residues along the sequence of COL1A1 (left) and COL1A2 (right). The modification occupancy at a specific primary structure position was semi-quantitatively calculated from the average modified/non modified values, extracted from the modificationSites.txt files of MQ[8].

Selected MS/MS spectra to assess occurrence of modifications

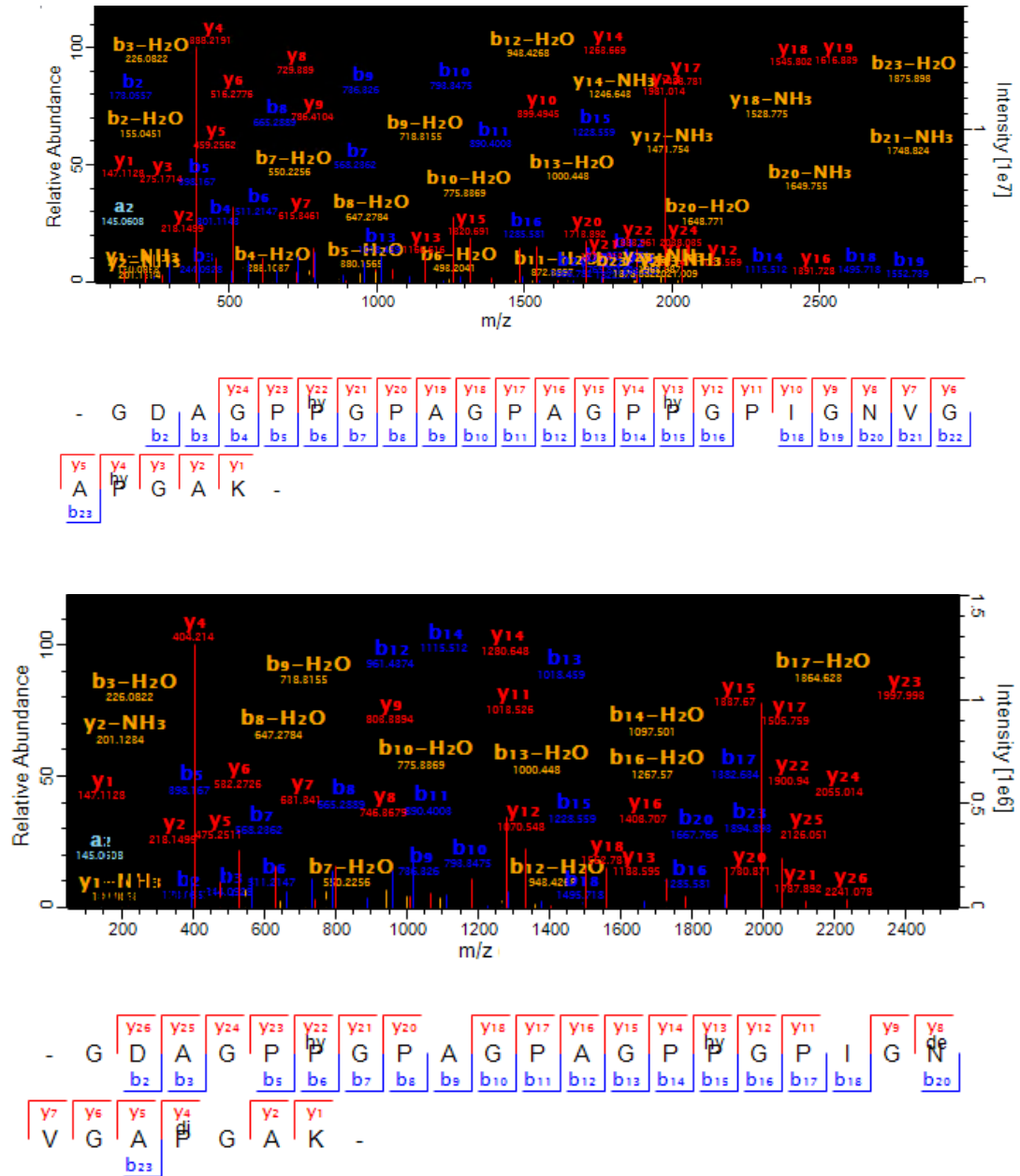
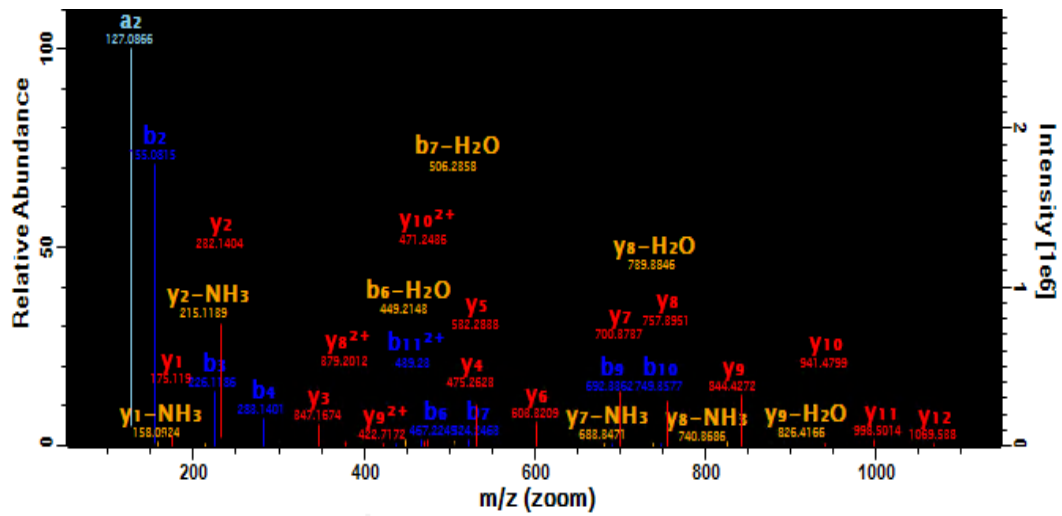
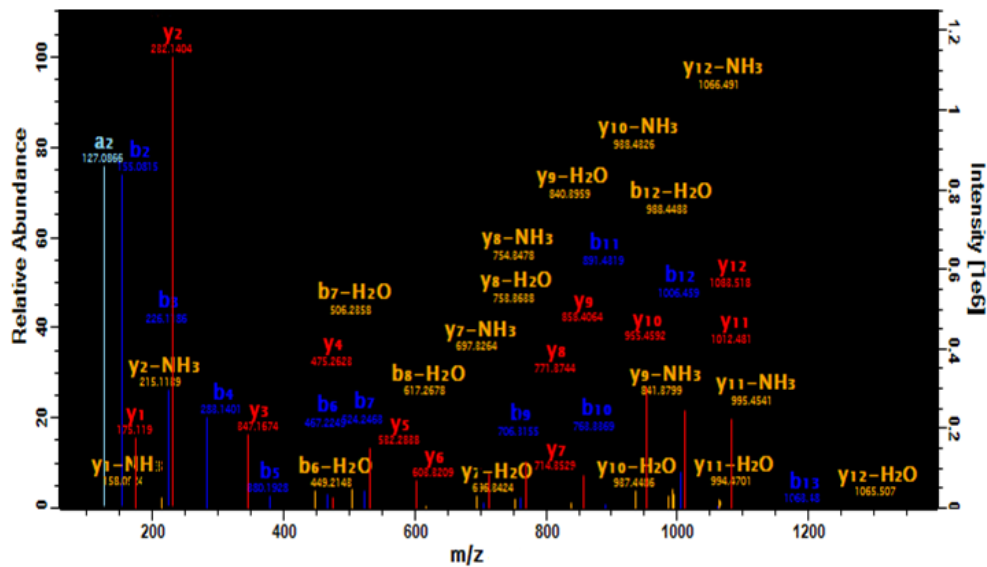


Figure S22a-b: MS/MS Spectra of COL1A1 GDAGPPGPAGPAGPPGPIGNVGAPGAK peptide . The underlined proline is detected as hydroxylated in the upper spectrum and di-oxidized (ΔM 31.989) in the lower spectrum.



- G P A G P S G P A G K D G R -
 b2 b3 b4 b6 b7 b9 b10 b11²



- G P A G P S G P A G K D G R -
 b2 b3 b4 b5 b6 b7 b9 b10 b11 b12 b13

Figure S23a-b: MS/MS Spectra of COL1A2 GPAGPSGPAGKDGR peptide. The underlined proline is detected as non modified in the upper spectrum and as oxidized to pyroglutamic acid (ΔM 13.979) in the lower spectrum.

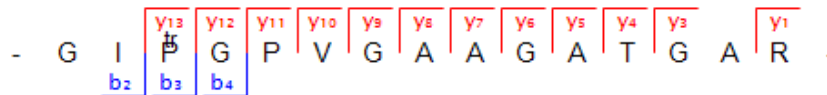
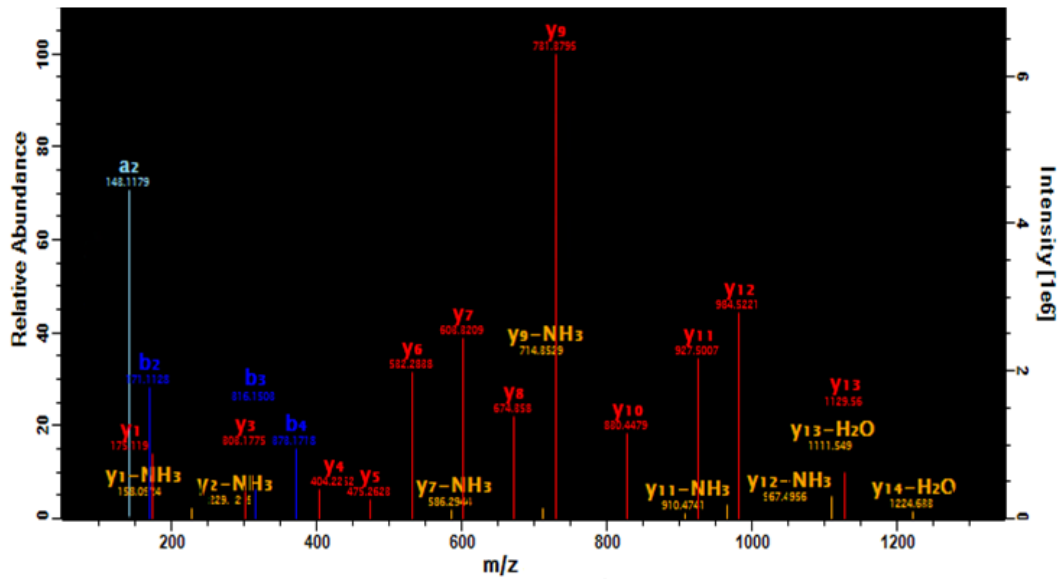
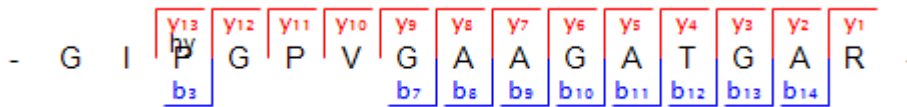
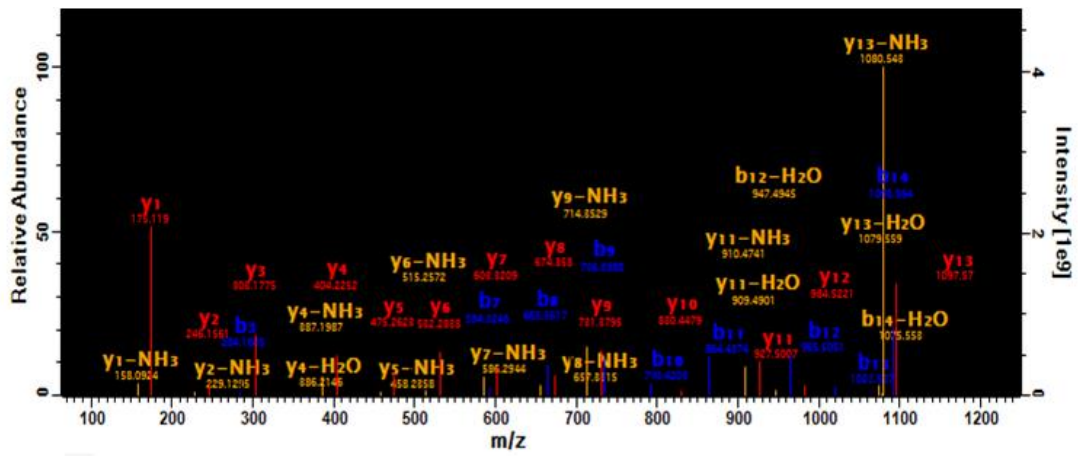


Figure S24a-b: MS/MS Spectra of COL1A2 GIPGPVGAAGATGAR peptide. The underlined proline is detected as hydroxylated in the upper spectra and as tri-oxidized (ΔM 47.985) in the lower spectrum.

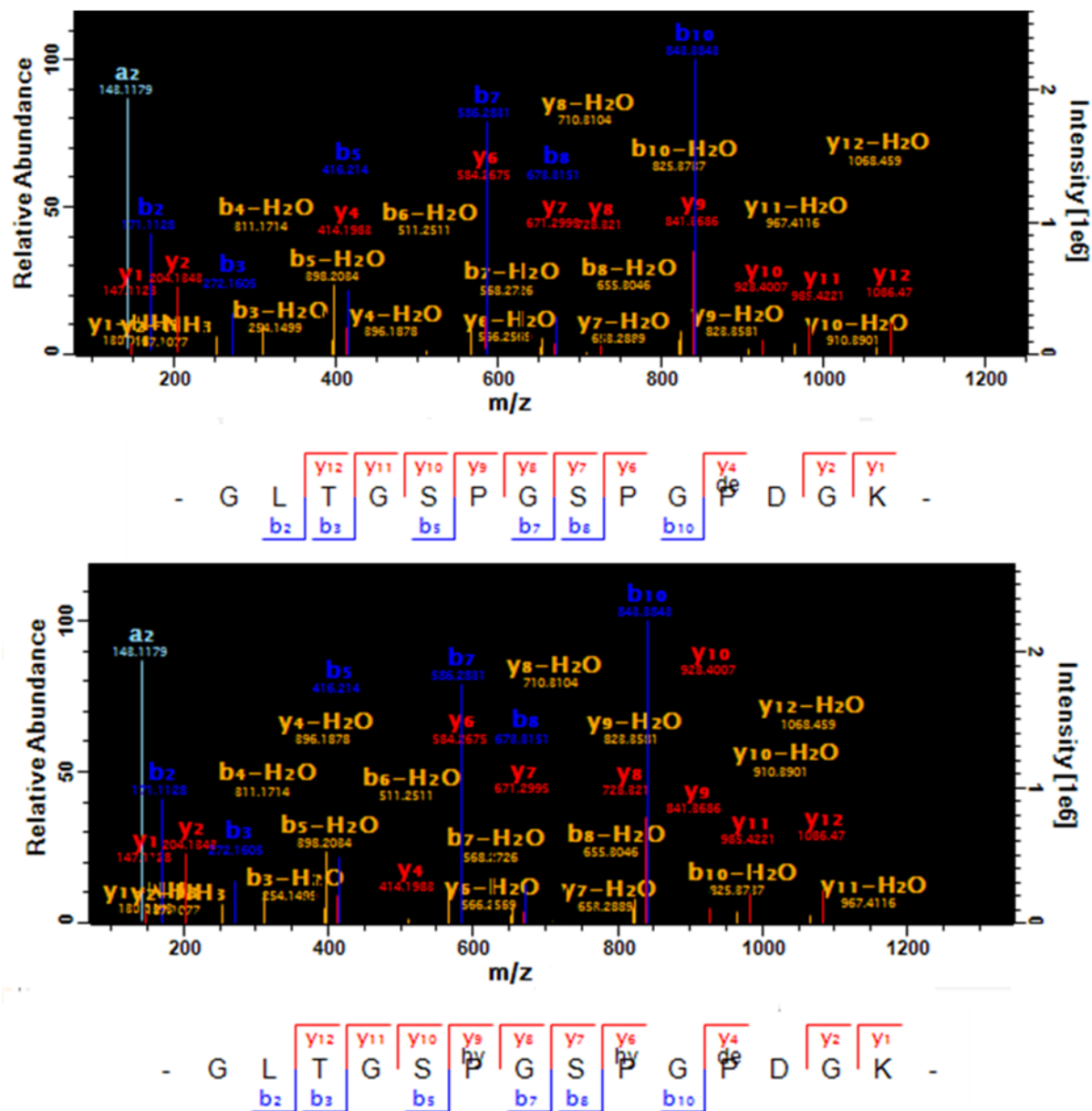


Figure S25a-b: MS/MS Spectra of COL1A1 GLTGSPGSPGDGK peptide. The underlined proline is detected in the upper spectra where hydroxylation (P) was set as variable modification with a mass shift of -2.016Da. When hydroxylation (P) was set as fixed modification (lower spectra), the underlined P was detected with a mass shift of -18.010 Da.

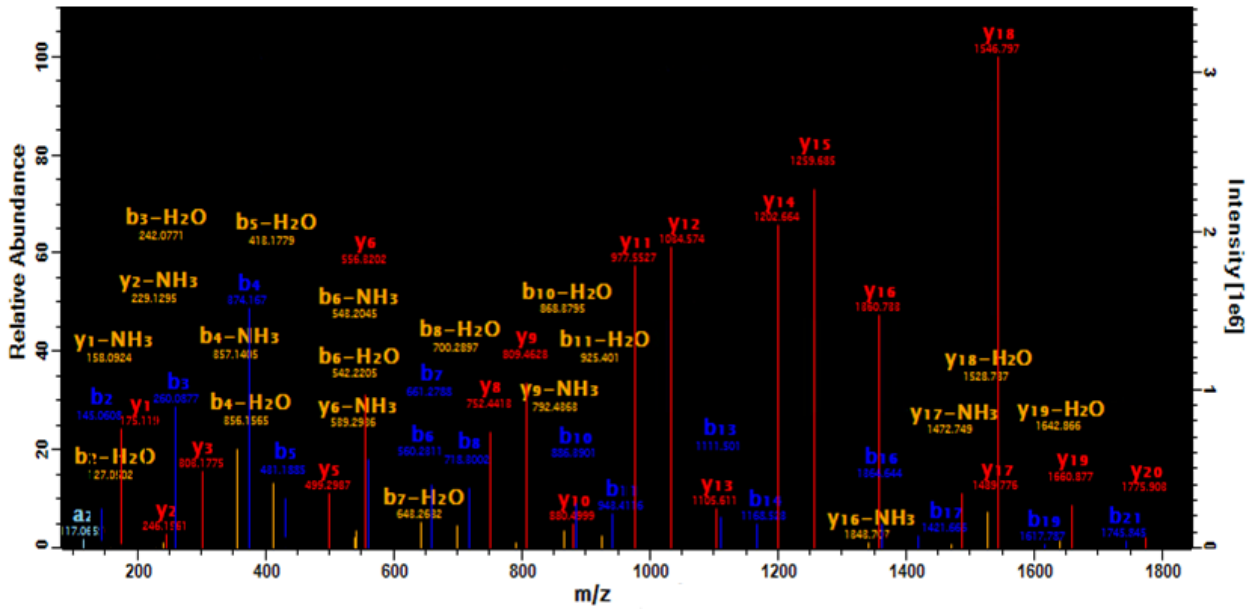


Figure S26: MS/MS Spectra of COL1A2 SGDRGETGPAGPAAGPVGPVGPVGR peptide. The underlined arginine is detected as ornithine ($\Delta M - 42.022$).

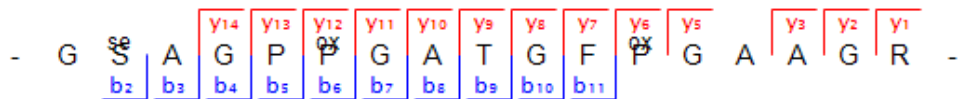
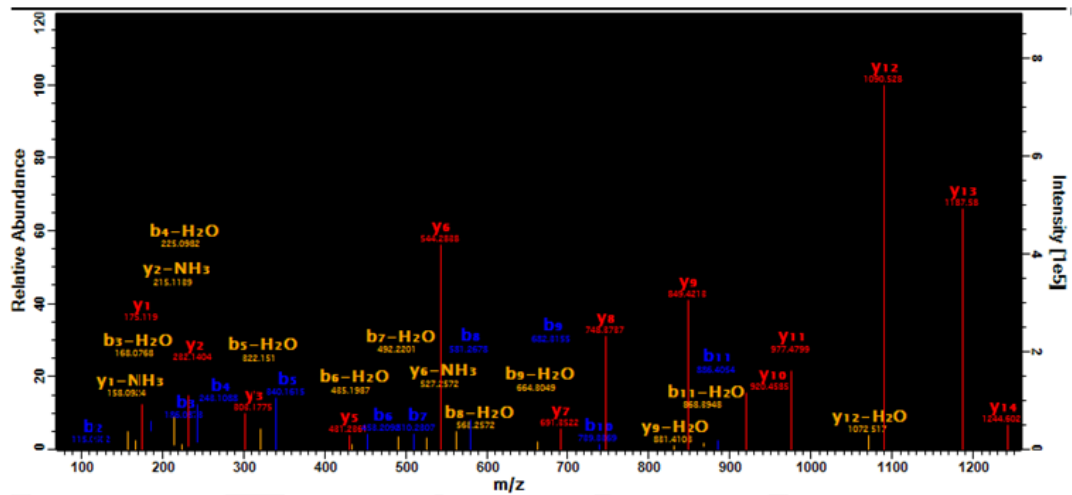
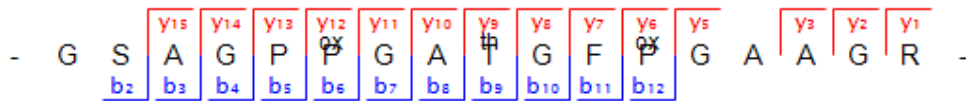
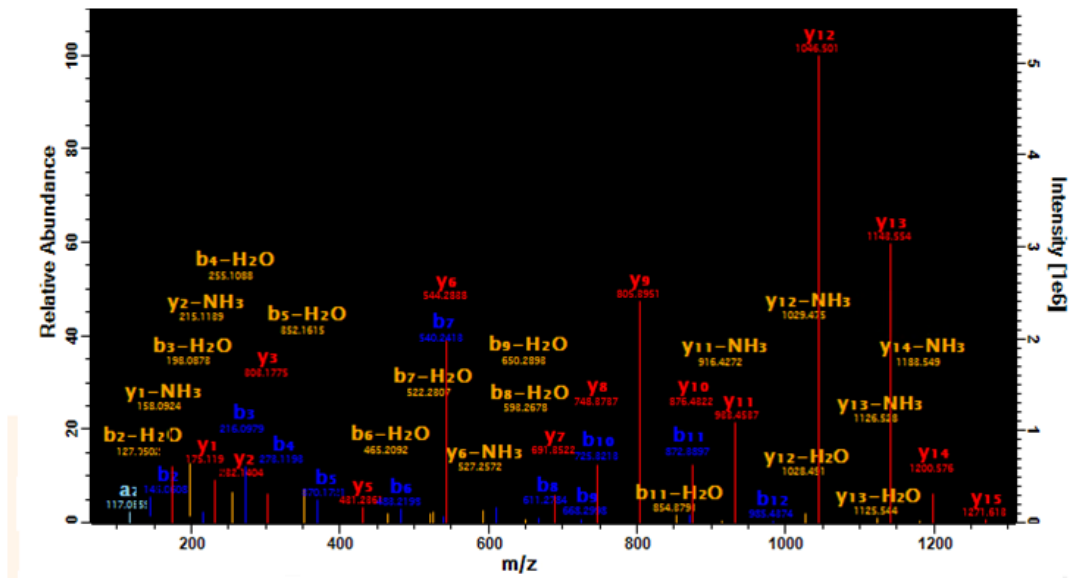


Figure S27a-b: MS/MS Spectra of COL1A1 GSAGPPGATGFPGAAGR peptide. The spectra demonstrate a Thr → Gly (ΔM -30.010) and a Ser → Gly (ΔM -44.026) substitution, respectively. The peptide with standard Ser and Thr was also identified.

References

- [1] R. Sawafuji, E. Cappellini, T. Nagaoka, A. K. Fotakis, R. R. Jersie-Christensen, J. V Olsen, K. Hirata, S. Ueda, *R. Soc. open Sci.* **2017**, *4*, 161004.
- [2] E. Cappellini, L. J. Jensen, D. Szklarczyk, A. Ginolhac, R. A. R. Da Fonseca, T. W. Stafford, S. R. Holen, M. J. Collins, L. Orlando, E. Willerslev, M. T. P. Gilbert, J. V. Olsen, *J. Proteome Res.* **2012**, *11*, 917–926.
- [3] J. Cox, M. Mann, *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- [4] S. Tyanova, T. Temu, J. Cox, *Nat. Protoc.* **2016**, *11*, 2301–2319.
- [5] M. M. Savitski, M. L. Nielsen, R. A. Zubarev, *Mol. Cell. Proteomics* **2006**, *5*, 935–948.
- [6] M. Mackie, P. R  ther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D. A. Pegg  , H. Howard, L. Harrison, L. J. Jensen, J. V. Olsen, E. Cappellini, *Angew. Chemie - Int. Ed.* **2018**, *57*, 7369–7374.
- [7] M. Ryk  r, B. Svensson, M. J. Davies, P. H  gglund, *J. Proteome Res.* **2017**, *16*, 3978–3988.
- [8] D. R. Martin, P. Dutta, S. Mahajan, S. Varma, S. M. Stevens, *Sci. Rep.* **2016**, *6*, 1–12.

3.3 Manuscript 2: Selection of peptide biomarkers in human bone collagen for the evaluation of deamidation (N, Q) and oxidation (M) by targeted proteomics.

Georgia Ntasi, Ismael Rodriguez Palomo, Addolorata De Chiaro, Gennaro Marino, Paolo Petrone and Leila Birolo

Abstract

Mass spectrometry (MS) is a powerful tool to analyze complex mixtures of proteins, discovering new modifications or characterizing and quantifying already annotated chemical modifications. The accurate quantitation of these modifications is vital in proteomic studies, revealing important information regarding 'protein's function, activity, and stability. One of the most frequent modifications is the deamidation of Asn and Gln. The transformation of asparagine to aspartic and/or iso-aspartic acid as also the transformation of glutamine to glutamic acid provokes a shift of 0.984 Da, thus rendering its quantification quite tricky. In this work, we used a dataset of (56) human bones from Herculaneum, Pompeii, Baia Scalandrone, Oplonti, and S.Paolo Belsito. A shotgun proteomic approach was applied to the samples, and collagen type I chains have always been identified with good scores and confidence. Screening of all the mass spectrometric data was performed in search of the most frequently detected and/or deamidated peptides, and the results were analyzed on a statistical basis to select the best peptides to be later used in the quantification of deamidation. The selected peptides were further validated by the analysis of deamidation (Asn, Gln) and oxidation (Met) in archaeological bones of published datasets, collected with different analytical procedures and by different mass spectrometry instrumentation, illustrating that this set of peptides can be used in any archaeological studies on human bone collagen.

Introduction

Deamidation is a spontaneous chemical modification that plays an essential role in protein degradation *in vivo*, so that has been proposed to represent a molecular clock [1–3]. It has been recently adopted as one of the signatures of authentic age in archaeological and paleontological bones as well as other archaeological remains or artistic objects [4–7]. It occurs on asparagine (Asn) and glutamine (Gln) amino acids resulting in aspartic/isoaspartic acid (Asp) and α -glutamic/ γ -glutamic acid (Glu), respectively. Bones are one of the most abundant mineralized tissues in the fossil record. Due to their structural, mechanical, and chemical properties, they resist over the years [1,2]. Collagen is the most abundant protein in bones (and in the whole body), and therefore the study of bone aging at a molecular level is frequently focused on collagen degradation [4,5]. In recent years, a particular focus was given to deamidation in bone collagen, and, as a

result, it is also considered in forensic science to estimate the post-mortem interval of unknown skeletal remains [8]. The conversion of asparagine and glutamine to aspartic and glutamic acids provokes a small mass shift of just +0.984 Da that, although detectable with MS-applied techniques, consists of a quite tricky target for quantification [5,9]. Most of the scientific approaches to quantify deamidation are, therefore, qualitative or semi-quantitative ones. [7,8,10–13] Targeted analyses such as multiple-reaction-monitoring (MRM) are alternative mass spectrometric approaches that allow a more rigorous quantitation [14] and have also been applied to the cultural heritage field, although seldomly [15], but its potentialities in determining deamidation already proved successful although challenging [15] [16]. Targeted approaches are invariably linked to the selection of adequate peptides. Herein we searched for peptide-candidates that can be used for deamidation (Asn, Gln) and oxidation (Met) quantitation studies in bone collagen. Using an extraordinary dataset of human bones from the archaeological sites of Pompeii, Herculaneum, and Oplontis, related to the same Vesuvius eruption of 79 DC, and a small number of bones from the pre-historical volcanic site of San Paolo Belsito, we analyzed LC-MS/MS data of tryptic digest of the collagen extracted in search for deamidation and oxidation spots that are always detectable, to be used as potential biomarkers. Some bone samples from coeval skeletal remains from the Campania archaeological site of Baia Scalandrone (II sec. AD, Roman imperial age, Puteoli, Naples, Italy) were also considered in an attempt to explore diagenetic factors that can affect collagen deamidation.

Materials and methods

Samples description

Fifty-six bone samples were collected from individuals unearthed in Pompeii (11), Herculaneum (22), Baia Scalandrone (12), Oplonti (8), and San Paolo Belsito (3) have been analyzed. Table 1 reports each specimen and related information.

Table 1: Specimens analysed.

Name	Type	Age	Site	Death
PC188	Elbow	79 AD	Pompeii	Vesuvius Eruption
PC189	Patella	79 AD	Pompeii	Vesuvius Eruption
PC190	Patella	79 AD	Pompeii	Vesuvius Eruption
PC193	Ribs	79 AD	Pompeii	Vesuvius Eruption
PC195	Femoral	79 AD	Pompeii	Vesuvius Eruption
PC196	Long bone	79 AD	Pompeii	Vesuvius Eruption
PC200	Clavicle	79 AD	Pompeii	Vesuvius Eruption
PC204	Femoral	79 AD	Pompeii	Vesuvius Eruption
PC205	Skull	79 AD	Pompeii	Vesuvius Eruption
PC208	Skull	79 AD	Pompeii	Vesuvius Eruption
PC209	Ulna	79 AD	Pompeii	Vesuvius Eruption
ECC104	Ribs	80 AD	Herculaneum	Vesuvius Eruption
EFEC104	Ilium	81 AD	Herculaneum	Vesuvius Eruption
ECC105	Ribs	82 AD	Herculaneum	Vesuvius Eruption
ECC184	Ribs	83 AD	Herculaneum	Vesuvius Eruption
EFPC106	Foot phalanx	79 AD	Herculaneum	Vesuvius Eruption
EFMC107	Hand plalanx	79 AD	Herculaneum	Vesuvius Eruption
ECC107	Ribs	79 AD	Herculaneum	Vesuvius Eruption
EFMC108	Hand plalanx	79 AD	Herculaneum	Vesuvius Eruption
ECC108	Ribs	79 AD	Herculaneum	Vesuvius Eruption
EFMC109	Hand plalanx	79 AD	Herculaneum	Vesuvius Eruption
ECC109	Ribs	79 AD	Herculaneum	Vesuvius Eruption
EFMC110	Hand plalanx	79 AD	Herculaneum	Vesuvius Eruption
EFMC111	Hand plalanx	79 AD	Herculaneum	Vesuvius Eruption
ECC111	Ribs	79 AD	Herculaneum	Vesuvius Eruption
EFMC112	Hand plalanx	79 AD	Herculaneum	Vesuvius Eruption
ECC112	Ribs	79 AD	Herculaneum	Vesuvius Eruption
EFPC113	Foot phalanx	79 AD	Herculaneum	Vesuvius Eruption
ECC113	Ribs	79 AD	Herculaneum	Vesuvius Eruption
ECC114	Ribs	79 AD	Herculaneum	Vesuvius Eruption
EFPC114	Foot phalanx	79 AD	Herculaneum	Vesuvius Eruption
EFMC115	Hand plalanx	79 AD	Herculaneum	Vesuvius Eruption
ECC115	Ribs	79 AD	Herculaneum	Vesuvius Eruption
BSCC180	Ribs	II sec AD	Scalandrone Bay	Normal death
BSFMC180	Hand phalanx	II sec AD	Scalandrone Bay	Normal death
BSC182	Ribs	II sec AD	Scalandrone Bay	Normal death
BSRC182	Radio	II sec AD	Scalandrone Bay	Normal death
BSC183	Ribs	II sec AD	Scalandrone Bay	Normal death

BSFMC183	Hand phalanx	II sec AD	Scalandrone Bay	Normal death
BSC184	Ribs	II sec AD	Scalandrone Bay	Normal death
BSFMC184	Hand phalanx	II sec AD	Scalandrone Bay	Normal death
BSC186	Ribs	II sec AD	Scalandrone Bay	Normal death
BSFMC186	Hand phalanx	II sec AD	Scalandrone Bay	Normal death
BSFMC187	Hand phalanx	II sec AD	Scalandrone Bay	Normal death
BSCC187	Ribs	II sec AD	Scalandrone Bay	Normal death
OPC213	Ribs	79 AD	Oplontis	Vesuvius Eruption
OPC214	Long bone	79 AD	Oplontis	Vesuvius Eruption
OPC215	Ribs	79 AD	Oplontis	Vesuvius Eruption
OPC216	Long bones	79 AD	Oplontis	Vesuvius Eruption
OPC217	Ribs	79 AD	Oplontis	Vesuvius Eruption
OPC218	Ilium	79 AD	Oplontis	Vesuvius Eruption
OPC226	Ribs	79 AD	Oplontis	Vesuvius Eruption
OPC228	Femoral	79 AD	Oplontis	Vesuvius Eruption
SPBC212OP	Long Bone	20 sec BC	San Paolo Bel Sito	Avellino pre-historical eruption
SPBC212OL	Flat bone of the shoulder	20 sec BC	San Paolo Bel Sito	Avellino pre-historical eruption
SPBC211	Tibia	20 sec BC	San Paolo Bel Sito	Avellino pre-historical eruption

Experimental procedures

Protein extraction and digestion

100 μ L of a solution of 0.5 M EDTA was added to the bone fragments (ca 300–800 μ g) for 10 days at RT, refreshing the solution every 2 days. After centrifugation for 2 min at 10,000 rpm in a benchtop microfuge, the supernatants were collected, and 10 μ L of a solution of 6 M Urea was added to the pellet and incubated for 10 min at RT, followed by sonication for 20 min. Urea was then 6-fold diluted with water. Finally, enzymatic digestion was carried out as in the minimally invasive proteomic analytical procedure described by Leo et al. [17]. Briefly, trypsin was added to a final concentration of 10 ng/ μ L to micro-samples (ca 300–800 μ g) as directly suspended in 50 μ L of Ambic 10 mM. After incubation at 37 °C for 16 h, the supernatants were recovered by centrifugation at 10,000 rpm, and the peptide mixture was filtered on 0.22 μ m PVDF membrane (Millipore), concentrated and purified using a reverse-phase C18 Zip Tip pipette tip as previously described [18].

LC-MS/MS and data analysis.

Peptides were eluted with 20 μ L of a solution made of 50% Acetonitrile, 50% Formic acid 0.1% in Milli-Q water, and analysed by LC-MS/MS. LC-MS/MS analyses were carried out on a 6520 Accurate-Mass Q-ToF LC/MS System (Agilent Technologies, Palo Alto, CA, USA) equipped with a 1200 HPLC System and a chip cube (Agilent Technologies). After loading, the peptide mixture was first concentrated and washed on a 40 nL enrichment column (Agilent Technologies chip), with 0.1% formic acid in 2% acetonitrile as eluent. The sample was then fractionated on a C18 reverse-phase capillary column (Agilent Technologies chip) at a flow rate of 400 nL/min, with a linear gradient of eluent B (0.1% formic acid in 95% acetonitrile) in A (0.1% formic acid in 2% acetonitrile) from 3% to 80% in 50 min. Peptide analysis was performed using the data-dependent acquisition of one MS scan (mass range from 300 to 2000 m/z) followed by MS/MS scans of the three most abundant ions in each MS scan. MS/MS spectra were measured automatically when the MS signal surpassed the threshold of 50,000 counts. Double and triple charged ions were preferably isolated and fragmented.

Data handling

The acquired MS/MS spectra were transformed in Mascot Generic files (.mgf) format and used to query the SwissProt database 2015_04 (548,208 sequences; 195,282,524 residues), with Homo sapiens as taxonomy restriction. A licensed version of Mascot software (www.matrixscience.com) version 2.4.0. was used with trypsin as enzyme; 3, as allowed number of missed cleavage; 10 ppm MS tolerance and 0.6 Da MS/MS tolerance; peptide charge from +2 to +3. No fixed chemical modification was inserted, but possible oxidation of methionines, deamidation at asparagines and glutamines, and hydroxylation on lysine and proline were considered as variable modifications. Only proteins presenting two or more peptides were considered as positively identified. Individual ion score threshold provided by Mascot software to evaluate the quality of matches in MS/MS data. Spectra with Mascot score below 25 were rejected.

Evaluation of the most frequently detected N, Q, and M in collagen chains type I.

Manual inspection of MS/MS data was performed in the samples where collagen protein sequence coverages exceed 20 % (33/56 samples). AA positions of asparagine, glutamine, and methionine were characterized in the selected samples as **NF** Not detected (Not Found), **X**: Detected only as unmodified, **D**: Detected only as modified, and **XD**: Detected both as modified and unmodified. To be considered for statistical analysis, we manually inspected the software assignment of the N or Q position under consideration in the fragmentation spectra in order to exclude any false positive matching or localization. To simplify the statistical analysis then we set 0 the NF characterization; thus, the non detected and as 1 the

modified positions (X, D, and XD characterization). Results were analysed on a statistical basis in order to identify which are the most frequently detected as deamidated (N, Q) and oxidized (M) peptides.

Evaluation of the deamidation N, Q, and oxidation (M) levels in collagen type I chains.

The acquired MS/MS spectra were re-analyzed in standard MaxQuant searches in a UniProt database (759,512 sequences, 37,179,137 residues) with *Homo sapiens* as the taxonomic restriction (20199 sequences, 928,813 residues). Samples have been analyzed as group datasets (5) based on the archaeological site. Deamidation of asparagine and glutamine residues (Asn, Gln), oxidation of methionine residues (Met), and hydroxylation of proline and lysine residues (Pro, Lys) were set as variable modifications. Additionally, MaxQuant's "evidence.txt" file was used for a positional evaluation of the relative deamidation and oxidation, using an in-house code available at <https://github.com/ismaRP/MSMSdeamidation>. For each asparagine, glutamine, and methionine position, the values are calculated by dividing the sum of intensities of the peptides containing the modification by the total intensity of all the peptides containing that position. Only peptides with high probability modification site assignments (>30) were considered. In this way, it was possible to evaluate whether the most frequently detected as deamidated (N, Q) and oxidized (M) positions of the previous step were also significantly modified.

Selection of the human collagen type I sequences to be used as biomarkers for deamidation and methionine oxidation.

Screening of mass data was performed in the 33 filtered bone samples, and the identified peptides containing the most frequently detected deamidation and methionine oxidation sites were listed. A systematic analysis was carried out in search of standard peptides to be used as markers. First of all, we considered only fragmentation spectra of good quality in order to avoid false positives by setting a high ion score of 25. The most observed m/z and ms/ms masses were listed for the common peptides that also have the same combination of modifications. In some cases, with a peptide containing more than one N, Q sites, specific m/z could discriminate deamidation position, but, in several cases, the same m/z was observed for both of the deamidation sites.

Final list of the peptide candidates as biomarkers for deamidation (N, Q) and methionine oxidation (M) in collagen chains type I.

There are several other aspects that were taken into consideration in the selection of peptides-candidates. Table 2 shows all the criteria that we used to filter the most frequently further detected peptides to lead finally to the set of the best candidates that can be used as target peptides in quantitation analysis such as MRM analysis.

Table 2: Selection criteria for deamidation (N, Q) and methionine oxidation (M) target peptides along with the sequences of COL1A1 and COL1A2 in the screening of the filtered samples (Protein sequence coverage >20).

Criteria	Description
Uniqueness	Peptides must be unique both for the protein sequence (Collagen) and for the taxonomy (<i>Homo sapiens</i>)
Peptide length	7-20 AA
Reliability	Exclusion of false positives , consideration of good fragmentation spectra with score above 25.
Number of N, Q	Preferably only one Gln or Asn position in each peptide
Hydrophobicity	We calculated the hydrophobicity based on the SSR Calc code http://hs2.proteome.ca/SSRCalc/Slope and we considered peptides in the range 10-30
Reactive residues	We tried to avoid the simultaneous presence of residues such as H, W,D, C that are susceptible to modifications during sample preparation

Validation of the peptide candidates in other datasets.

The selected peptides with the listed ms and ms/ms masses were tested in other archaeological bones of four independent already published datasets, where collagen type I have been identified. The tests were carried out on samples of different age, unrelated to the Vesuvius eruption, and that have been processed with different collagen extraction protocols and analysed with different instrumentation. Specifically, peptides were tested in the samples: H-162, H-142 (dated between 1657-1683 AD) reported in [13], 44HE950-RADFORD, 7NCE98A-WOODVILLE (dated between 1790-1850 AD) from [19], FWLeakey_1554 (dated between 40-70 ka BP) from [20] and the sample AR-30_A (dated between 40-50 ka BP) from [8]. The raw files of the aforementioned samples were downloaded from each free publication's deposition platform and were analyzed in separate MaxQuant runs in a UniProt database (759,512 sequences, 37,179,137

residues) with *Homo sapiens* as the taxonomic restriction (20199 sequences, 928,813 residues). Deamidation of asparagine and glutamine residues (Asn, Gln), oxidation of methionine residues (Met), and hydroxylation of proline and lysine residues (Pro, Lys) were set as variable modifications, and carbamidomethylation of cysteine was set as a fixed modification. Protein identifications were supported by a false discovery rate (FDR) of 0.01 applied (same FDR for dependent peptides when applied), and proteins were considered as identified only with least 2 different non-overlapping peptides above 70 ion score threshold. Contaminant proteins were assessed using the contamination.fasta provided by MQ which includes common laboratory contaminants [21] (see MaxQuant Downloads -contaminants.fasta,) and can be found at http://www.coxdocs.org/doku.php?id=maxquant:start_downloads.htm, n.d. These protein hits were excluded from further analysis. Only peptides with score higher than 40 were considered and manual inspection of ms/ms spectra was performed as further validation. The selected N,Q and M sites were classified in each of the sample as: XD: when the selected peptides was detected both in the modified (deamidated/oxidized) and non modified form; D: when the selected peptides was detected only in the modified form, X: when the selected peptides was detected only as unmodified; and NF: when the peptide was not found at all. Finally, the selected peptides were used for a semi quantitative evaluation of the modification level using an in-house script available at <https://github.com/ismaRP/MSMSdeamidation>. For each asparagine, glutamine and methionine position, the values are calculated by dividing the sum of intensities of the signals corresponding to the modified peptide by the total intensity of all the peptides containing that position.

Results & Discussion

A shotgun proteomics approach by LC-MS/MS was applied to skeletal samples excavated from Pompeii (11), Herculaneum (22), Baia Scalandrone (12), Oplontis (8), and San Paolo Bel Sito (3) archaeological sites. Very stringent criteria for protein identification were used: only peptides with scores higher than 25 were considered and proteins were considered as identified only when two or more peptides have been detected. Type I collagen was identified in all samples, and the sequence coverage was used as the first qualitative criteria of whether the specific sample could be used in the following statistical evaluation of deamidation (Asn, Gln). Individuals with collagen sequence coverages below 20 were rejected (figure S1). Thirty-three samples were finally selected, and manual inspection of MS/MS data was performed on peptides containing asparagine, glutamine, and methionine. Figure 1 represents the frequency of asparagine and glutamine sites detection along the chains of collagen type I. Surprisingly, despite the extremely high variability of the samples (different parts of the body, different archaeological sites, a different type of death) there are some positions that are always detected and others that are much less, and this profile is similar for all the groups. This means that some collagen regions are easily accessible and/or detectable, and, conversely, all the positions that are less detected belong to the same regions. For instance, Asn at position 229 and Gln 250, 639, 643, and 661 are all in two poorly covered regions in COL1A1. Also, from a statistical analysis perspective, it results that COL1A1 is a better candidate since many Asn and Gln of COL1A2 are poorly detected. This might also be due to the stoichiometric double concentration of COL1A1 in respect to COL1A2 since type I collagen is composed of two identical $\alpha 1$ (I) chains and one genetically different $\alpha 2$ (I) chain [22]. Interestingly, we can observe that less frequently detected regions are always followed by some among the most frequently detected ones, suggesting that there might be a structural explanation, and this is well-wishing for collagen degradation studies. This pattern is the same in the samples of all analyzed datasets. The manual evaluation of methionine oxidation led to very similar results, meaning that some Met positions were not detected in any of the samples and others always detected, and, most interestingly, always detected as oxidized (Fig 2). Interestingly, Met oxidation results seem to nicely complement data on deamidation sites since undetected Met positions of COL1A1 belong to "poorly covered" collagen regions.

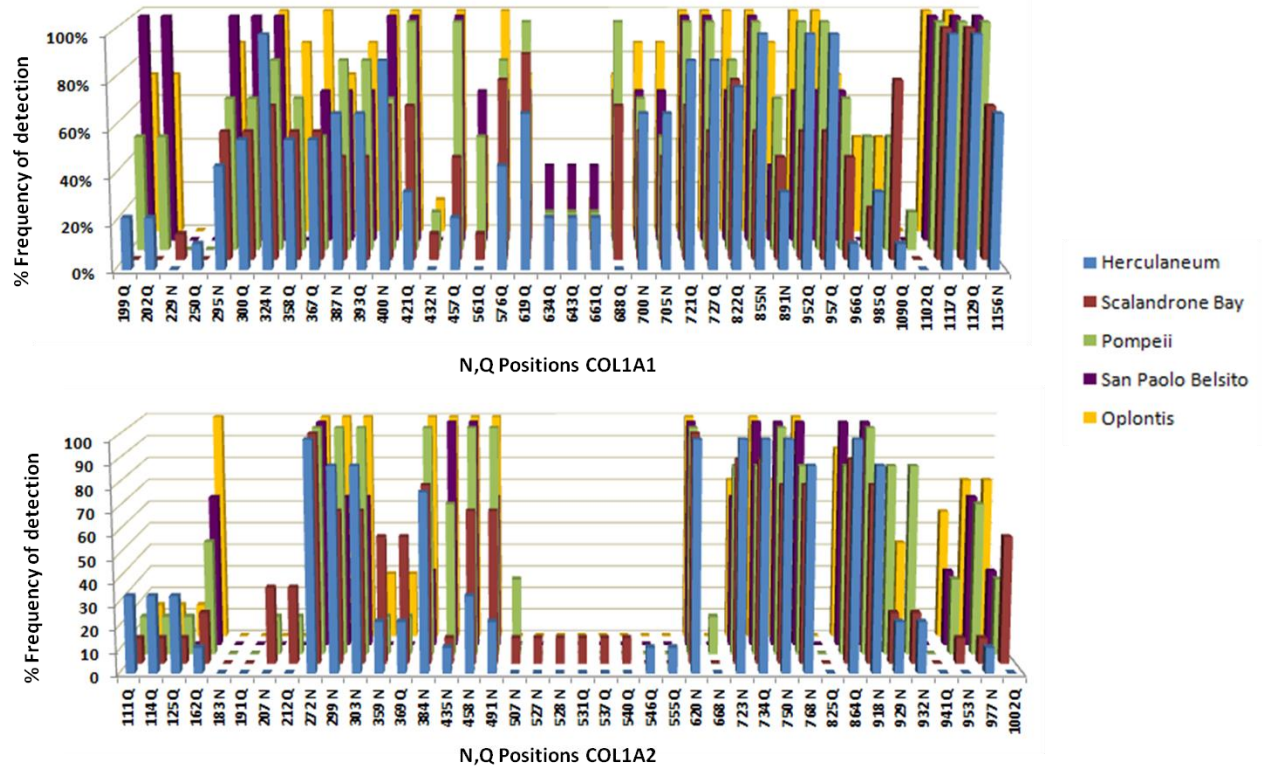


Figure 1: Bar chart of the average levels of Asn, Gln positions detection (regardless deamidation or any other possible modification within the peptides) in the chains of type I collagen (COL1A1, upper panel, COL1A2, lower panel) in the 33 samples of Pompeii, Herculaneum, Baia Scalandrone, Oplontis, and San Paolo Bel Sito.

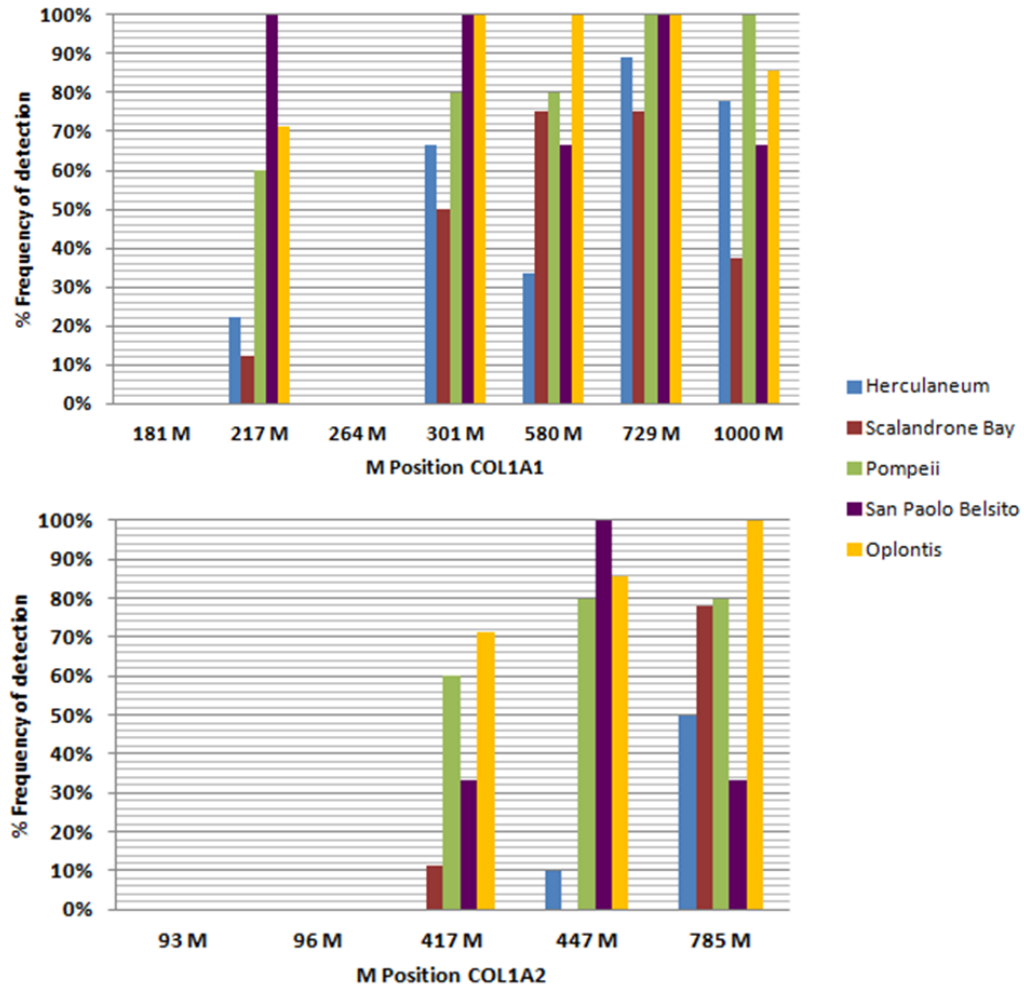


Figure 2: Bar chart of the average levels of Met positions detection (regardless of oxidation or any other possible modification within the peptides) in the chains of type I collagen (COL1A1; upper panel, COL1A2, lower panel) in the 33 samples of Pompeii, Herculaneum, Baia Scalandrone, Oplontis, and San Paolo Bel Sito.

Subsequently, manual inspection of MS/MS spectra was performed to evaluate Asn and Gln deamidation and Met oxidation, and the results are reported in tables S1-S12. Moreover, we evaluated whether the Asn, Gln, and Met positions detected as most modified are always identified with the same peptide, showing that only a few of them are identified with the same sequence and same combination of modifications (we have to keep in mind the high level of hydroxyproline peculiar of collagen). The most frequently observed m/z and ms/ms sequences of COL1A1 and COL1A2 are reported in tables S13-S14. In the case of Met oxidation, almost all the positions were equivalent, always detected with the same peptide, and always detected in the oxidized form (Table S15). Hydrophobicity was estimated for all peptides (<http://hs2.proteome.ca/SSRCalc/Slope>), as it significantly affects the peptide detectability in MRM

experiments, the most common and sensible targeted analysis. More analytically, long peptides 'display a linear relationship between signal intensity and peptide amount, often leading to inaccurate quantification results. Actually, the non-linearity results from a different absorption to the hydrophobic surface in the ESI droplet that leads to unstable responses [14, 16]. Intending to plan a quantitative study of deamidation, we take into account that ultimate peptides should be not too hydrophobic, should ideally contain only one Gln or Asn (preferably Gln since deamidation at Gln are slower [9] and defined concurrent post-translational modifications (in case of collagen, defined hydroxylation of prolines state). Based on the criteria above, tables 3 and 4 present the ultimate list of the best candidate peptides that can be used for deamidation (Asn, Gln) quantitation studies. Subsequently, a positional evaluation of deamidation (Asn, Gln) and oxidation (Met) was performed by using an in-house script (see materials and methods) to further validate if the positions detected as most deamidated and oxidated are significantly modified. The script performs a relative quantification based on peptide intensities among the modified and non-modified forms. As figures S2 and S3 show, some of the above frequency evaluation positions can be considered hot spots of deamidation (Asn, Gln) and oxidation (Met).

Table 3: List of the best peptide candidates that can be used as biomarkers of deamidation in human COL1A1 (P02452) and COL1A2 (P08123), according to a statistical evaluation of 33 bone samples of Pompeii, Herculaneum, BaiaScalandrone, Oplontis and San Paolo Bel Sito. Z: Charge; PL: Peptide Length; H: Hydrophobicity. The mass of the corresponding deamidated form can be calculated with the addition of 0.984 Da in the M value of the unmodified peptide.

Protein	N,Q Position	Peptide	Z	PL	H	Calculated M (Da)	Observed m/z
COL1A1 (P02452)	367 Q	R.GSEGPQGVR.G	2	10	8.09	885.416	443.7139
	400 N	K.GANGAPGIAGAPGFPGAR.G + 3 Hydroxylation(P)	3	19	23.93	1584.751	529.2607
	421 Q	R.GPSGPQGPGGPPGPK.G + Hydroxylation(P)	2	16	12.19	1301.622	651.8212
	457 Q	K.GEPGPVGVQPPGPAGEEGK.R + 2Hydroxylation(P)	3	21	21.04	2002.957	668.6648
	561Q	K.TGPPGPAGQDGRGPPGPPGAR.G + + 4Hydroxylation(P)	3	23	14.39	2055.958	686.3318

576 Q	R.GQAGVMGFPGPK.G + Hydroxylation(P)	2	13	22.11	11160.55	581.73
619 Q	K.DGEAGAQQPPGPAGPAGER.G + Hydroxylation(P)	2	20	15.14	1705.752	853.8904
688 Q	R.GVQGGPPGAGPR.G+ Hydroxylation(P)	2	13	11.52	1104.554	553.2881
891N	R.VGPPGSGNAGPPGPPGAGK.E + 3 Hydroxylation(P)	2	22	17.33	1811.866	906.9444
966 Q	R.GVVGLPGQR.G + Hydroxylation(P)	2	10	17.58	897.489	449.7549
1156 N	K.DGLNGLPGPIGPPGPR.G+ 3 Hydroxylation(P)	2	17	30.44	1560.776	781.3994

COL1A2 (P08123)	272 N	K.GEIGAVGNAGPAGPR.G	2	19	20.83	1546.771	774.3924
	384 N	R.GPNGEAGSAGPPGPPGLR.G + 2 Hydroxylation(P)	2	20	20.99	1618.756	810.3844
	458 N	R.GLPGSPGNIGPAGK.E +2 Hydroxylation(P)	2	15	17.68	1252.627	627.3252
	491 N	R.GEPGNIGFPGPK.G+ 2 Hydroxylation(P)	2	13	22.45	1200.563	601.2912
	918 N	R.GPPGAVGSPGVNGAPGEAGR.D + 3 Hydroxylation(P)	2	21	18.91	1750.81	876.4174

Table 4 List of the most frequently detected Met containing peptides in human COL1A1 (P02452) and COL1A2 (P08123), according to a statistical evaluation of 33 bone samples of Pompeii, Herculaneum, Baia Scalandrone, Oplontis, and San Paolo Bel Sito. The most frequently detected (>90%) masses are reported with and without deamidation. Z: Charge; PL: Peptide Length; H: Hydrophobicity. The mass of the corresponding oxidized form can be calculated with the addition of 15.998 Da in the M⁺ value of the unmodified peptide.

Collagen chain	M Position	Peptide	Z	PL	H	Calculated M (Da)	Observed m/z
COL1A1 (P02452)	580 M	R.GQAGVMGFPGPK.G+Deamidation(N,Q) + Hydroxylation(P)	2	13	22.11	1161.546	581.7713
	1000 M	R.GPPGPMGPPGLAGPPGESGR.G + 2 Hydroxylation(P)	2	21	24.91	1815.854	908.9252
COL1A2 (P08123)	417 M	R.AGVMGPPGSR.G + Hydroxylation(P)	2	11	13.21	943.451	472.7182
	447 M	R.GPNGDAGRPGEPGLMGPR.G + Deamidation(N,Q) + Hydroxylation(P)	3	19	18.87	1766.798	884.393
	785 M	R.GDGGPPGMTGFPGAAGR.G + 2 Hydroxylation(P)	2	18	21.7	1532.665	767.3355

Final validation of the selected peptides was performed by evaluating their occurrence in six already published archaeological bones that have been extracted with different protocols and analyzed by different instrumentation. In particular, four to six archaeological bones are much younger than the samples from the Campania region, dated to 18-19th century AD (reference numbers: H-162, H-142 from [13], and 44HE950-RADFORD, 7NCE98A-WOODVILLE from [19] while the other two dates back between 40.-70ka BC attributed to *Homo neanderthalensis* (reference numbers: AR-30A from [8] and FWLeakey_1554, from [20]). All samples were analyzed in separate MaxQuant runs, and the presence of the selected peptides was also manually checked. As table 5 illustrates in the older samples (AR-30 and FWLeakey_1554), all the peptides were detected both in the modified and non-modified form. Very interestingly, in the younger samples (18-19th century AD), some peptides were even not deamidated (Asn, Gln) or oxidized (Met), and in some cases, the Asn, Gln positions were detected with intense signal even in positions not covered in older samples (AR-30A from [8] and FWLeakey_1554, from [20]). In order to have a more quantitative evaluation of deamidation (Asn, Gln) and oxidation (M) levels of test bones, the selected peptides were used to calculate deamidation and oxidation at single sites along collagen alpha-1 (I) and alpha-2 (I) chains, using the in-house script described above. As figure S4 demonstrates, the level of deamidation varies among the samples with the younger bones (reference numbers: H-162, H-142 from [13], and 44HE950-RADFORD, 7NCE98A-WOODVILLE from [19] to be less deamidated in comparison to the older ones (reference numbers: AR-30A from [8] and FWLeakey_1554, from [20]). Therefore, the selected peptides can be used to distinguish collagen deamidation in samples of different ages and be considered in the evaluation of collagen degradation over the years. Moreover, figure S5 reports the samples' oxidation levels using the selected peptides. Also, in this case, the older samples (reference numbers: AR-30A from [8] and FWLeakey_1554, from [20]) are more oxidized in comparison to younger ones, confirming that also methionine oxidation is worth considering in degradation studies since it increases with age in body tissues, due to biological and/or diagenetic aging. Nevertheless, both deamidation and methionine oxidation are sensitive to several environmental factors, from humidity to the chemical characteristics of the burial soil, as well as to sample manipulation.

Table 5: Evaluation of the occurrence of the selected biomarkers in test cases of archaeological bones. X indicates that the peptide was detected as unmodified (deamidation (Asn, Gln) or oxidation (Met)) ;D indicates that the peptide was detected as deamidated or oxidize; X-D indicates that both modified (deamidated or oxidized, respectively) and unmodified forms were detected; NF. indicates that the peptide was not identified at all.

Protein	N, Q Position	Peptide	H-142	H-162	44HE950-RADFORD	7NCE98A-WOODVILLE	FWLeakey_1554	AR-30
COL1A1 (P02452)	367 Q	R.GSEGPQGV.R	X	NF	NF	X	XD	XD
	421 Q	R.GPSGPQGGP PGPK.G + Hydroxylation(P)	X	X	X	X	XD	XD
	457 Q	K.GEPGPVGVQGP PGPAGEEGK.R + 2Hydroxylation(P)	NF	XD	X	X	XD	XD
	561Q	K.TGPPGAGQDG RPGPPGGAR.G + 4Hydroxylation(P)	XD	XD	X	X	XD	D
	576 Q	R.GQAGVMGFPGP K.G + Hydroxylation(P)	XD	XD	XD	XD	XD	XD
	580 M	R.GQAGVMGFPGP K.G +Deamidation(N, Q) + Hydroxylation(P)	X	XD	D	D	XD	XD
	619 Q	K.DGEAGAQQPPG PAGPAGER.G +Deamidation(N, Q) + Hydroxylation(P)	NF	X	X	X	XD	XD
	688 Q	R.GVQGGPAGP R.G + Hydroxylation(P)	XD	XD	X	X	XD	XD
	891N	R.VGPPGSGNAG PPGPPGAGK.E + 3 Hydroxylation(P)	XD	XD	X	X	XD	XD
	966 Q	R.GVVGLPGQR.G + Hydroxylation(P)	D	XD	XD	XD	XD	XD
	1000 M	R.GPPGPMGPPGL AGPPGESGR.G +	X	NF	XD	XD	XD	XD

		2 Hydroxylation(P)						
	1156 N	K.DGLNGLPGPIGP PGPR.G + 3 Hydroxylation(P)	X	XD	D	D	XD	XD

Conclusions

Asn and Gln deamidation and Met oxidation can be considered as molecular signatures of protein damage. We analyzed coeval ancient bone remains from Pompeii, Herculaneum and Oplontis, and bones from Baia Scalandrone S. Paolo Belsito with a bottom-up proteomic approach. During the Vesuvius eruption in AD 79, Pompeii and Oplontis were covered by a layer of 6 meters deep consisting of ash and cinder, although victims from Herculaneum were exposed to the pyroclastic surge high temperature [23,24]. Also, human bones related to a pre-historical eruption of Avellino (Avellino 3800 Myr, Naples, Italy) and human bones from a coeval skeletal population from the Campanian region were also considered (Baia Scalandrone site, II sec. AD, Roman imperial age, Puteoli, Naples, Italy). Despite the incredibly high temperatures during the eruption and the post-mortem factors that led to collagen aging, bone collagen was preserved in most of the samples as demonstrated by the high sequence coverages obtained in the herein presented analyses.

The samples with sequence coverages more than 20 % were used for statistical analysis, and manual interpretation of collagen chains type I was performed for an accurate evaluation of deamidation (Asn, Gln) and oxidation (Met). In a statistical analysis perspective, in some collagen zones, Asn, Gln, and Met are better detected than in others. At the same time, it is quite interesting that the poorly detectable zones are immediately followed by the best detectable ones, suggesting a structural consideration behind. Subsequently, the statistically most detected Asn, Gln, and Met were selected, and a second search was performed to investigate if these "hot spots" of detection also consist of hot spots of deamidation. In particular, we calculated the deamidation (Asn, Gln) and oxidation (Met) values at single sites along the collagen alpha-1 (I) and alpha-2 (I) chains, discovering that indeed the most frequently detected sites were also significantly modified. For this purpose, an in-house script based on peptides' intensities was used. Subsequently, we investigated whether these positions are always detected with the same peptide and the same combination of modifications. In up to 90% of the selected samples (33), the most deamidated and oxidized peptides are detected with the same peptides, and the most observed ms and ms/ms masses were listed. A series of criteria were applied for the selection of the final peptide-candidates that were eventually tested on already published archaeological bones. Interestingly, in the older samples (< 40ka BC), all of the peptides were identified both in the modified and unmodified form, and in the younger samples (18th

century), that are generally less deamidated, not all of the selected peptides were deamidated, observing in some cases that deamidated Asn, Gln belongs to longer peptides. Afterwards, the selected peptides were used for the calculation of deamidation (Asn, Gln) and oxidation (Met) using the in-house script described above. According to the fig S4 and S5 the samples demonstrate different modification levels that chemical modification increases from the younger to the older in the age samples. This fact confirms the suitability of the selected peptides for a more accurate quantitative study. It is worth mentioning that the peptide selection in this work results from a mass screening on experimental data. Usually, the literature's quantitation methods use peptides based on theoretical parameters regarding peptide ionization [25–27]. Therefore, we propose that this list of peptides can be used for quantitation studies, although the validation of their occurrence in different bone collagen matrices is still in progress.

Abbreviations

COL1A1: Collagen Type I Alpha 1 Chain; COL1A2: Collagen Type I Alpha 2 Chain; EDTA: Ethylenediaminetetraacetic acid; ESI: Electrospray ionization; MRM: multiple reaction monitoring;

Author Contributions

GN, LB, AD, GM, EC designed research; GN, AD, IRP carried out the experimental research; GN, IPR, PPP, EC LB analysed data; GN, IPR, PPP, EC LB wrote the paper.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area).

References

- [1] RC Hill, MJ Wither, T. Nemkov, A. Barrett, A. D'Alessandro, M. Dzieciatkowska, K.C. Hansen, Preserved proteins from extinct bison latifrons identified by tandem mass spectrometry; Hydroxylysine glycosides are a common feature of ancient collagen, *Mol. Cell. Proteomics*. 14 (2015) 1946–1958. <https://doi.org/10.1074/mcp.M114.047787>.
- [2] R.C. Dobberstein, M.J. Collins, O.E. Craig, G. Taylor, K.E.H. Penkman, S. Ritz-Timme, Archaeological collagen: Why worry about collagen diagenesis?, *Archaeol. Anthropol. Sci.* 1 (2009) 31–42. <https://doi.org/10.1007/s12520-009-0002-7>.
- [3] C. Wadsworth, M. Buckley, Proteome degradation in fossils: Investigating the longevity of protein survival in ancient bone, *Rapid Commun. Mass Spectrom.* 28 (2014) 605–615. <https://doi.org/10.1002/rcm.6821>.
- [4] C. Solazzo, J. Wilson, J.M. Dyer, S. Clerens, J.E. Plowman, I. Von Holstein, P. Walton Rogers, E.E. Peacock, M.J. Collins, Modeling deamidation in sheep α -keratin peptides and application to archeological

wool textiles, *Anal. Chem.* 86 (2014) 567–575. <https://doi.org/10.1021/ac4026362>.

[5] G. Leo, I. Bonaduce, A. Andreotti, G. Marino, P. Pucci, M.P. Colombini, L. Birolo, Deamidation at asparagine and glutamine As a major modification upon deterioration/aging of proteinaceous binders in mural paintings, *Anal. Chem.* 83 (2011) 2056–2064. <https://doi.org/10.1021/ac1027275>.

[6] NL van Doorn, J. Wilson, H. Hollund, M. Soressi, M.J. Collins, Site-specific deamidation of glutamine: a new marker of bone collagen deterioration, *Rapid Commun. Mass Spectrom.* 26 (2012) 2319–2327. <https://doi.org/10.1002/rcm.6351>.

[7] J. Wilson, N.L. Van Doorn, M.J. Collins, Assessing the extent of bone degradation using glutamine deamidation in collagen, *Anal. Chem.* 84 (2012) 9041–9048. <https://doi.org/10.1021/ac301333t>.

[8] F. Welker, M. Hajdinjak, S. Talamo, K. Jaouen, M. Dannemann, F. David, M. Julien, M. Meyer, J. Kelso, I. Barnes, S. Brace, P. Kamminga, R. Fischer, B.M. Kessler, J.R. Stewart, S. Pääbo, M.J. Collins, J.J. Hublin, Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 11162–11167. <https://doi.org/10.1073/pnas.1605834113>.

[9] E.R. Schroeter, T.P. Cleland, Glutamine deamidation: An indicator of antiquity, or preservational quality?, *Rapid Commun. Mass Spectrom.* 30 (2016) 251–255. <https://doi.org/10.1002/rcm.7445>.

[10] M. Mackie, P. Rütger, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Peggie, H. Howard, L. Harrison, L.J. Jensen, J. V. Olsen, E. Cappellini, Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting, *Angew. Chemie - Int. Ed.* 57 (2018) 7369–7374. <https://doi.org/10.1002/anie.201713020>.

[11] T.P. Cleland, E.R. Schroeter, M.H. Schweitzer, Biologically and diagenetically derived peptide modifications in moa collagens, *Proc. R. Soc. B Biol. Sci.* 282 (2015) 20150015–20150015. <https://doi.org/10.1098/rspb.2015.0015>.

[12] N. Procopio, A.T. Chamberlain, M. Buckley, Exploring Biological and Geological Age-related Changes through Variations in Intra- and Intertooth Proteomes of Ancient Dentine, *J. Proteome Res.* 17 (2018) 1000–1013. <https://doi.org/10.1021/acs.jproteome.7b00648>.

[13] R. Sawafuji, E. Cappellini, T. Nagaoka, A.K. Fotakis, R. Rakownikow, Jersie-Christensen, J. V Olsen, K. Hirata, S. Ueda, Proteomic profiling of archaeological human bone, (n.d.). <https://doi.org/10.1098/rsos.161004>.

[14] V. Vidova, Z. Spacil, A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition, *Anal. Chim. Acta.* 964 (2017) 7–23. <https://doi.org/10.1016/j.aca.2017.01.059>.

[15] R. Vinciguerra, A. Illiano, A. De Chiaro, A. Carpentieri, A. Lliveras-Tenorio, I. Bonaduce, G. Marino, P. Pucci, A. Amoresano, L. Birolo, Identification of proteinaceous binders in paintings: A targeted proteomic approach for cultural heritage, *Microchem. J.* 144 (2019) 319–328. <https://doi.org/10.1016/j.microc.2018.09.021>.

[16] P. Hao, S.S. Adav, X. Gallart-Palau, S.K. Sze, Recent advances in mass spectrometric analysis of protein deamidation, *Mass Spectrom. Rev.* 36 (2017) 677–692. <https://doi.org/10.1002/mas.21491>.

[17] G. Leo, L. Cartechini, P. Pucci, A. Sgamellotti, G. Marino, L. Birolo, Proteomic strategies for the identification of proteinaceous binders in paintings, *Anal. Bioanal. Chem.* 395 (2009) 2269–2280. <https://doi.org/10.1007/s00216-009-3185-y>.

[18] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A.R. Da Fonseca, T.W. Stafford, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, M.T.P. Gilbert, J. V. Olsen, Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, *J. Proteome Res.* 11 (2012) 917–926. <https://doi.org/10.1021/pr200721u>.

[19] T.P. Cleland, Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics, *J. Proteome Res.* 17 (2018) 3976–3983. <https://doi.org/10.1021/acs.jproteome.8b00637>.

[20] L.T. Lanigan, M.; Mackie, S.; Feine, J.-J.; Hublin, R.W. Schmitz, A.; Wilcke, . . Welker, Multi-protease analysis of Pleistocene bone proteomes, *J. Proteomics.* 228 (2020).

<https://doi.org/10.1016/j.jprot.2020.103889>.

[21] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372. <https://doi.org/10.1038/nbt.1511>.

[22] E. Vuorio, B. De Crombrughe, The family of collagen genes, *Annu. Rev. Biochem.* 59 (1990) 837–872. <https://doi.org/10.1146/annurev.bi.59.070190.004201>.

[23] G. Mastrolorenzo, P.P. Petrone, M. Pagano, A. Incoronato, P.J. Baxter, A. Canzanella, L. Fattore, Archaeology: Herculaneum victims of Vesuvius in AD 79, *Nature*. 410 (2001) 769–770. <https://doi.org/10.1038/35071167>.

[24] P. Petrone, P. Pucci, A. Vergara, A. Amoresano, L. Birolo, F. Pane, F. Sirano, M. Niola, C. Buccelli, V. Graziano, A hypothesis of sudden body fluid vaporization in the 79 AD victims of Vesuvius, *PLoS One*. 13 (2018). <https://doi.org/10.1371/journal.pone.0203210>.

[25] Q. Chen, Y. Jiang, Y. Ren, M. Ying, B. Lu, Peptide Selection for Accurate Targeted Protein Quantification via a Dimethylation High-Resolution Mass Spectrum Strategy with a Peptide Release Kinetic Model, *ACS Omega*. 5 (2020) 3809–3819. <https://doi.org/10.1021/acsomega.9b02002>.

[26] N.C. Van De Merbel, Protein quantification by LC-MS: A decade of progress through the pages of *Bioanalysis*, *Bioanalysis*. 11 (2019) 629–644. <https://doi.org/10.4155/bio-2019-0032>.

[27] E. Song, Y. Gao, C. Wu, T. Shi, S. Nie, T.L. Fillmore, A.A. Schepmoes, M.A. Gritsenko, W.J. Qian, R.D. Smith, K.D. Rodland, T. Liu, Targeted proteomic assays for quantitation of proteins identified by proteogenomic analysis of ovarian cancer, *Sci. Data*. 4 (2017) 1–13. <https://doi.org/10.1038/sdata.2017.91>.

Supporting Information

Selection of the best samples to be used for evaluation of deamidation (N,Q) and methionine oxidation (M) level.

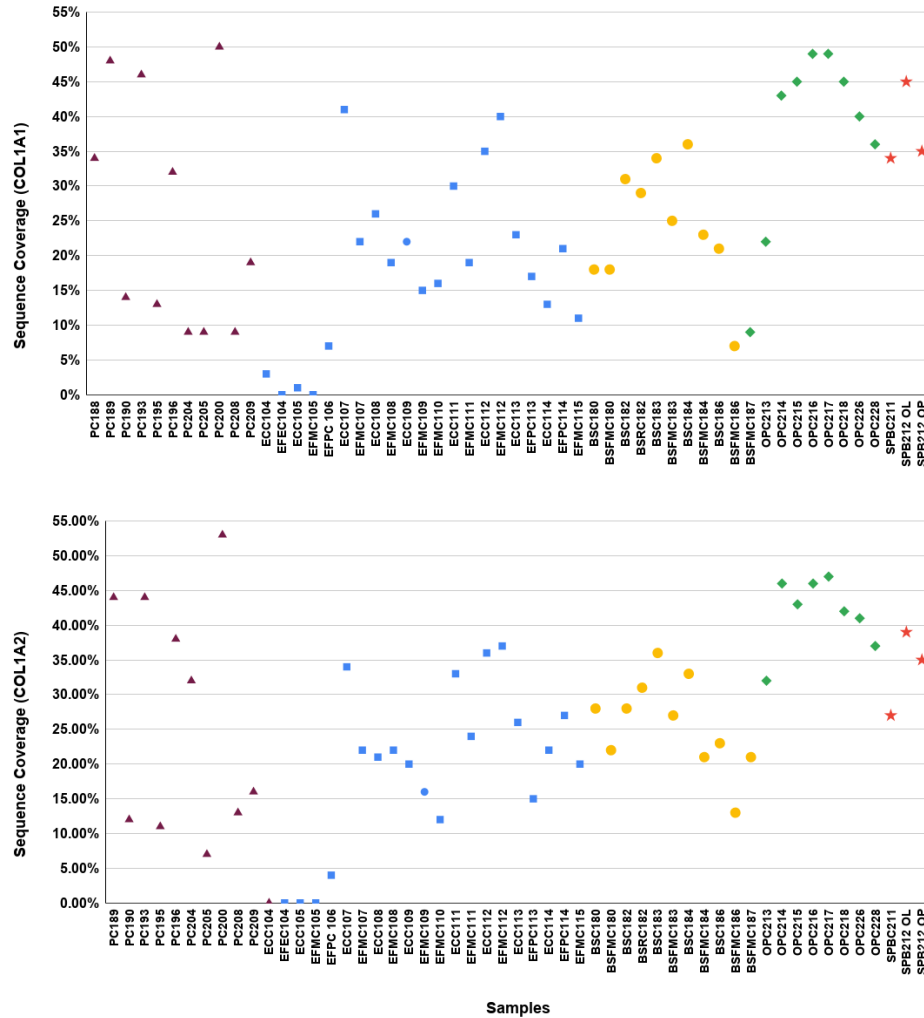


Figure S1: Schematic representation of the protein sequence coverages of COL1A1 (upper panel) and COL1A2 (lower panel) in the human bone samples of Pompeii, Herculaneum, Baia Scalandrone, Oplontis and San Paolo Bel Sito.

Evaluation of the most frequently detected N,Q and M in collagen chains type I.

Table S1: Deamidation sites along the sequence of collagen alpha-1(I) in Pompeii samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	PC188	PC189	PC193	PC196	PC200	PC204
Collagen Sequence Coverage	34%	48%	46%	32%	50%	28%
Protein score	1363	2341	2278	1206	2651	964
199 Q	X	X	X	N	N	N
202 Q	X	X	D	N	N	N
229 N	N	N	N	N	N	N
250 Q	N	N	N	N	N	N
295 N	X	X	N	X	XD	N
300 Q	XD	X	N	X	XD	N
324 N	N	XD	XD	XD	X	X
358 Q	N	XD	X	XD	XD	N
367 Q	N	XD	X	N	XD	N
387 N	X	XD	X	X	XD	N
393 Q	XD	XD	XD	D	XD	N
400 N	N	D	XD	X	X	N
421 Q	D	XD	XD	X	XD	D
432 N	N	N	N	N	XD	N
457 Q	X	XD	XD	XD	XD	D
561 Q	D	D	D	D	XD	D
576 Q	X	XD	XD	XD	XD	XD
619 Q	XD	XD	XD	X	XD	X
634 Q	N	N	X	N	N	N
643 Q	N	N	X	N	N	N
661 Q	N	N	D	N	N	N
688 Q	D	XD	XD	X	XD	D
700 N	XD	XD	N	XD	X	N
705 N	D	N	N	XD	XD	N
721 Q	XD	X	XD	XD	XD	XD
727 Q	XD	X	XD	XD	XD	X
822 Q	D	XD	XD	XD	N	XD
855N	X	XD	XD	XD	XD	X
891N	D	D	N	N	XD	D
952 Q	X	XD	XD	D	XD	X
957 Q	X	XD	XD	X	XD	X

966 Q	X	XD	XD	X	X	X
985 Q	X	X	N	X	N	N
1090 Q	N	X	X	N	D	N
1102 Q	N	N	N	N	X	N
1117 Q	X	X	XD	X	X	X
1129 Q	XD	XD	XD	XD	XD	XD
1156 N	XD	XD	D	X	XD	XD

Table S2: Deamidation sites along the sequence of collagen alpha-1(II) in Pompeii samples. with collagen PSQ >20. When the site is labeled as X indicates that the site was detected in the unmodified form, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated forms were detected. N are the not detected peptides.

Position AA	PC188	PC189	PC193	PC196	PC200	PC204
Protein score	1124	2451	2441	1244	2841	1252
Collagen Sequence Coverage	22%	44%	44%	38%	53%	32%
111 Q	N	N	N	X	N	N
114 Q	N	N	N	D	N	N
125 Q	N	N	N	X	N	N
162 Q	N	XD	X	N	D	N
183 N	N	N	N	N	N	N
191 Q	N	N	N	N	N	N
207 N	N	N	N	N	XD	N
212 Q	N	N	N	N	D	N
272 N	X	XD	XD	D	XD	D
299 N	X	XD	XD	X	XD	XD
303 N	X	XD	XD	X	XD	XD
359 N	N	N	D	N	N	N
369 Q	N	N	X	N	N	N
384 N	X	X	XD	X	XD	XD
435 N	N	XD	XD	N	XD	D
458 N	X	XD	D	XD	XD	X
491 N	X	XD	XD	XD	X	XD
507 N	N	N	N	D	XD	N
527 N	N	N	N	N	N	N
528 N	N	N	N	N	N	N
531 Q	N	N	N	N	N	N
537 Q	N	N	N	N	N	N
540 Q	N	N	N	N	N	N

546 Q	N	N	N	N	N	N
555 Q	N	N	N	N	N	N
620 N	D	XD	XD	XD	XD	D
668 N	N	N	N	N	X	N
723 N	N	XD	XD	D	XD	X
734 Q	N	XD	XD	D	XD	XD
750 N	D	XD	XD	D	XD	X
768 N	D	XD	XD	N	XD	X
825 Q	N	N	N	N	N	N
864 Q	N	XD	XD	X	XD	XD
918 N	XD	XD	XD	D	XD	XD
929 N	D	X	N	D	XD	X
932 N	D	X	N	XD	XD	D
941 Q	N	N	N	N	N	N
953 N	N	N	N	D	X	N
977 N	N	XD	D	D	XD	N
1002 Q	N	N	D	N	XD	N

Table S3: Deamidation sites along the sequence of collagen alpha-1(I) in Herculeaneum samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	ECC10 7	EFMC1 07	ECC108	ECC109	ECC111	ECC112	EFMC11 2	ECC113	EFPC11 4
Protein score	1791	535	761	592	1184	1314	1451	545	476
Collagen Sequence Coverage	41%	22%	26%	22%	30%	35%	40%	23%	21%
199 Q	N	N	N	N	N	X	X	N	N
202 Q	N	N	N	N	N	X	XD	N	N
229 N	N	N	N	N	N	N	N	N	N
250 Q	X	N	N	N	N	N	N	N	N
295 N	N	N	N	X	D	XD	XD	N	N
300 Q	N	N	X	X	X	XD	X	N	N
324 N	XD	XD	X	XD	XD	XD	XD	XD	D
358 Q	X	XD	X	D	N	N	XD	N	N
367 Q	XD	N	N	X	N	XD	XD	N	X
387 N	X	N	N	X	X	XD	X	N	X
393 Q	X	N	N	X	X	XD	XD	N	X
400 N	D	X	X	XD	XD	XD	XD	D	N

421 Q	XD	N	N	N	D	N	XD	N	N
432 N	N	N	N	N	N	N	N	N	N
457 Q	X	N	N	N	N	X	N	N	N
561 Q	N	N	N	N	N	N	N	N	N
576 Q	XD	N	N	N	X	XD	X	N	N
619 Q	XD	X	XD	N	XD	XD	X	N	N
634 Q	N	N	X	N	N	N	N	X	N
643 Q	N	N	X	N	N	N	N	X	N
661 Q	N	N	X	N	N	N	N	X	N
688 Q	N	N	N	N	N	N	N	N	N
700 N	N	XD	XD	XD	N	XD	XD	N	XD
705 N	N	XD	D	XD	N	XD	XD	N	XD
721 Q	X	X	XD	XD	X	XD	XD	N	X
727 Q	X	XD	XD	XD	X	XD	XD	N	XD
822 Q	X	N	X	XD	XD	XD	XD	N	D
855N	X	X	XD	X	XD	XD	D	XD	X
891N	D	D	N	N	N	N	XD	N	N
952 Q	X	X	XD	XD	XD	XD	X	X	X
957 Q	XD	X	XD	XD	XD	XD	X	X	XD
966 Q	D	N	N	N	N	N	N	N	N
985 Q	N	X	N	X	D	N	N	N	N
1090 Q	X	N	N	N	N	N	N	N	N
1102 Q	N	N	N	N	N	N	N	N	N
1117 Q	X	X	XD	X	XD	XD	XD	XD	X
1129 Q	XD	XD	XD	XD	XD	XD	XD	XD	XD
1156 N	XD	N	N	N	XD	XD	D	XD	X

Table S4: Deamidation sites along the sequence of collagen alpha-1(II) in Herculaneum samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	ECC1 07	EFMC1 07	ECC10 8	ECC10 9	EFMC1 09	ECC11 1	ECC11 2	EFMC1 12	ECC11 3	EFPC1 14
Protein score	1670	675	711	627	449	1358	1430	1422	1056	971
Collagen Sequence Coverage	34%	22%	21%	20%	16%	33%	36%	37%	26%	27%
111 Q	N	N	N	N	N	X	X	D	N	N
114 Q	N	N	N	N	N	X	X	X	N	N
125 Q	N	N	N	N	N	X	X	D	N	N
162 Q	D	N	N	N	N	N	N	N	N	N
183 N	N	N	N	N	N	N	N	N	N	N
191 Q	N	N	N	N	N	N	N	N	N	N
207 N	N	N	N	N	N	N	N	N	N	N
212 Q	N	N	N	N	N	N	N	N	N	N
272 N	D	D	XD	D	N	XD	XD	XD	X	X
299 N	X	N	XD	X	X	X	XD	XD	XD	XD
303 N	XD	N	XD	D	XD	XD	XD	D	XD	XD
359 N	N	N	N	N	N	XD	X	N	N	N
369 Q	N	N	N	N	N	XD	X	N	N	N
384 N	XD	XD	XD	XD	N	N	X	XD	N	D
435 N	N	N	N	N	N	N	XD	N	N	N
458 N	XD	N	N	N	N	N	D	X	N	N
491 N	XD	N	N	N	N	N	N	XD	N	N
507 N	N	N	N	N	N	N	N	N	N	N
527 N	N	N	N	N	N	N	N	N	N	N
528 N	N	N	N	N	N	N	N	N	N	N
531 Q	N	N	N	N	N	N	N	N	N	N
537 Q	N	N	N	N	N	N	N	N	N	N
540 Q	N	N	N	N	N	N	N	N	N	N
546 Q	N	N	N	N	N	X	N	N	N	N
555 Q	N	N	N	N	N	D	N	N	N	N
620 N	XD	D	XD	XD	XD	XD	XD	XD	XD	XD
668 N	N	N	N	N	N	N	N	N	N	N
723 N	XD	XD	XD	X	XD	XD	XD	XD	XD	XD
734 Q	XD	XD	XD	X	X	XD	XD	XD	XD	XD
750 N	D	XD	D	D	D	XD	XD	XD	XD	XD
768 N	D	X	N	XD	D	XD	XD	XD	XD	XD

825 Q	N	N	N	N	N	N	N	N	N	N
864 Q	XD	X	XD	X	X	XD	XD	XD	XD	XD
918 N	XD	XD	X	D	D	XD	XD	XD	N	X
929 N	X	N	N	N	N	N	N	N	N	D
932 N	XD	N	N	N	N	N	N	N	N	D
941 Q	N	N	N	N	N	N	N	N	N	N
953 N	N	N	N	N	N	N	N	N	N	N
977 N	D	N	N	N	N	N	N	N	N	N
1002 Q	N	N	N	N	N	N	N	N	N	N

Table S5: Deamidation sites along the sequence of collagen alpha-1(I) in Scaladrone Bay samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	BSC182	BSRC182	BSC183	BSFMC1 83	BSC184	BSFMC1 84	BSC186	BSC187
Protein score	1071	1193	1326	967	1448	775	445	3242
Collagen Sequence Coverage	31%	29%	34%	25%	36%	23%	21%	55%
199 Q	N	N	N	N	N	N	N	D
202 Q	N	N	N	N	N	N	N	N
229 N	N	N	N	N	N	N	N	XD
250 Q	N	N	N	N	N	N	N	XD
295 N	X	N	N	X	X	X	N	XD
300 Q	X	N	N	X	X	X	N	N
324 N	N	N	X	X	XD	D	X	N
358 Q	N	X	N	X	X	N	D	D
367 Q	XD	X	D	X	N	N	N	N
387 N	D	N	N	N	X	X	N	X
393 Q	X	N	N	N	D	X	N	D
400 N	N	XD	XD	D	D	XD	N	X
421 Q	XD	XD	XD	XD	D	N	N	XD
432 N	N	N	N	N	N	N	N	XD
457 Q	N	N	X	X	XD	N	N	XD
561 Q	N	N	N	N	N	N	N	XD
576 Q	XD	XD	XD	XD	XD	X	N	XD
619 Q	X	XD	XD	N	XD	XD	D	XD
634 Q	N	N	N	N	N	N	N	XD
643 Q	N	N	N	N	N	N	N	XD

661 Q	N	N	N	N	N	N	N	XD
688 Q	D	X	N	X	D	XD	N	XD
700 N	XD	N	N	XD	X	N	X	XD
705 N	X	N	N	XD	D	N	N	X
721 Q	XD	X	XD	XD	D	N	N	X
727 Q	X	X	N	XD	D	N	N	D
822 Q	X	X	X	N	X	X	XD	X
855N	X	X	XD	N	X	N	N	XD
891N	N	N	D	D	XD	N	N	XD
952 Q	N	X	X	N	X	N	XD	XD
957 Q	N	XD	XD	N	XD	N	X	XD
966 Q	N	D	X	XD	N	N	N	XD
985 Q	N	N	N	N	N	N	X	X
1090 Q	X	XD	X	X	D	X	N	XD
1102 Q	N	N	N	N	N	N	N	XD
1117 Q	X	X	XD	X	X	X	X	XD
1129 Q	XD	XD	XD	XD	XD	X	XD	X
1156 N	D	D	XD	N	XD	N	D	X

Table S6: Deamidation sites along the sequence of collagen alpha-1(II) in Scaladrone Bay samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	BSC182	BSRC182	BSC183	BSFMC183	BSC184	BSFMC184	BSC186	BSC187
Protein score	1069	1353	1482	878	1301	866	768	3157
Collagen Sequence Coverage	28%	31%	36%	27%	33%	21%	23%	54%
111 Q	N	N	N	N	X	N	N	N
114 Q	N	N	N	N	X	N	N	N
125 Q	N	N	N	N	X	N	N	N
162 Q	N	X	N	X	N	N	N	N
183 N	N	N	N	N	N	N	N	N
191 Q	N	N	N	N	N	N	N	N
207 N	N	N	N	N	N	D	D	X
212 Q	N	N	N	N	N	D	D	D
272 N	XD	D	D	XD	XD	X	X	XD
299 N	N	X	X	N	XD	X	X	XD
303 N	N	D	D	N	D	D	D	XD
359 N	N	N	X	N	X	X	X	X

369 Q	N	N	X	N	D	X	X	X
384 N	N	X	X	N	X	X	X	XD
435 N	N	N	N	N	N	N	N	D
458 N	XD	XD	XD	X	XD	N	N	X
491 N	XD	XD	XD	XD	XD	N	N	X
507 N	N	N	N	N	N	N	N	D
527 N	N	N	N	N	N	N	N	XD
528 N	N	N	N	N	N	N	N	X
531 Q	N	N	N	N	N	N	N	XD
537 Q	N	N	N	N	N	N	N	XD
540 Q	N	N	N	N	N	N	N	X
546 Q	N	N	N	N	N	N	N	N
555 Q	N	N	N	N	N	N	N	N
620 N	D	X	XD	XD	XD	X	X	X
668 N	N	N	N	N	N	N	N	N
723 N	XD	X	XD	X	XD	XD	XD	XD
734 Q	XD	XD	XD	X	XD	XD	XD	X
750 N	D	D	X	N	D	XD	XD	XD
768 N	X	XD	XD	N	XD	XD	XD	XD
825 Q	N	N	N	N	N	N	N	N
864 Q	D	N	XD	XD	X	XD	XD	XD
918 N	XD	XD	X	X	XD	N	N	XD
929 N	N	N	N	N	XD	N	N	XD
932 N	N	N	N	N	D	N	N	XD
941 Q	N	N	N	N	N	N	N	N
953 N	N	N	N	N	N	N	N	X
977 N	N	N	N	N	N	N	N	XD
1002 Q	X	X	X	X	N	N	N	X

Table S7: Deamidation sites along the sequence of collagen alpha-1(I) in Oplontis samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	OPC214	OPC215	OPC216	OPC217	OPC218	OPC226	OPC228
Protein score	1887	2128	2322	2196	1849	1623	1626
Collagen Sequence Coverage	43%	45%	49%	49%	45%	40%	36%
199 Q	N	X	X	XD	X	X	N
202 Q	N	X	X	D	D	X	N
229 N	N	N	N	N	N	N	N
250 Q	N	N	N	N	N	N	N
295 N	X	X	X	X	X	X	N
300 Q	X	XD	XD	XD	X	X	N
324 N	XD	XD	XD	XD	XD	X	XD
358 Q	XD	XD	XD	XD	XD	XD	N
367 Q	D	XD	D	XD	XD	X	D
387 N	D	XD	XD	XD	XD	N	N
393 Q	D	XD	XD	XD	XD	N	XD
400 N	XD	XD	XD	XD	XD	XD	XD
421 Q	XD	XD	XD	XD	XD	XD	XD
432 N	X	N	N	N	N	N	N
457 Q	XD	XD	XD	XD	XD	XD	X
561 Q	N	N	D	XD	X	N	N
576 Q	XD	XD	XD	XD	XD	XD	XD
619 Q	N	XD	XD	XD	N	XD	XD
634 Q	N	N	N	N	N	N	X
643 Q	N	N	N	N	N	N	XD
661 Q	N	N	N	N	N	N	XD
688 Q	D	N	XD	XD	XD	D	N
700 N	XD	XD	XD	XD	XD	X	N
705 N	XD	XD	XD	XD	XD	D	N
721 Q	XD	XD	XD	XD	XD	XD	XD
727 Q	XD	X	XD	XD	XD	X	X
822 Q	X	XD	XD	XD	X	XD	XD
855N	XD	XD	XD	XD	X	XD	XD
891N	D	XD	XD	XD	XD	X	N
952 Q	X	XD	XD	XD	X	X	XD
957 Q	XD	XD	XD	XD	X	X	XD
966 Q	D	N	XD	XD	XD	N	X

985 Q	N	X	X	N	X	N	N
1090 Q	N	X	N	X	X	N	N
1102 Q	N	N	N	N	N	N	N
1117 Q	X	X	X	X	X	X	X
1129 Q	XD	XD	XD	XD	XD	XD	XD
1156 N	XD	XD	D	XD	N	XD	X

Table S8: Deamidation sites along the sequence of collagen alpha-1(II) in Oplontis samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	OPC214	OPC215	OPC216	OPC217	OPC218	OPC226	OPC228
Protein score	2236	1940	2181	2451	1783	2121	1766
Collagen Sequence Coverage	46%	43%	46%	47%	42%	41%	37%
111 Q	N	N	N	X	N	N	N
114 Q	N	N	N	X	N	N	N
125 Q	N	N	N	D	N	N	N
162 Q	XD	XD	D	XD	X	XD	D
183 N	N	N	N	N	N	N	N
191 Q	N	N	N	N	N	N	N
207 N	N	N	N	N	N	N	N
212 Q	N	N	N	N	N	N	N
272 N	XD	XD	XD	XD	XD	XD	XD
299 N	X	XD	XD	XD	D	XD	XD
303 N	XD	XD	XD	XD	D	XD	XD
359 N	X	N	XD	N	N	N	N
369 Q	X	N	D	N	N	N	N
384 N	XD	XD	XD	X	D	XD	XD
435 N	D	D	XD	XD	XD	X	X
458 N	XD	XD	XD	XD	XD	XD	X
491 N	XD	XD	XD	XD	XD	XD	X
507 N	N	N	N	N	N	N	N
527 N	N	N	N	N	N	N	N
528 N	N	N	N	N	N	N	N
531 Q	N	N	N	N	N	N	N
537 Q	N	N	N	N	N	N	N
540 Q	N	N	N	N	N	N	N
546 Q	N	N	N	N	N	N	N

555 Q	N	N	N	N	N	N	N
620 N	XD	XD	XD	XD	XD	XD	XD
668 N	N	N	N	N	N	N	N
723 N	XD	XD	XD	N	N	X	X
734 Q	XD	XD	XD	XD	XD	XD	XD
750 N	D	N	XD	XD	D	XD	N
768 N	X	XD	XD	XD	XD	XD	XD
825 Q	N	N	N	N	N	N	N
864 Q	N	XD	XD	XD	D	XD	XD
918 N	N	XD	XD	XD	N	XD	XD
929 N	N	D	D	D	N	N	N
932 N	N	XD	XD	X	N	N	N
941 Q	N	N	N	N	N	N	N
953 N	D	X	N	N	X	D	N
977 N	XD	N	XD	XD	XD	XD	N
1002 Q	X	D	D	XD	XD	N	D

Table S9: Deamidation sites along the sequence of collagen alpha-1(I) in San Paolo Belsito samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	SPBC211	SPB212 OL	SPB212 OP
Protein score	1146	1951	1305
Collagen Sequence Coverage	34%	45%	35%
199 Q	X	X	X
202 Q	D	X	X
229 N	N	N	N
250 Q	N	N	N
295 N	X	X	XD
300 Q	X	D	XD
324 N	X	XD	XD
358 Q	N	N	N
367 Q	D	D	N
387 N	X	XD	N
393 Q	X	XD	N
400 N	D	D	X
421 Q	X	XD	XD
432 N	N	N	N

457 Q	X	XD	XD
561 Q	N	X	D
576 Q	N	XD	XD
619 Q	X	XD	N
634 Q	N	XD	N
643 Q	N	D	N
661 Q	N	D	N
688 Q	D	X	N
700 N	XD	N	XD
705 N	D	N	XD
721 Q	D	XD	D
727 Q	XD	XD	X
822 Q	XD	XD	N
855N	X	XD	X
891N	N	N	X
952 Q	N	XD	X
957 Q	N	XD	XD
966 Q	N	XD	XD
985 Q	N	N	N
1090 Q	N	N	N
1102 Q	N	N	N
1117 Q	X	X	X
1129 Q	X	XD	XD
1156 N	XD	D	X

Table S10: Deamidation sites along the sequence of collagen alpha-1(II) in San Paolo Belsito with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not deamidated, while D indicates that the site was observed in the deamidated form. X-D indicates that both deamidated and non deamidated form were detected. N are the not detected peptides.

Position AA	SPBC211	SPB212 OL	SPB212 OP
Protein score	1069	1948	1523
Collagen Sequence Coverage	27%	39%	35%
111 Q	N	N	N
114 Q	N	N	N
125 Q	N	N	N
162 Q	N	XD	X
183 N	N	N	N
191 Q	N	N	N
207 N	N	N	N
212 Q	N	N	N
272 N	X	XD	XD
299 N	X	XD	N
303 N	XD	XD	N
359 N	N	N	N
369 Q	N	N	N
384 N	N	X	N
435 N	D	XD	X
458 N	XD	XD	XD
491 N	N	XD	XD
507 N	N	N	N
527 N	N	N	N
528 N	N	N	N
531 Q	N	N	N
537 Q	N	N	N
540 Q	N	N	N
546 Q	N	N	N
555 Q	N	N	N
620 N	X	XD	XD
668 N	N	N	N
723 N	D	N	XD
734 Q	D	XD	D
750 N	X	XD	X
768 N	X	X	X
825 Q	N	N	N

864 Q	X	XD	D
918 N	X	XD	X
929 N	N	N	N
932 N	N	N	N
941 Q	N	N	N
953 N	N	X	N
977 N	N	D	D
1002 Q	N	D	N

Table S11: Methionine oxidation sites along the sequence of collagen alpha-1(I) in all samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not oxidized, while O indicates that the site was observed in the oxidized form. OX indicates that both oxidized and non oxidized form were detected. N are the not detected peptides.

Sample Name	181 M	217 M	264 M	301 M	580 M	729 M	1000 M
BSCC182	N	N	N	O	O	O	N
BSRC182	N	N	N	N	O	O	O
BSC183	N	N	N	N	OX	O	O
BSFMC183	N	N	N	O	OX	O	N
BSCC184	N	N	N	O	O	O	N
BSFMC184	N	N	N	O	OX	N	N
BSCC186	N	N	N	N	N	O	O
BSCC187	N	O	N	N	N	N	N
SPBC211	N	O	N	O	N	O	O
SPBC212 OL	N	O	N	O	O	O	O
SPBC212 OP	N	O	N	O	O	O	N
OC214	N	N	N	XO	XO	XO	N
OC215	N	O	N	O	O	O	O
OC216	N	O	N	O	O	O	O
OC217	N	O	N	O	XO	O	XO
OC218	N	O	N	O	O	O	O
OC226	N	O	N	O	XO	O	O
OC228	N	N	N	X	O	O	O
PC188	N	O	N	XO	N	O	O
PC189	N	O	N	O	O	O	XO
PC193	N	O	N	N	O	O	XO
PC196	N	N	N	O	O	O	O
PC200	N	N	N	X	O	O	O
ECC107	X	N	N	X	N	O	O

EFMC107	N	N	N	N	N	XO	O
ECC108	N	N	N	O	N	XO	N
ECC109	N	N	N	XO	N	XO	O
ECC111	N	N	N	X	O	O	O
ECC112	N	O	N	XO	XO	O	X
EFMC112	N	O	N	O	XO	XO	O
ECC113	N	N	N	N	N	N	X
EFPC114	N	N	N	N	N	O	N

Table S12: Methionine oxidation sites along the sequence of collagen alpha-1(II) in all samples with collagen protein sequence coverage >20. When the site is labeled as X indicates that the site was detected but is not oxidized, while O indicates that the site was observed in the oxidized form. OX indicates that both oxidized and non oxidized form were detected. N are the not detected peptides.

Sample Name	93 M	96 M	417 M	447 M	785 M
BSCC182	N	N	O	N	O
BSRC182	N	N	N	N	O
BSC183	N	N	N	N	O
BSFMC183	N	N	N	N	O
BSCC184	N	N	N	N	O
BSFMC184	N	N	N	N	O
BSCC186	N	N	N	N	N
BSFMC186	N	N	N	N	O
BSCC187	N	N	N	N	N
SPBC211	N	N	N	O	N
SPBC212 OL	N	N	N	O	N
SPBC212 OP	N	N	O	O	O
OC214	N	N	O	O	XO
OC215	N	N	O	O	O
OC216	N	N	O	O	O
OC217	N	N	O	XO	O
OC218	N	N	O	O	O
OC226	N	N	N	N	XO
OC228	N	N	N	O	O
PC188	N	N	O	N	N
PC189	N	N	O	O	O
PC193	N	N	O	O	O
PC196	N	N	N	O	O
PC200	N	N	N	O	O

ECC107	N	N	N	N	X
EFMC107	N	N	N	N	O
ECC108	N	N	N	N	N
ECC109	N	N	N	N	N
ECC111	N	N	N	N	O
EFMC111	N	N	N	N	O
ECC112	N	N	N	O	N
EFMC112	N	N	N	N	XO
ECC113	N	N	N	N	N
EFPC114	N	N	N	N	N

Evaluation of the deamidation (N,Q) and oxidation (M) modification levels along collagen chain type I.



Figure S2: 3D visualization of deamidation level (N,Q) along the sequence of collagen alpha-1 (I) and collagen alpha-2 (I) chains (upper panel) and, for greater clarity, zoomed in the zones of the selected peptides (lower panel), in the groups of Pompeii, Herculaneum, Baia Scalandrone bay, Oplontis and San Paolo Bel Sito. The label size indicates the relative intensity of each position in the single samples.

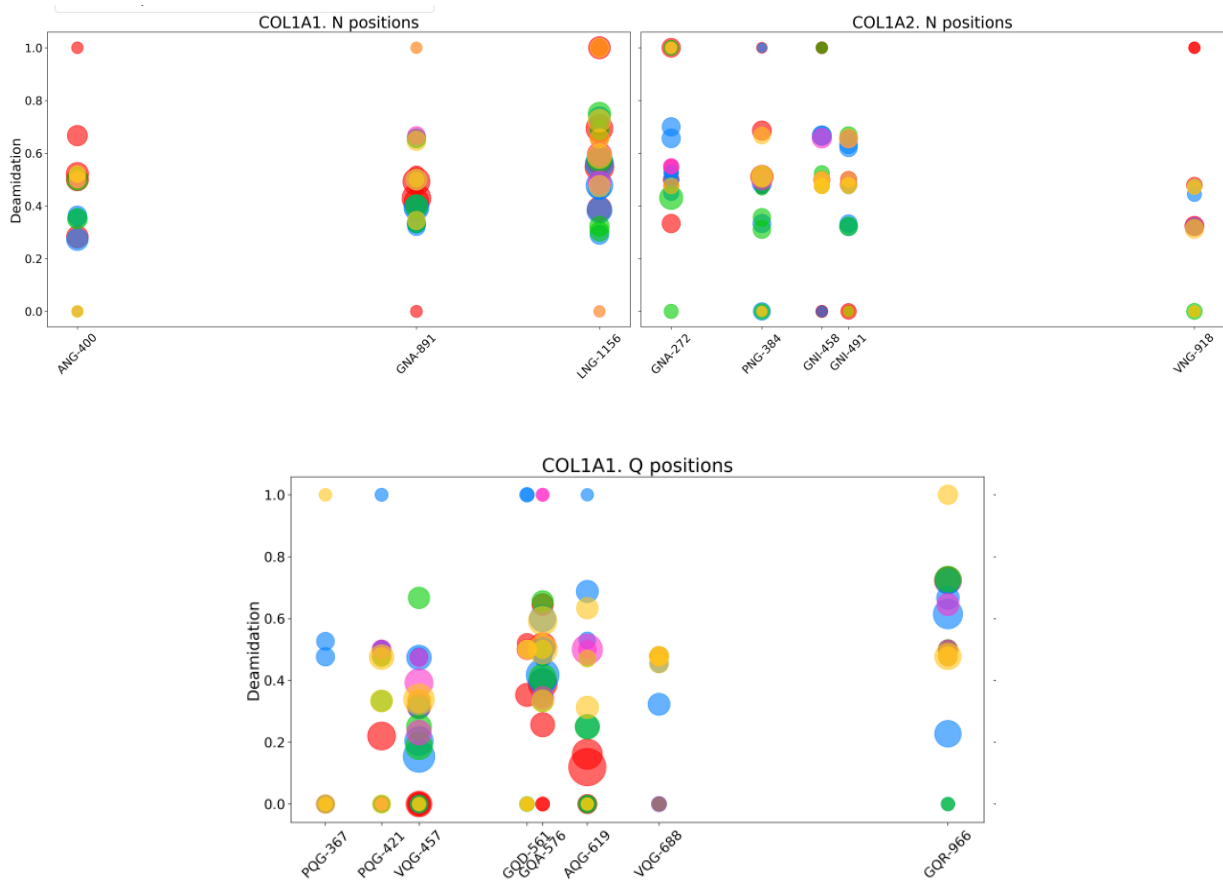


Figure S2: 3D visualization of deamidation level (N,Q) zoomed in the zones of the selected peptides, along the sequence of collagen alpha-1 (I) and collagen alpha-2 (I) chains in the groups of Pompeii, Herculaneum, Baia Scalandrone bay, Oplontis and San Paolo Bel Sito. The label size indicates the relative intensity of each position in the single samples.

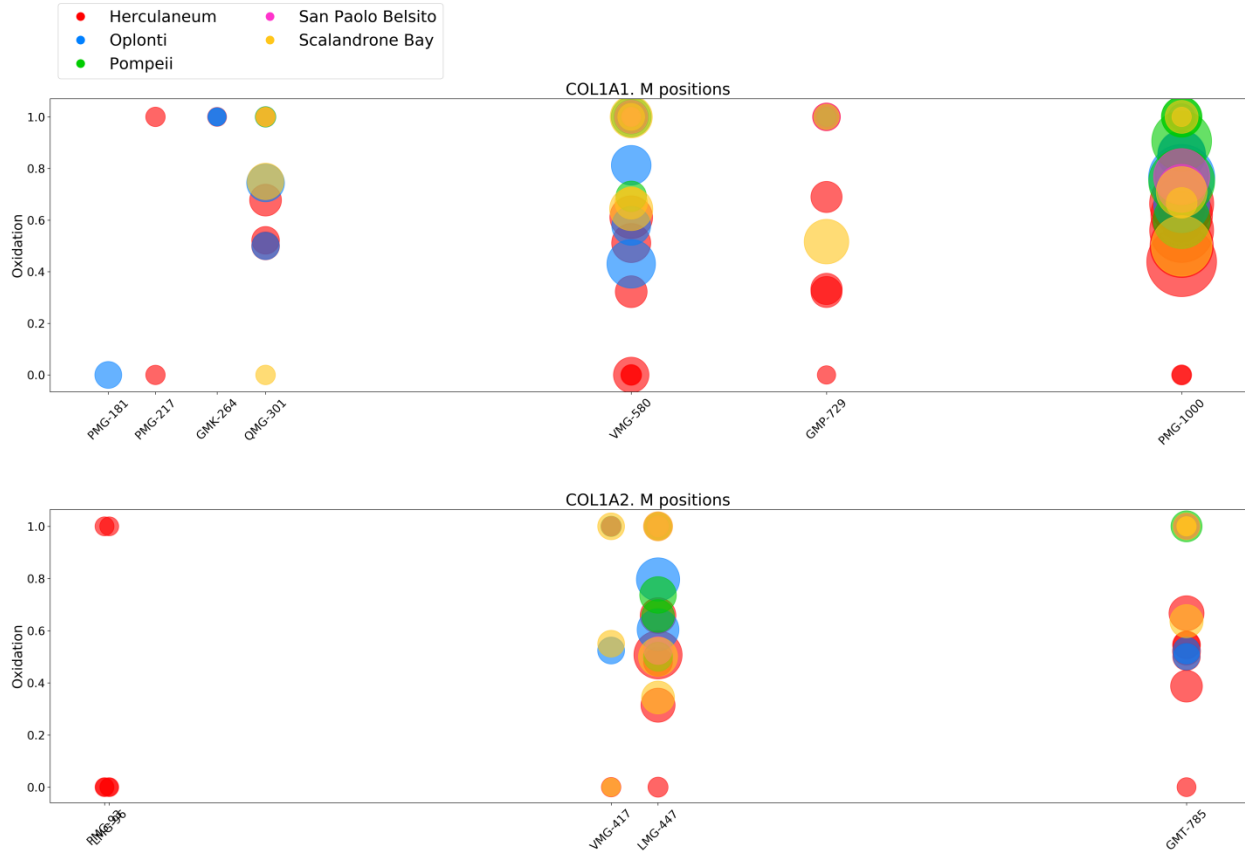


Figure S3: 3D visualization of oxidation level (M) along the sequence of collagen alpha-1 (I) and collagen alpha-2 (I) chains (upper panel) and, for greater clarity, zoomed in the zones of the selected peptides (lower panel), in the groups of Pompeii, Herculaneum, Baia Scalandrone bay, Oplontis and San Paolo Bel Sito. The label size indicates the relative intensity of each position in the single samples.

Selection of the human collagen type I sequences to be used as biomarkers for deamidation and methionine oxidation.

Table S13: Selection of the most frequently detected Asn and Gln containing peptides in COL1A1 (P02452) from the 33 samples of Pompeii, Herculaneum, Baia Scalandrone, Oplontis and San Paolo Bel Sito. The most frequently masses are reported. Z: Charge ; PL: Peptide Length ; H: Hydrophobicity. The mass of the corresponding deamidated form can be calculated with the addition of 0.984 Da in the M value of the unmodified peptide. Please notice that in some cases, the unmodified form is already once or double deamidated.

N,Q Position	Peptide	Z	PL	H	Calculated M (Da)	Observed m/z
199Q	R.GLPGPPGAPGPQGFQGGPEGEPGASGPM GPR.G + Oxidation(M) + 5 Hydroxylation(P)	3	34	27.51	3099.385	1034.128
202 Q						
295 N	K.GDAGPAGPKGEPGSPGENGAPQMGR.G + Oxidation(M) + 4 Hydroxylation(P)	3	28	14.66	2524.074	842.358
300 Q						
324 N	R.GNDGATGAAGPPGPTGPAGPPGFPGAVGAK .G) + 3 Hydroxylation (P)	2	31	24.05	2547.185	1274.599
358 Q	R.GNDGATGAAGPPGPTGPAGPPGFPGAVGAK GEAGPQGPR.G + 4 Hydroxylation(P)	3	40	22.87	3412.589	1138.537
367 Q	R.GSEGPQGVR.G	2	10	8.09	885.416	443.713
	R.GSEGPQGVRGEPGPPGAGAAGPAGNPGA DGQPGAK.G + Deamidation (N,Q) + 4 Hydroxylation(P)	3	37	16.53	3183.431	1062.14302
387 N	R.GEPGPPGAGAAGPAGNPGADGQPGAK.G + 4 Hydroxylation(P)	2	28	17.69	2315.027	1158.527
393N	R.GEPGPPGAGAAGPAGNPGADGQPGAK.G + Deamidation (N,Q) + 4 Hydroxylation(P)	2	28	17.69	2316.031	1159.023
400 N	K.GANGAPGIAGAPGFPGAR.G + 3 Hydroxylation(P)	3	19	23.93	1584.751	529.261

421 Q	R.GPSGPQQGPPGPPGPK.G + Hydroxylation(P)	2	16	12.19	1301.622	651.821
457 Q	K.GEPGPVGVQGGPPGAGEEGK.R + 2Hydroxylation(P)	3	21	21.04	2002.957	668.665
561Q	K.TGPPGPAGQDGRPGPPGPPGAR.G + 4Hydroxylation(P)	3	23	14.39	2055.958	686.332
576 Q	R.GQAGVMGFPGPK.G + Hydroxylation(P)	2	13	22.11	1160.551	581.282
619 Q	K.DGEAGAQQPPGPAGPAGER.G + Hydroxylation(P)	2	20	15.14	1705.752	853.89
688 Q	R.GVQGGPPGAGPR.G +Deamidation(N,Q) + Hydroxylation(P)	2	13	11.52	1104.554	553.288
700 N	R.GANGAPGNDGAKGDAGAPGAPGSQGAPGL QGMPGER.G + Oxidation(M) + 5 Hydroxylation(P)	3	37	18.5	3239.399	1080.814
705 N	R.GANGAPGNDGAKGDAGAPGAPGSQGAPGL QGMPGER.G +Deamidation (N,Q)+ Oxidation(M) + 5 Hydroxylation(P)	3	37	18.5	3240.383	1081.142
721 Q	R.GANGAPGNDGAKGDAGAPGAPGSQGAPGL QGMPGER.G +2Deamidation (N,Q)+ Oxidation(M) + 5 Hydroxylation(P)	3	37	18.5	3241.367	1081.47
727 Q	K.GDAGAPGAPGSQGAPGLQGMPGER.G + Oxidation(M) + 4 Hydroxylation(P)	2	25	21.74	2214.945	1108.487
822 Q	R.GAPGDRGEPGPPGAGFAGPPGADGQPGAK .G + 4 Hydroxylation(P)	3	31	20.5	2702.218	901.744
855N	K.GDAGPPGPAGPAGPPGPIGNVGPAGAK.G + 3 Hydroxylation(P)	2	28	22.75	2280.11	1141.065
891N	R.VGPPGPSGNAGPPGPPGAGK.E + 3 Hydroxylation(P)	2	22	17.33	1811.866	906.944
952 Q	K.GSPGADGPAGAPGTPGPQGIAGQR.G + 2 Hydroxylation(P)	2	25	18.69	2103.979	1053.002
957 Q	K.GSPGADGPAGAPGTPGPQGIAGQR.G +Deamidation (N,Q) + Hydroxylation(P)	2	25	18.69	2088.968	1045.502
966 Q	R.GVVGLPGQR.G + Hydroxylation(P)	2	10	17.58	897.489	449.755
1117 Q	R.GFSLQGGPPGPPGSPGEGQPSGASGPAGPR.	3	31	23.34	2688.239	897.082

	G + 2Hydroxylation(P)					
1129 Q	R.GFSGLQGPPGPPGSPGEGQPSGASGPAGPR. G+Deamidation(N,Q) + Hydroxylation(P)	3	31	23.34	2673.268	892.092
1156 N	K.DGLNGLPGPIGPPGPR.G + 3 Hydroxylation(P)	3	17	30.44	1560.776	781.399

Table S14: Selection of the most frequently detected Asn and Gln containing peptides in COL1A2 (P08123) from the 33 samples of Pompeii, Herculaneum, Baia Scalandrone, Oplontis and San Paolo Bel Sito. The most frequently observed masses are reported. Z: Charge ; PL: Peptide Length ; H: Hydrophobicity. The mass of the corresponding deamidated form can be calculated with the addition of 0.984 Da in the M value of the unmodified peptide. Please notice that in some cases, the unmodified form is already deamidated.

N,Q Position	Peptide	Z	PL	H	Calculated M (Da)	Observed m/z
272 N	K.GEIGAVGNAGPAGPAGPR.G	2	19	20.83	1546.771	774.3924
299 N	R.GEVGLPGLSGPVGPPGNPGAN GLTGAK.G + 3 Hydroxylation(P)	2	28	31.09	2433.2	1217.618
303 N	+Hydroxylation(K)					
384 N	R.GPNGEAGSAGPPGPPGLR.G + 2 Hydroxylation(P)	2	20	20.99	1618.756	810.3844
458 N	R.GLPGSPGNIGPAGK.E + 2 Hydroxylation(P)	2	15	17.68	1252.627	627.3252

491 N	R.GEPGNIGFPGPK.G + 2 Hydroxylation(P)	2	13	22.45	1200.563	601.2912
620 N	R.GPSGPPGPDGNKGEPGVVGAV GTAGPSGPSGLPGER.G + 3 Hydroxylation(P)	3	37	25.89	3212.519	1607.267
723 N	R.GEVGPAGPNGFAGPAGAAGQP GAK.G+ Hydroxylation(P)	2	25	22.89	2049.973	1025.991
734 Q	R.GEVGPAGPNGFAGPAGAAGQP GAK.G+ Deamidation(N,Q) + Hydroxylation(P)	2	25	22.89	2050.957	1026.477
750 N	K.GENGVVGPTGPVGAAGPAGPN GPPGPAGSR.G + Hydroxylation(P)	2	31	22.43	2566.239	1284.125
768 N	K.GENGVVGPTGPVGAAGPAGPN GPPGPAGSR.G + Deamidation (N,Q)+ Hydroxylation(P)	2	31	22.43	2567.222	1284.615

864 Q	K.GPSGEAGTAGPPGTPGPQGLL APGILGLPGSR.G + 4 Hydroxylation(P)	3	50	36.77	2596.475	986.4588
918 N	R.GPPGAVGSPGVNGAPGEAGR. D + 3 Hydroxylation(P)	2	21	18.91	1750.81	876.4174

Table S15: Selection of methionine containing peptides in COL1A1(P02452) and COL1A2 (P08123) from the 33 samples of Pompeii, Herculaneum, Baia Scalandrone, Oplontis and San Paolo Bel Sito. The most frequently observed masses are reported. Z: Charge ; PL: Peptide Length ; H: Hydrophobicity. The mass of the corresponding deamidated form can be calculated with the addition of 15.998 Da in the M value of the unmodified peptide. Please notice tha in some cases, the unmodified form is already deamidated.

Collagen chain	M Position	Peptide	Z	PL	H	Calculated M (Da)	Observed m/z
COL1A1	217 M	R.GLPGPPGAPGPQGFQ GPPGEPGEPGASGPMG PR.G + 5 Hydroxylation(P)	3	34	27.51	3084.385	1029.123
(P02452)	301 M	K.GDAGPAGPKGEPGSP GENGAPGQMGPR.G + Deamidation(N,Q) + 4 Hydroxylation(P)	3	28	14.66	2509.074	837.352
	580 M	R.GQAGVMGFPGPK.G + Deamidation(N,Q) + Hydroxylation(P)	2	13	22.11	1161.546	581.7713

	729 M	R.GANGAPGNDGAKGD AGAPGAPGSQGAPGLQ GMPGER.G + Deamidation(N,Q) + 5 Hydroxylation(P)	3	37	18.5	3224.399	1075.809
	1000 M	R.GPPGPMGPPGLAGPP GESGR.G + 2 Hydroxylation(P)	2	21	24.91	1815.854	908.9252
COL1A2 (P08123)	417 M	R.AGVMGPPGSR.G + Hydroxylation(P)	2	11	13.21	943.451	472.7182
	447 M	R.GPNGDAGRPGEPGL MGPR.G + Hydroxylation(P)	3	19	18.87	1766.798	884.393
	785 M	R.GDGGPPGMTGFPGA AGR.G + 2 Hydroxylation(P)	2	18	21.7	1532.665	767.3355

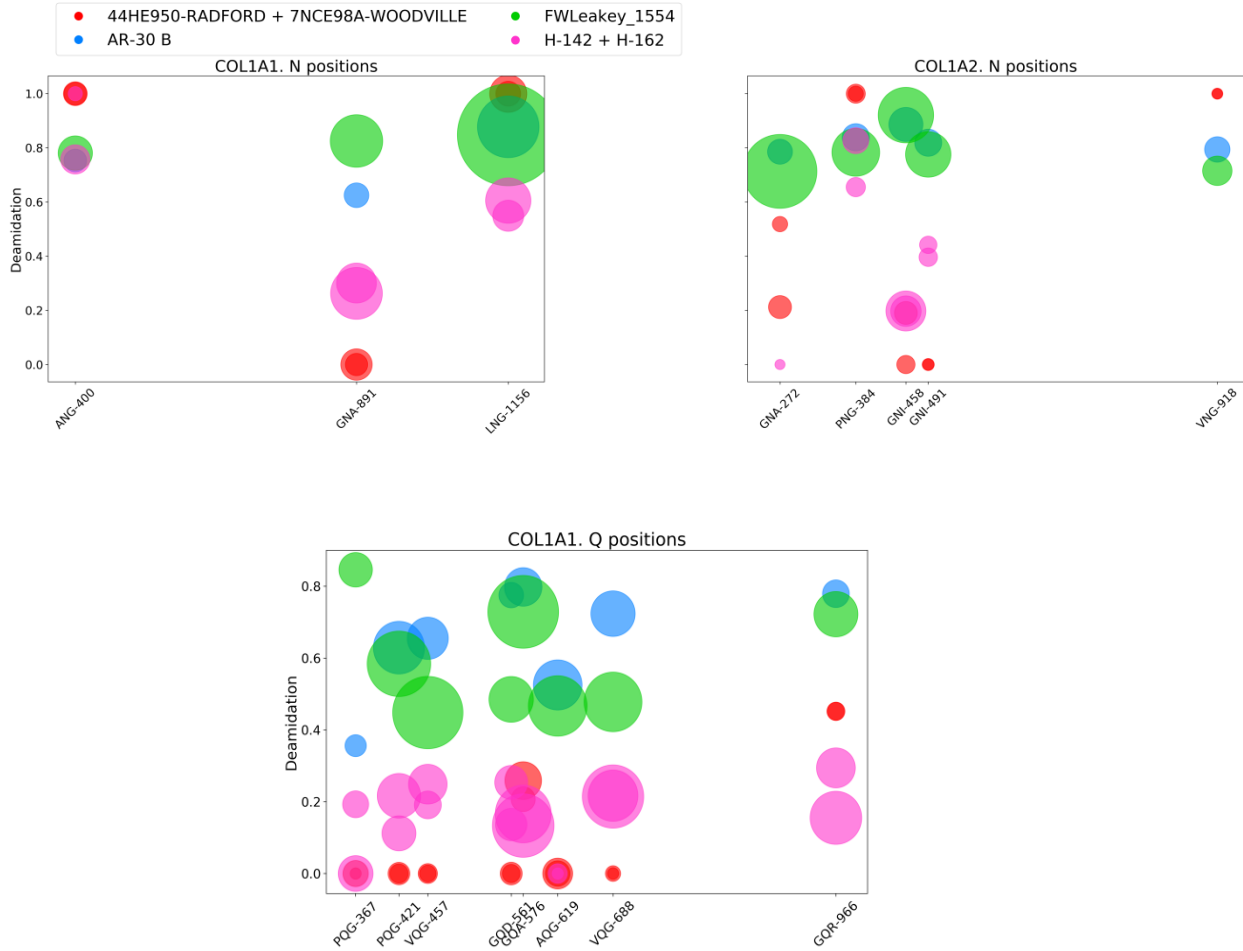


Figure S4: 3D visualization of deamidation (N,Q) along the sequence of collagen alpha-1 (I) and collagen alpha-2 (I) chains (upper panel) and, for greater clarity, zoomed in the zones of the selected peptides (lower panel), in test archaeological bones; reference numbers: H-162, H-142 from [1], 44HE950-RADFORD, 7NCE98A-WOODVILLE from [2] AR-30A from [3], and FWLeahey_1554 from [4]. The values were calculated was obtained only for the selected peptides of the table S16. The label size indicates the relative intensity of each position in the single sample.

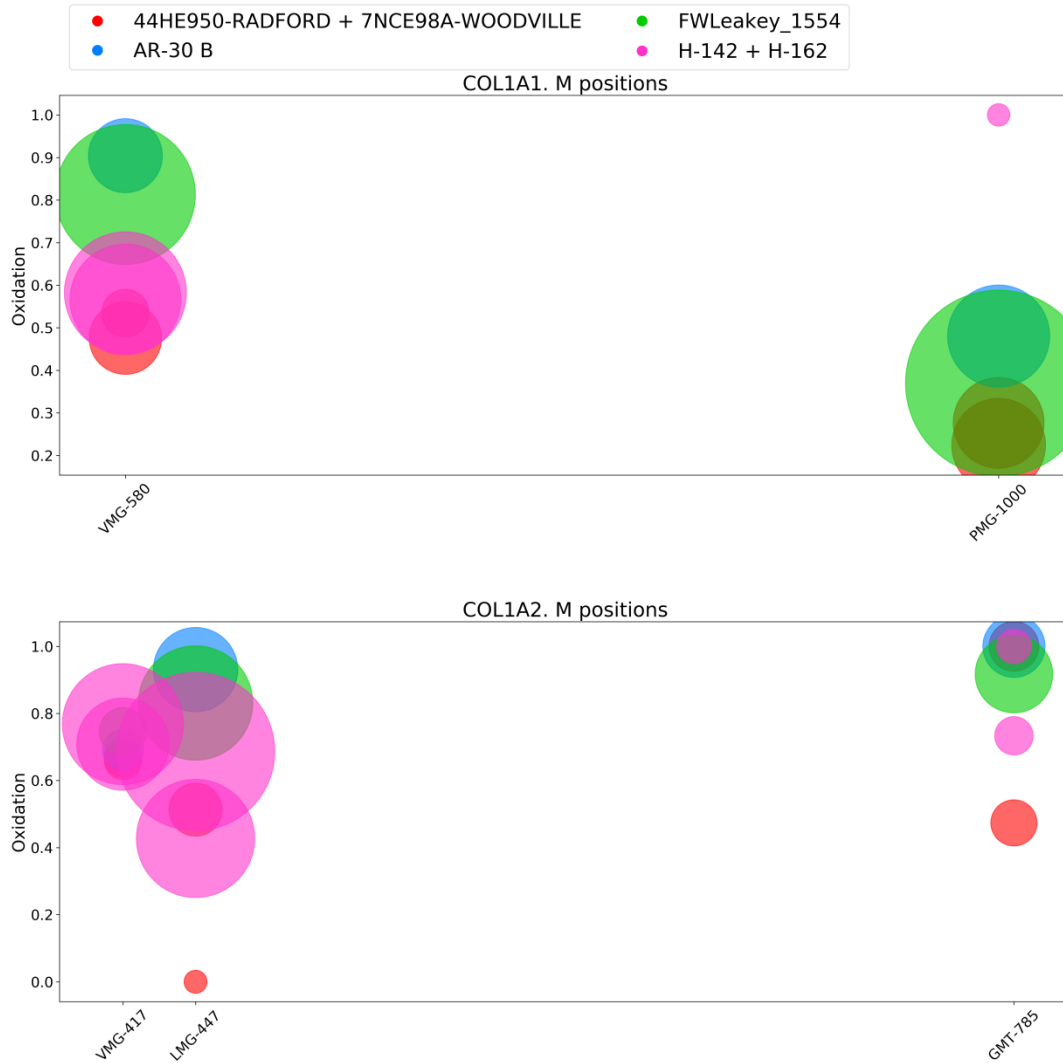


Figure S5: 3D visualization of oxidation (M) along the sequence of collagen alpha-1 (I) and collagen alpha-2 (I) chains (upper panel) and, for greater clarity, zoomed in the zones of the selected peptides (lower panel), in test archaeological bones; reference numbers: H-162, H-142 from [1], 44HE950-RADFORD, 7NCE98A-WOODVILLE from [2], AR-30A from [3], and FWLeakey_1554 from [4]. The values were calculated only for the selected peptides of the table S16. The label size indicates the relative intensity of each position in the single sample.

References

- [1] R. Sawafuji, E. Cappellini, T. Nagaoka, A.K. Fotakis, R. Rakownikow Jersie-Christensen, J. V Olsen, K. Hirata, S. Ueda, Proteomic profiling of archaeological human bone, (n.d.). <https://doi.org/10.1098/rsos.161004>.
- [2] T.P. Cleland, Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics, *J. Proteome Res.* 17 (2018) 3976–3983. <https://doi.org/10.1021/acs.jproteome.8b00637>.
- [3] F. Welker, M. Hajdinjak, S. Talamo, K. Jaouen, M. Dannemann, F. David, M. Julien, M. Meyer, J. Kelso, I. Barnes, S. Brace, P. Kamminga, R. Fischer, B.M. Kessler, J.R. Stewart, S. Pääbo, M.J. Collins, J.J. Hublin, Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 11162–11167. <https://doi.org/10.1073/pnas.1605834113>.
- [4] L.T. Lanigan, M. Mackie, S. Feine, J.J. Hublin, R.W. Schmitz, A. Wilcke, M.J. Collins, E. Cappellini, J. V. Olsen, A.J. Taurozzi, F. Welker, Multi-protease analysis of Pleistocene bone proteomes, *J. Proteomics.* 228 (2020). <https://doi.org/10.1016/j.jprot.2020.103889>.

3.4 Manuscript 3: Molecular characterization of animal glues for restoration purposes

Georgia Ntasi, Sara Sbriglia, Rosanna Pitocchi ,Chiara Melchiorre, Laura dello Iorio, Giancarlo Fatigati, Ilaria Bonaduce, Andrea Carpentieri, Gennaro Marino and Leila Birolo. Manuscript in preparation

Abstract

Animal glues are widely used in restoration, serving as adhesives, binders, and consolidants for organic and inorganic materials [1,2]. Their performance is intrinsically linked to the adhesive properties of the original collagen source, which determines the glue's chemical, physical, and mechanical properties. Collagen modifications, in terms of chemical modifications including fragmentation is strongly affected by the actual different recipes for the extraction and preparation from the original animal tissue. In order to molecularly characterize a set of artificial and commercial glues, a wide-broad protocol for the extraction of proteins, lipids, and sugars was used. GC-MS analysis demonstrated the co presence of natural and/or synthetic adhesives that might be involved in the glue performances. A shotgun proteomic approach provided the animal origin, even when in mixture, discriminating also between hide and strong glue on the basis of the presence of collagen type III, which is absent in bone and present in connective skin/hide tissues. Proteomic data were used to investigate chemical modifications, ranging from deamidation to methionine oxidation and backbone cleavage as main signatures of collagen damage. SDS-PAGE analysis revealed a different behavior between hide and strong glues in terms of acid-soluble collagen degradation products, whereas EGA-MS analysis is now in progress for a further molecular characterization, associated also with the glues' different physical properties.

Introduction

The first evidence of the use of natural substances as adhesive dates back to 4000 B.C. as archaeologists found in a prehistoric burial site a pottery vessel that had been repaired with a sticky resin from tree sap. A simple procedure for making animal glue was reported in 2000 B.C, while the first commercial glue factory was started in the 1700s in Holland that produced animal glue from hides, and the first patent was issued for a fish glue in Britain about 50 years after. Thereafter, there was a flourishing of patents for recipes of glues made from animal bones, fish, starch, and milk proteins[3].

The term animal glue usually refers to adhesive prepared from mammalian connective tissues, namely bones, skin/hide or sinew: upon treatment with acids, alkalis or hot water, the otherwise insoluble collagen, the main protein constituent of all these tissues, becomes soluble. Hide glues are from tannery waste, while

bone glues, of course, are made from animal bones. Fish glues, that are the forerunners of all household glues, are by-products of fish skin, and are also made of collagen. Collagen-based animal glues have been widely used in artworks, as adhesives, binding media for pigments and consolidants for organic and inorganic materials.

Collagen in its natural state is a triple helix protein, with the peculiar Gly-X-Y repetitive sequence and a uniquely high content of Pro and Hyp, that make it easily recognizable in the protein universe. It is naturally water insoluble and it has to be transformed into soluble gelatin to be used as adhesive, binder or consolidants for organic and inorganic materials.

Performances depend upon the original collagen source, but are also strongly affected by the extraction and preparation procedures, since both the specific sequence and the processing of the starting material influence the resistance of the polipeptidic chains to hydrolysis and the partial denaturation of the triple helix that occurs upon cooling in the gelatination process [3].

These factors have a significant influence on the properties of the glue, in terms of glue viscosity, strength, and overall mechanical behavior [4]. Conservators' choice of the commercial glue to be used mainly relies on their experience and often lacks a scientific characterization that could help them in drawing a suitably informed decision.

In this respect, this paper is aimed to provide a diagnostic tool that can be used to molecularly characterize commercial animal glues from different sources. Proteomics is gaining now a popular momentum in the field of diagnostic tools for cultural heritage and apart from being on the side of historians in the characterization of materials and conservation state of work of arts and archaeological remains, it can also be useful for the operators in the restoration campaigns in the selections of the most appropriate/best materials to be used. Because of mass spectrometry high accuracy and sensibility, proteomics has already been widely used for collagen analysis to discriminate species from bone fragments from different species [5–7], to identify proteinaceous binders in paintings [8,9] and gildings [10]. A multi-technique approach, with the combination of mass spectrometric techniques (GC-MS, LC-MS/MS and EGA-MS) and denaturing electrophoresis was applied to the animal glues, in order to provide a broad range molecular characterization. In this way we were able to give a complete picture in terms of proteins and other organic molecules (sugars and lipids) but also to discover different collagen chemical modification profiles between hide and strong glues.

Results

Protein Identification

A shotgun proteomics approach using LC-MS/MS was applied to 19 animal glues, provided either from the restoration laboratory of University Suor Orsola Benincasa in Naples or from the restoration laboratory of the Museo del Prado in Madrid (table S1). It has to be noticed that animal glues is an intriguing kind of sample and their protein characterization is quite challenging. It is in fact quite easy to identify the presence of collagen, because of its quite peculiar and unique sequence in the panorama of protein, but identify the actual collagen origin and its chemical modifications are the real challenges. This originates from the extremely high sequence similarity among the species. As a result, in some cases, collagen discrimination between two animal species (i.e., bovine and porcine collagen, for instance) is actually based only on a very few unique peptides. Many bioinformatic tools characterize collagen materials using peptides that are unique for the protein family (i.e., COL1A1 of mammals) but common for the animal species of the same family. In the case of animal glues, this could be quite tricky, since most of the animals commonly used to produce them belong to the same family (of course fish glues are an exception). Since we mainly focused our attention on hide and strong glues, and genome information for characterization of fish collagen are still poor, collagen identification was performed, taking into consideration only the unique peptides that differentiate the species, although, in the total number of identified peptides, we included both unique and common peptides. It is possible to discriminate between hide and bone glues on the basis of the presence of collagen type III, which is absent in bone made glue.

The results of protein identifications are shown in table 1, where the samples were divided into strong glues, when only collagen of type I was detected, and into hide glues when collagen of type III was identified. As a result, the majority of the samples (13 out of 19) are characterized as hide glues and four were identified as strong glues. We should remind that collagen extraction from the animal bones (strong glues) requires a more vigorous extraction procedure than the preparation from animal skin (hide glues).

Although the label on nine glues reported the term “rabbit glue”, we identified the presence of collagen from *Oryctolagus cuniculus* only in four samples, and actually only the sample labeled as *Rabbit glue (of the first postwar)* appears to be the only pure rabbit glue, most likely because is the only purely artificial glue among the samples. All the other rabbit glues contain some bovine and sheep collagen, and the *rabbit glue 7* contains also some porcine and fish glue. All the rabbit glues are hide glues since collagen of type III was

identified in all samples, as expected, since the original recipe wants the hide glue being prepared by prolonged boiling animal skin.

Rather surprisingly, among the samples labeled as fish glues, only in *sturgeon glue* we identified some fish collagen. but it's not attributed to the Acipenseridae family but instead to Scyliorhinidae family as collagen type I of *Scyliorhinus canicula* has been detected.

The samples tagged as strong glues were consistent with their labels as no collagen of type III was detected. In particular, all of them are mixtures of bovine with porcine collagen although in some cases donkey collagen has been identified too. All the protein identification details for the single samples are reported in the supplementary material (S2-S20).

Table 1: Collagen chains that have been identified in the animal glue samples. MS/MS spectra were searched by Mascot MS/MS Ion search using COLLE database, considering deamidation on Gln and Asn, oxidation on Met, pyro- Glu formation at Gln at the N-terminus of peptides as variable modifications. Details of identifications are reported in the supplementary materials.

Name of the sample	Taxonomy	Collagen a1(I)	Collagen a2(I)	Collagen a1(III)
Rabbit glue (postwar)	<i>Oryctolagus cuniculus</i>	*	*	*
Rabbit Glue Totten (60s)	<i>Bos taurus</i>	*	*	*
Rabbit glue (CTS)	<i>Bos taurus</i>	*	*	*
Rabbit glue (Pearls)	<i>Bos taurus</i>	*	*	*
Rabbit glue (Giosi)	<i>Bos taurus</i>	*	*	*
	<i>Sus scrofa</i>	*	*	*
Rabbit glue 10	<i>Bos taurus</i>	*	*	*
Rabbit glue 2	<i>Oryctolagus cuniculus</i>	*	*	
	<i>Ovis aries</i>		*	*
	<i>Capra hircus</i>	*		
Rabbit glue 3	<i>Bos taurus</i>	*	*	*
	<i>Oryctolagus cuniculus</i>	*	*	*
	<i>Ovis aries</i>		*	
Rabbit glue 7	<i>Oryctolagus cuniculus</i>	*		
	<i>Sus scrofa</i>	*	*	*
Rabbit glue 6	<i>Bos taurus</i>	*	*	*
	<i>Oryctolagus cuniculus</i>	*	*	
Fish glue (flakes)	<i>Sus scrofa</i>	*	*	*
Fish glue 4	<i>Bos taurus</i>	*	*	*
Fish glue 5	<i>Bos taurus</i>	*	*	*

Sturgeon fish glue (pearls)	<i>Scyliorhinus canicula</i>	*	*	
	<i>Sus scrofa</i>	*	*	
Strong glue (pearls)	<i>Bos taurus</i>	*	*	
	<i>Sus scrofa</i>	*	*	
Strong glue (old manufacturing)	<i>Bos taurus</i>	*	*	
	<i>Sus scrofa</i>	*	*	
	<i>Equus asinus</i>	*	*	
Strong glue 8	<i>Bos taurus</i>	*	*	
	<i>Sus scrofa</i>	*	*	
	<i>Equus asinus</i>		*	
Strong glue 9	<i>Bos taurus</i>	*	*	
	<i>Sus scrofa</i>	*		
	<i>Ovis aries</i>	*	*	
	<i>Equus asinus</i>	*	*	
Strong glue 1	<i>Bos taurus</i>	*	*	
	<i>Sus scrofa</i>		*	
	<i>Equus asinus</i>	*	*	

Molecular characterization of organic molecules as additives.

Organic, natural, or synthetic compounds have been used as additives in the preparation of animal glues for several purposes, as reported in literature [11]. For instance, oils and resins have been used to fix the mechanical properties of the animal glues, whereas carbohydrates are usually used to increase gel strength and viscosity. The GC-MS analysis was carried out for a more in-depth characterization of other organic component in the glues under examination.

The wide range of samples from different sources available as dataset allowed to evaluate on a significant set number of independent glues the substantial variability of different recipes that can be used in the preparation of animal glues. A multistep protocol was applied, and sugars and lipids were identified. Lipids were identified as fatty acid methyl esters (FAMES) (tables S22 and S24) derivatives, whereas monosaccharides (tables S23 and S25) as trimethylsilyl derivatives.

Several steroid compounds have been detected in all samples, indicating their animal origin. Also, compounds of abietanic skeleton were detected, demonstrating the use of resins of *Pinaceae* family in the glue preparation (Table S20). More specifically, in the *Rabbit (Postwar)* glue that is the only handmade animal glue, a derivative of valeric acid was detected, a major component of *Valeriana officinalis* [12], thus indicating the addition of natural plants (Table S21). The presence of natural sources is also confirmed from the detection of heterocyclic compounds such as *1H-Indene*, *2-butyl-5-hexyloctahydro-* present in natural sources [13–18]. Interestingly, in most of the glues provided by the University of Suor Orsola, adipic acid was identified, a compound that is found in beets and sugar cane. On the other hand, in case of the animal glues

provided by the museum in Madrid, the identification of synthetic chemical compounds such as eurucamide, demonstrated a more recent and industrial animal glue preparation (Table S21). In the same samples, stevioside was detected, a glycoside that is a major component of stevia plant [19].

Also, several monosaccharides such as glucose mannose, arabinose, were detected in the most of the cases (Tables S23 and S25). Moreover, palmitic and stearic acid were commonly identified in all the samples (Tables S22 and S24), whereas the co-presence of fatty acids such as oleic and margaric acid with sugars such as galactose, might suggest the presence of some milk derivatives or confirm the presence of bovine milk such as in case of the animal glue “old manufacturing” where casein of *Bos taurus* was identified (table S.8) . However, most of the identified sugars and lipids are so common and can be found in several materials of both animal and vegetable origin, that their identification does not help much the molecular characterization and discrimination of the glues, at this stage. Some more information would be gained from a quantitative analysis, currently in progress.

Diagenetically induced modifications in animal glues

Modifications of amino acids, such as oxidation of Met, deamidation of Gln and Asn, as well as backbone cleavage of the polipeptidic chains, are degradation phenomena commonly observed and routinely searched for in aged proteins [20, 21]. We evaluated deamidation occurrence in these set of samples keeping in mind the suggestion by [22] that deamidation should be viewed as a global indicator of sample preservational quality, since rates and levels of deamidation are affected by several chemical and environmental factors. The harsher conditions of collagen extraction in the preparation of strong glues in comparison to hide glues might have been imprinted also in the level of collagen deamidation.

For the evaluation of deamidation we mainly focused our attention on glutamine (Gln) deamidation since the reaction in case of asparagine (Asn) is faster and easily provoked during the sample manipulation, thus losing its informational utility. For the calculation of deamidation raw LC-MSMS data were analyzed by MaxQuant software with an in-house script based on peptide intensities for semi quantitative evaluation, as reported in [20].

The deamidation levels of strong and hide glues are reported in figure S1. Deamidation of hide glues is extremely variable, ranging between 2-98%. On average, we nevertheless can say that strong glues are less deamidated, with values in the range between 2 and 12% glues.

Also, quite interestingly, we observed that the most deamidated among the hide glues (Fish 4, Rabbit (Pearls), Rabbit (CTS), Rabbit (Totin) are those that contain only bovine collagen, that from now and on will

be called “pure animal glues”, to discriminate them from the mixed-collagen hide glues, that contain collagen from more than one species. To further investigate this aspect, we evaluated the deamidation level in the glues that contain a single type of collagen. Figure S2 presents the deamidation levels of the bovine sequences in the pure bovine glues, whereas Fig S3 the levels of pure rabbit and porcine glues, respectively. Figure S2 demonstrates that all collagen chains of the pure bovine glues are extensively deamidated in comparison to the rabbit (Fig S3 A) and porcine (Fig S3 B) collagen chains. Therefore, it seems that deamidation might be influenced by the type of the contained animal glue. To further investigate it, we calculated the deamidation level of the bovine collagen sequences contained in the mixed glues (Figure S4). From the comparison of the values between the figure S1 and S4 it seems as the deamidation was mostly ascribable to the bovine sequences. This will require further investigation in the future. However, not all of the animal glues contain bovine collagen, so we can't use the elevated deamidation levels of *Bos taurus* for a comparative study among mixed and pure animal glues. For this reason a different approach was adopted.

As a further investigation, we manually combined the peptides on the basis of the protein family and not of the animal species, and deamidation was calculated again. Fig S5-S7 report the deamidation of the three more abundant protein families: collagen-type I chain 1, collagen-type I chain 2 and collagen-type III chain 1. From the comparison of the S2-A and S5 we can observe: collagen-type I chain 1 of the mixed glues are less deamidated in comparison to the peptides of the pure bovine glues. This observation is confirmed also from the comparison of the other protein families of mixed glues with the same protein families of the pure glues: fig S6 versus fig S2B and fig S7 versus fig S2 C. Mixed-collagen glues are eventually much less deamidated than pure glues.

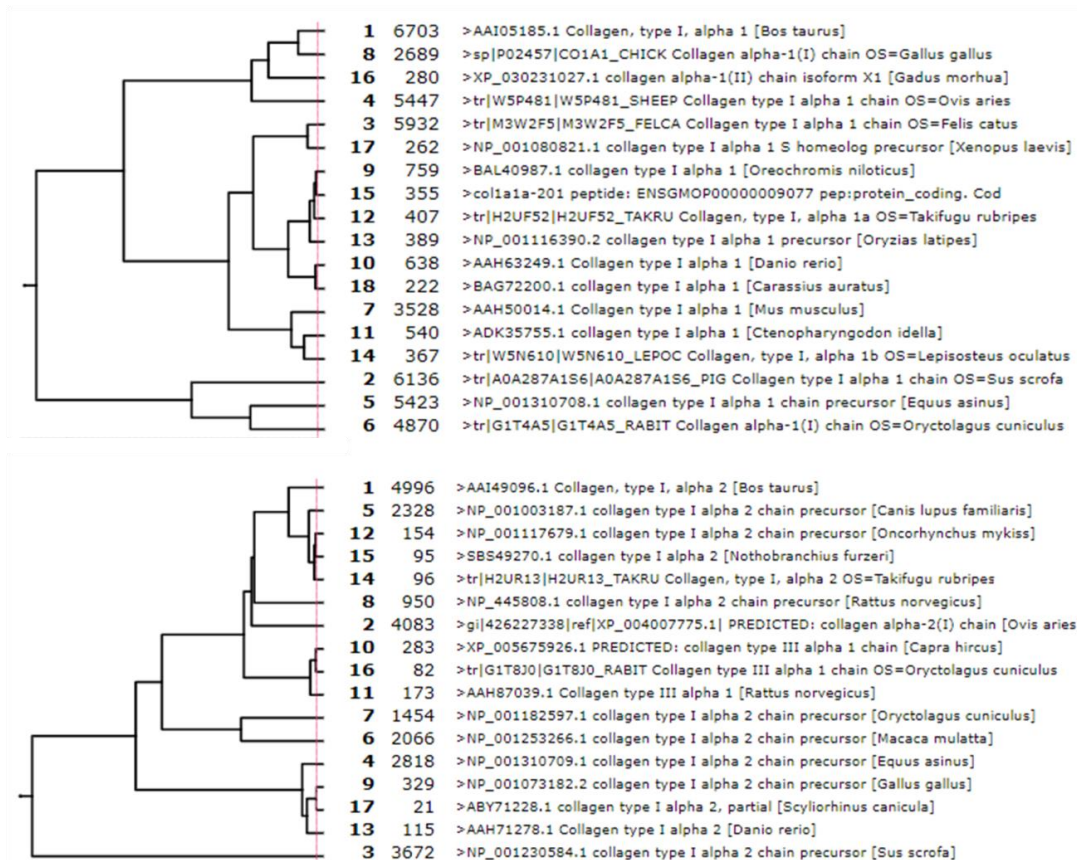


Figure 1: Diagrammatic illustration of the two biggest collagen protein families in case of animal glues.

Structural evaluation of collagen deamidation.

A positional evaluation of deamidation (NQ) was performed to investigate on any possible three-dimensional effect on deamidation of collagen in the glue. As first approach, given the high diversity in the composition of the animal glues we focused our attention to the animal glues that contain only *Bos taurus* collagen (pure bovine animal glues).

The evaluation of deamidation level was provided “positionally”, by checking manually the fragmentation spectra of the peptides containing Asn and Gln, as identified by MASCOT software. The positions of Asn and Gln that were detected as unmodified were characterized as X, the positions that were found only deamidated as D, the position that were both as deamidated and unmodified as DX and, finally, the positions that were not detected at all, nor unmodified nor deamidated, as “NF”. This simple approach allows to gain a better vision of collagen degradation in the principal chains of collagen $\alpha 1(I)$ and collagen $\alpha 2(I)$ (Fig1).

A similar approach was applied also for the evaluation of oxidation level of (M). In this case, the oxidized positions were characterized as OX and the non-oxidized that have been detected as unmodified, as X.

Finally, both the evaluation of methionine and deamidation were combined in a unique table that presents a picture of the collagen damage in an amino acid order.

As tables 2 and 3 demonstrates, that are some regions that are identified mainly as modified, which are followed by regions that are not detected. It was quite interesting that differently from deamidation, almost all the detected peptides containing methionine were found as oxidized. Moreover, the methionine positions that were not detected, seem to belong to regions where asparagines and glutamines were not detected to, suggesting a common structural issue. For instance, in the collagen I chain $\alpha 1$ the positions 300 M, 579 M, 728M and 999M were detected always oxidized, although the other positions belong to regions at the beginning or the end of the collagen chains. On the other hand, in the collagen type I chain 2 the position 445 M was detected always as oxidized, while all the other positions belong to regions that were not covered at all. From the positional evaluation of deamidation (NQ), it seems that a very similar percent of deamidation, as described in the previous paragraph, in the collagen chains of *Bos taurus* is attributed to same collagen zones in the principal structural chain. Regardless the sample preparation of the single animal glue, the positional evaluation of collagen degradation shows a similarity for this type of collagen. Some further insights might come from considering in more details the structural effect, namely by searching for any structural feature within the chain that might influence the deamidation level at specific position within the sequence.

Table 2: Positional evaluation of deamidation (N,Q) and oxidation (M) along Collagen $\alpha 1(I)$ sequence of *Bos taurus* in all the pure bovine animal glues. NF: Not detected; X:Detected; D/O:Detected as modified; DX/OX : Detected both as modified and unmodified.

AA Position	Rabbit(CTS)	Rabbit(Pearls)	Rabbit(Totten)	Fish 4	Fish 5	Rabbit 10
162 Q	NF	NF	NF	NF	NF	NF
180 M	NF	NF	NF	NF	NF	NF
198 Q	X	X	DX	DX	DX	DX
201 Q	D	D	DX	DX	DX	DX
216 M	NF	NF	NF	NF	NF	NF
228 N	D	D	D	NF	NF	NF
249 Q	NF	NF	NF	NF	NF	NF
263 M	NF	NF	NF	NF	NF	NF
294 N	DX	D	D	DX	DX	NF
299 Q	DX	D	DX	DX	DX	D

300 M	OX	OX	O	O	O	O
323 N	DX	D	DX	DX	DX	DX
357 Q	DX	D	DX	DX	D	D
366 Q	D	D	D	DX	DX	D
386 N	DX	D	DX	DX	DX	DX
392 Q	DX	D	DX	DX	DX	DX
399 N	DX	D	DX	DX	DX	DX
420 Q	DX	D	DX	DX	DX	DX
431 N	D	D	D	NF	NF	NF
456 Q	DX	DX	DX	DX	DX	DX
560 Q	DX	D	DX	D	DX	DX
575 Q	DX	D	DX	D	DX	DX
579 M	OX	O	O	O	OX	O
618 Q	DX	D	DX	DX	DX	DX
633 Q	NF	NF	NF	NF	DX	NF
642 Q	NF	NF	NF	NF	DX	NF
660 Q	NF	NF	NF	NF	DX	NF
687 Q	DX	DX	DX	DX	DX	DX
699 N	DX	NF	DX	NF	NF	NF
704 N	DX	DX	DX	NF	NF	NF
720 Q	DX	D	DX	NF	DX	DX
726 Q	DX	X	DX	NF	DX	DX
728 M	OX	O	O	O	O	O
821 Q	DX	D	DX	DX	DX	DX
854 N	DX	D	DX	D	DX	DX
890 N	DX	D	DX	D	DX	DX
951 Q	DX	DX	DX	DX	DX	DX
956 Q	DX	D	DX	DX	DX	DX
965 Q	DX	D	DX	DX	DX	DX
984 Q	NF	NF	NF	NF	NF	NF
999 M	OX	OX	OX	O	OX	O
1089 Q	NF	NF	NF	NF	NF	NF
1101 Q	NF	NF	NF	NF	NF	NF
1116 Q	DX	DX	DX	DX	DX	DX

1128 Q	DX	DX	DX	DX	DX	DX
1155 N	DX	DX	DX	DX	DX	DX
1202 Q	NF	NF	NF	NF	NF	NF
1205 Q	NF	NF	NF	NF	NF	NF
1263 M	NF	NF	NF	NF	NF	NF
1292 M	NF	NF	NF	NF	NF	NF
1329 M	NF	NF	NF	NF	NF	NF
1357 M	NF	NF	NF	NF	NF	NF
1376 M	NF	NF	NF	NF	NF	NF

Table 3: Positional evaluation of deamidation (N,Q) and oxidation (M) along Collagen $\alpha 2(I)$ sequence of *Bos taurus* in all the pure bovine animal glues. NF: Not detected; X:Detected; D/O:Detected as modified; DX/OX : Detected both as modified and unmodified.

AA Position	Rabbit(CTS)	Rabbit(Pearls)	Rabbit(Totten)	Fish 4	Fish 5	Rabbit 10
80 Q	NF	NF	NF	NF	NF	NF
91 M	NF	NF	NF	NF	NF	NF
94 M	NF	NF	NF	NF	NF	NF
109 Q	NF	NF	NF	DX	DX	DX
112 Q	NF	NF	X	DX	DX	DX
123 Q	NF	DX	DX	DX	DX	DX
160 Q	NF	DX	NF	D	DX	DX
181 N	NF	NF	NF	NF	NF	DX
189 Q	D	NF	DX	NF	NF	NF
205 N	D	DX	DX	NF	NF	NF
210 Q	NF	DX	DX	NF	NF	NF
270 N	DX	NF	X	D	D	NF
297 N	DX	DX	DX	DX	DX	DX
301 N	D	DX	DX	DX	DX	DX
357 N	NF	NF	NF	NF	DX	DX
366 Q	NF	NF	DX	NF	DX	DX
399 N	NF	NF	NF	NF	NF	NF
415 M	NF	NF	NF	O	X	O
433 N	DX	DX	DX	D	DX	DX
445 M	OX	OX	OX	O	X	OX

456 N	D	DX	DX	D	DX	DX
489 N	D	DX	X	DX	DX	DX
525 N	NF	NF	DX	DX	DX	DX
526 N	NF	NF	DX	DX	DX	DX
529 Q	NF	NF	DX	DX	DX	DX
535 Q	NF	DX	X	DX	DX	X
538 Q	NF	NF	DX	DX	DX	X
544 Q	D	NF	X	DX	DX	DX
553 Q	NF	NF	DX	D	DX	DX
618 N	D	DX	DX	D	DX	DX
688 N	D	DX	DX	D	DX	DX
721 N	D	DX	DX	D	D	DX
732 Q	DX	DX	DX	D	DX	DX
748 N	D	DX	DX	D	DX	DX
766 N	D	DX	DX	DX	DX	DX
823 Q	NF	DX	NF	NF	NF	NF
862 Q	NF	DX	DX	DX	DX	DX
909 N	D	DX	DX	DX	DX	DX
912 N	DX	DX	DX	DX	DX	DX
916 N	D	DX	DX	DX	D	DX
927 N	D	NF	DX	DX	NF	NF
930 N	NF	NF	NF	NF	NF	NF
939 Q	NF	NF	NF	NF	NF	NF
951 N	DX	NF	NF	NF	NF	DX
964 Q	DX	DX	DX	DX	DX	DX
975 N	NF	DX	DX	D	D	NF
1000 Q	NF	DX	NF	NF	DX	NF
1024 N	D	NF	DX	DX	DX	DX
1027 Q	D	DX	DX	DX	DX	DX
1039 Q	DX	NF	DX	DX	DX	DX
1068 Q	DX	DX	DX	DX	DX	DX
1081 Q	NF	NF	NF	NF	NF	NF
1084 Q	NF	NF	NF	NF	NF	NF
1186 M	NF	NF	NF	NF	NF	NF

1248 M	NF	NF	NF	NF	NF	NF
1255 M	NF	NF	NF	NF	NF	NF
1277 M	NF	NF	NF	NF	NF	NF

Backbone cleavage

Backbone cleavage of the polipeptidic chain is an expected degradation feature in ancient proteins [23, 24]. Damage at the backbone can be evaluated since, upon trypsin hydrolysis, semi-tryptic peptides will be generated, with trypsin cleavage site only at one end. The number of semitryptic peptides was semi-quantitatively evaluated by counting peptide to spectrum matches (PSMs) and dividing the PSMs of semitryptic peptides to the total PSMs of identified peptides, including both tryptic and semitryptic peptides. We evaluated the backbone cleavage occurrence in the single collagen chains in all strong and hide the samples. According to fig 2, strong glues contain a much higher number of semitryptic peptides in comparison to hide glues. Furthermore, more semitryptic peptides were found in mixed-collagen glues than in pure animal glues, suggesting that backbone cleavage took place in all polipeptidic chains of each glue. Therefore is related glue preparation procedures.

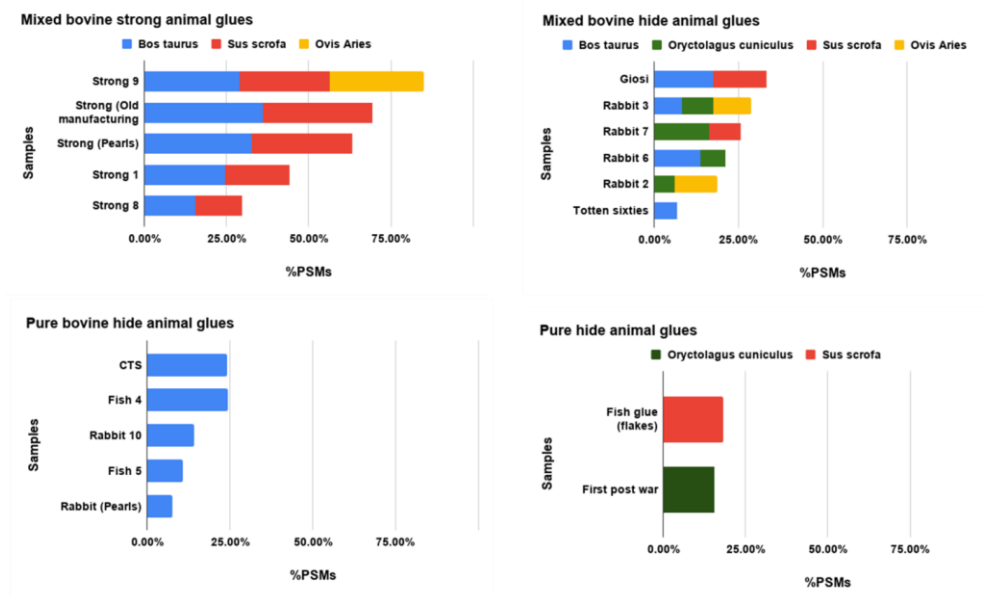


Figure 2: Grouping of the occurrence of backbone damage in the pure and mixed collagen animal glues. The occurrence of hydrolytic cleavages was evaluated by calculating the PSMs of semitryptic peptides normalized by the total number of PSMs for that chain (tryptic plus semitryptic).

Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS–PAGE)

Collagen is a quite challenging material for SDS gel, because of its gelling properties, its high molecular weight (collagen is formed by two alpha 1, and one alpha 2 chains, of a molecular weight ~100 kDa each), and because of cross links between the chains, impairing protein migration and separation [25,26]. We focused our attention to the search of relatively low molecular weight bands (below 100 kDa) that may indicate some collagen degradation.

Collagen extraction was performed as reported in *Hong et al 2017* [27] with slight modifications and acid soluble collagen (ASC) fractions were loaded onto a SDS gel. Figure 3 demonstrates the ASC different patterns among some pure and mixed-collagen hide and strong glues. In particular, mixed-collagen glues demonstrated several bands that might be considered as products of collagen degradation. These bands are not intense in case of pure hide and mixed-collagen strong glues, or in the rabbit pure glue that is dated to 60s and is artificial animal glue. Therefore, the results suggest that mixed-collagen hide glues are more susceptible to backbone cleavage, resulting in an extensive presence of high molecular weight fragments.

An SDS-gel on the heat soluble collagen (HSC) fraction, extracted within the same overall procedure, is needed, to further investigate the occurrence of macroscopic collagen degradation.

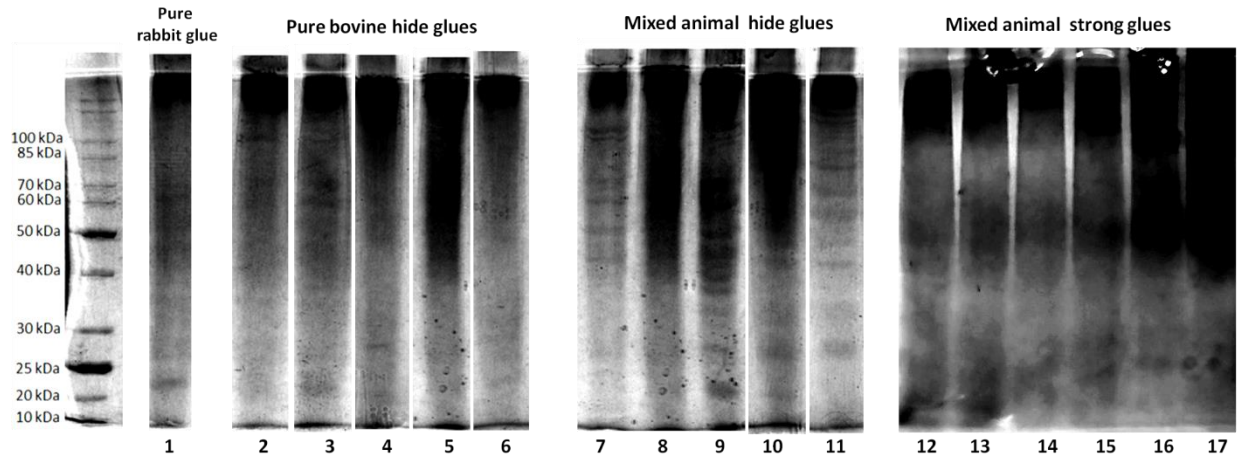


Figure 3: 7% SDS-PAGE showing ASC from different animal glues. **M.W:** molecular weight standards; **Pure rabbit glue:** 1: Rabbit (Totten 60s). **Pure bovine glues:** 2: Rabbit 10; 3: Rabbit (Postwar); 4: Fish 4; 5: Rabbit (Pearls) 6: Rabbit (CTS). **Mixed-collagen animal glues:** 7: Rabbit (Giosi); 8: Rabbit 3; 9: Fish (flakes); 10: Rabbit 6; 11: Rabbit 7. **Mixed animal strong animal glues:** 12: Strong 1; 13:Strong (Pearls); 14:Strong 8; 15: Strong 9; 16:Strong (Old manufacturing) 17: Sturgeon (Flakes). We shall remind that the labels do not always reflect the actual glue composition.

Conclusions

This work focused on the molecular characterization of commercial animal glues used for restoration purposes, provided by University of Suor Orsola Benincasa in Naples, and by the Museum del Prado in Madrid. A bottom-up approach was firstly applied for the identification of the collagen origin, verifying that in several cases the label did not correspond to the actual animal origin of the glue. The identification of collagen type III could be used to molecularly classify the glue between hide and strong glues. Collagen of type III origins from the use of skin waste and were used to enhance glue properties; for example, altering the degree of viscosity. Also, GC-MS analysis revealed information on the presence of other organic molecules, such as lipids and sugars, of animal origin, biological markers of milk, and plant sources too. Also, the identification of synthetic compounds in the glues from the museum in Madrid, are indicative of a rather modern and industrial preparation, contrary to the glues provided by University of Suor Orsola where only natural additives were identified.

Screening of MS/MS data was performed to investigate if the different procedures in the glue preparation have affected collagen degradation. Modifications of amino acids, such as oxidation of methionine, deamidation of asparagine and glutamine, as well as the backbone cleavage, were evaluated as commonly

detected degradation phenomena in aged proteins. AS far as the deamidation of Asn and Gln is concerned, we suggest that deamidation in the case of animal collagen should be collectively calculated for each family of collagen proteins (i.e., all type I collagen sequences) and not for each animal species in the single sample. The reason lies in the similarity of collagen between species. Specifically, with MASCOT or any bioinformatics tool, common peptides are systematically assigned to the protein that has been identified with the highest number of peptides. As a result, and because, in a mixture, peptides that are in common between different species cannot of course be distinguished, the actual calculated deamidation value might be consistently affected if we strive to separate the single chain deamidation. Taking into consideration collectively all the peptides of a protein family, bypasses this problem, leading to more reliable evaluation of deamidation. In our case, mixed-collagen animal glues were found less deamidated than pure animal glues. Furthermore, the collagen chains of *Bos taurus* in hide-mixed collagen glues were found less deamidated in respect to the protein families of the glues, contained pure bovine glues. This observation will deserve further investigation in the future.

An inspection on the structural distribution of deamidation was carried out for the pure animal glues. We observed that some regions appear significantly more deamidated and, interestingly these regions are followed by not detected zones that are fully missing in the identification, suggesting that some three-dimensional arrangement might affect the local deamidation level and overall degradation of the collagen chain. A similar behavior was observed when analyzing methionine oxidation as other degradation parameter. The frequency of backbone cleavage is generally considered another degradation feature of ancient proteins. The number of semitryptic peptides was higher in the strong than in the hide glues, suggesting that the harsher conditions to extract collagen from bones increases backbone fragmentation. However, SDS-PAGE seems to suggest from the analysis of the acid soluble collagen fractions (ACS) of mixed-collagen hide glues that the preparation of mixed collagen glues also led to an evident polypeptide chain fragmentation. More information will come from independent molecular characterization by EGA in the next few months in collaboration with Prof. Ilaria Bonaduce from the University of Pisa, to shed more light on this multi-factorial and complicate material.

Author Contributions

Georgia Ntasi & Sara Sbriglia carried out the proteomic and GC-MS experimental procedures and proteomic data analysis Andrea Carpentieri and Chiara melchiorre, performed the interpretation of the GC-MS data, Rosanna Pitocchi carried out the SDS-PAGE analysis, Ilaria Bonaduce, carried out the EGA-MS analysis. Leila

Birolò, Andrea Carpentieri, Ilaria Bonaduce Gennaro Marino designed the experiments, All authors wrote the paper.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area).

References

- [1] A. Emblem, M. Hardwidge, Adhesives for packaging, in: Packag. Technol., Elsevier, 2012: pp. 381–394. <https://doi.org/10.1533/9780857095701.2.381>.
- [2] Adhesives, in: Build. Decor. Mater., Elsevier, 2011: pp. 325–341. <https://doi.org/10.1533/9780857092588.325>.
- [3] S. Ebnesajjad, Handbook of Adhesives and Surface Preparation: Technology, Applications and Manufacturing, Elsevier Science, 2010. <https://doi.org/10.1016/B978-1-4377-4461-3.10008-2>.
- [4] Schellmann Nanke C., Animal Glues: A review of their key properties relevant to conservation, Rev. Conserv. 3630 (2007) 55–66. <https://doi.org/10.1179/sic.2007.52.Supplement-1.55>.
- [5] M. Buckley, M. Collins, J. Thomas-Oates, J.C. Wilson, Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 23 (2009) 3843–3854. <https://doi.org/10.1002/rcm.4316>.
- [6] F. Welker, M. Soressi, W. Rendu, J.-J. Hublin, M. Collins, Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France, J. Archaeol. Sci. 54 (2015) 279–286. <https://doi.org/10.1016/j.jas.2014.12.010>.
- [7] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J. V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.-J. Hublin, L. Orlando, I. Barnes, R.D.E. MacPhee, Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, Nature. 522 (2015) 81–84.
- [8] G. Leo, L. Cartechini, P. Pucci, A. Sgamellotti, G. Marino, L. Birolò, Proteomic strategies for the identification of proteinaceous binders in paintings, Anal. Bioanal. Chem. 395 (2009) 2269–2280. <https://doi.org/10.1007/s00216-009-3185-y>.
- [9] W. Fremout, M. Dhaenens, S. Saverwyns, J. Sanyova, P. Vandenabeele, D. Deforce, L. Moens, Tryptic peptide analysis of protein binders in works of art by liquid chromatography–tandem mass spectrometry, Anal. Chim. Acta. 658 (2010) 156–162. <https://doi.org/10.1016/j.aca.2009.11.010>.
- [10] S. Dallongeville, M. Koperska, N. Garnier, G. Reille-Taillefert, C. Rolando, C. Tokarski, Identification of animal glue species in artworks using proteomics: Application to a 18th century gilt sample, Anal. Chem. 83 (2011) 9431–9437. <https://doi.org/10.1021/ac201978j>.
- [11] N.C. Schellmann, Animal glues: a review of their key properties relevant to conservation, Stud. Conserv. 52 (2007) 55–66. <https://doi.org/10.1179/sic.2007.52.supplement-1.55>.
- [12] C. Wang, Z. Zheng, X. Deng, X. Ma, S. Wang, J. Liu, Y. Liu, J. Shi, Flexible and powerful strategy for qualitative and quantitative analysis of valepotriates in *Valeriana jatamansi* Jones using high-performance liquid chromatography with linear ion trap Orbitrap mass spectrometry, J. Sep. Sci. 40 (2017) 1906–1919. <https://doi.org/10.1002/jssc.201601406>.
- [13] heterocyclic compounds in plants indole - Search Results - PubMed, (n.d.).

[https://pubmed.ncbi.nlm.nih.gov/?term=heterocyclic compounds in plants indole&page=2](https://pubmed.ncbi.nlm.nih.gov/?term=heterocyclic+compounds+in+plants+indole&page=2) (accessed December 11, 2020).

[14] B. Muszyńska, H. Ekiert, I. Kwiecień, A. Maślanka, R. Zodi, L. Beerhues, Comparative analysis of therapeutically important indole compounds in *in vitro* cultures of *hypericum perforatum* cultivars by HPLC and TLC analysis coupled with densitometric detection, *Nat. Prod. Commun.* 9 (2014) 1437–1440. <https://doi.org/10.1177/1934578x1400901009>.

[15] M. Kawaguchi, K. Syono, The excessive production of indole-3-acetic acid and its significance in studies of the biosynthesis of this regulator of plant growth and development, *Plant Cell Physiol.* 37 (1996) 1043–1048. <https://doi.org/10.1093/oxfordjournals.pcp.a029051>.

[16] T.K.O. Nguyen, P. Marcelo, E. Gontier, R. Dauwe, Metabolic markers for the yield of lipophilic indole alkaloids in dried woad leaves (*Isatis tinctoria* L.), *Phytochemistry.* 163 (2019) 89–98. <https://doi.org/10.1016/j.phytochem.2019.04.006>.

[17] M. Chripkova, F. Zigo, J. Mojzis, Antiproliferative effect of indole phytoalexins, *Molecules.* 21 (2016). <https://doi.org/10.3390/molecules21121626>.

[18] M. Erb, N. Veyrat, C.A.M. Robert, H. Xu, M. Frey, J. Ton, T.C.J. Turlings, Indole is an essential herbivore-induced volatile priming signal in maize, *Nat. Commun.* 6 (2015). <https://doi.org/10.1038/ncomms7273>.

[19] A. Bassoli, L. Merlini, SWEETENERS | Intensive, in: *Encycl. Food Sci. Nutr.*, Elsevier, 2003: pp. 5688–5695. <https://doi.org/10.1016/b0-12-227055-x/01172-x>.

[20] M. Mackie, P. Rütter, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Peggie, H. Howard, L. Harrison, L.J. Jensen, J. V. Olsen, E. Cappellini, Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting, *Angew. Chemie - Int. Ed.* 57 (2018) 7369–7374. <https://doi.org/10.1002/anie.201713020>.

[21] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A.R. Da Fonseca, T.W. Stafford, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, M.T.P. Gilbert, J. V. Olsen, Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, *J. Proteome Res.* 11 (2012) 917–926. <https://doi.org/10.1021/pr200721u>.

[22] E.R. Schroeter, T.P. Cleland, Glutamine deamidation: An indicator of antiquity, or preservational quality?, *Rapid Commun. Mass Spectrom.* 30 (2016) 251–255. <https://doi.org/10.1002/rcm.7445>.

[23] S. Orsini, A. Yadav, M. Dilillo, L.A. McDonnell, I. Bonaduce, Characterization of Degraded Proteins in Paintings Using Bottom-Up Proteomic Approaches: New Strategies for Protein Digestion and Analysis of Data, *Anal. Chem.* 90 (2018) 6403–6408. <https://doi.org/10.1021/acs.analchem.8b00281>.

[24] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: From bones to paintings ☆, *Microchem. J.* 126 (2016) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.

[25] T. HAYASHI, Y. NAGAI, The Anomalous Behavior of Collagen Peptides on Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis Is Due to the Low Content of Hydrophobic Amino Acid Residues¹, *J. Biochem.* 87 (1980) 803–808. <https://doi.org/10.1093/oxfordjournals.jbchem.a132809>.

[26] J. Wu, Z. Li, X. Yuan, P. Wang, Y. Liu, H. Wang, Extraction and isolation of type I, III and V collagens and their SDS-PAGE analyses, *Trans. Tianjin Univ.* 17 (2011) 111–117. <https://doi.org/10.1007/s12209-011-1543-2>.

[27] H. Hong, S. Chaplot, M. Chalamaiah, B.C. Roy, H.L. Bruce, J. Wu, Removing Cross-Linked Telopeptides Enhances the Production of Low-Molecular-Weight Collagen Peptides from Spent Hens, *J. Agric. Food Chem.* 65 (2017) 7491–7499. <https://doi.org/10.1021/acs.jafc.7b02319>.





Supporting information

Material and methods



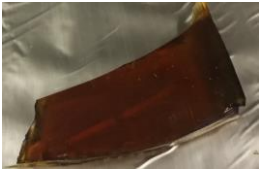



Samples description

18 animal glues of different age, composition, viscosity and hardness, provided by the restoration laboratory of Suor Orsola Benincasa in Naples, and by Museo del Prado in Madrid, have been analyzed and characterized by a proteomic approach.. The name of the samples along with their reported source and age are reported in the Table 1. The photos of all the samples are reported in the figure S1).

Table S1: Animal glue's information and provenience

Animal glue Name	Age	Source
Rabbit glue (postwar) 	After first World war	Suor Orsola Benincasa
Rabbit Glue Totten (sixties) 	60s	Suor Orsola Benincasa
Rabbit glue (CTS) 	60s	Suor Orsola Benincasa
Rabbit glue (pearls) 	60s	Suor Orsola Benincasa

<p>Rabbit glue (Giosi)</p> 	<p>60s</p>	<p>Suor Orsola Benincasa</p>
<p>Rabbit glue 10</p> 	<p>Modern</p>	<p>Museo del Prado</p>
<p>Rabbit glue 2</p> 	<p>Modern</p>	<p>Museo del Prado</p>
<p>Rabbit glue 3</p> 	<p>Modern</p>	<p>Museo del Prado</p>
<p>Rabbit glue 7</p> 	<p>Modern</p>	<p>Museo del Prado</p>
<p>Rabbit glue 6</p> 	<p>Modern</p>	<p>Museo del Prado</p>
<p>Fish glue (flakes)</p> 	<p>60s</p>	<p>Suor Orsola Benincasa</p>
<p>Fish glue 4</p> 	<p>Modern</p>	<p>Museo del Prado</p>

<p>Fish glue 5</p> 	<p>Modern</p>	<p>Museo del Prado</p>
<p>Sturgeon fish glue (pearls)</p> 	<p>60s</p>	<p>Suor Orsola Benincasa</p>
<p>Strong glue (pearls)</p> 	<p>60s</p>	<p>Suor Orsola Benincasa</p>
<p>Strong glue (old manufacturing)</p> 	<p>60s</p>	<p>Suor Orsola Benincasa</p>
<p>Strong glue 8</p> 	<p>Modern</p>	<p>Museo del Prado</p>
<p>Strong glue 9</p> 	<p>Modern</p>	<p>Museo del Prado</p>
<p>Strong glue 1</p> 	<p>Modern</p>	<p>Museo del Prado</p>

Experimental procedures

Organic molecule extraction and analysis by GC-MS

Organic molecule extraction was performed by a multistep extraction protocol as described in [1]. Briefly, different organic solvents from a strong basic (~12) to a strong acidic (~3) pH have been used to separate organic molecules on the basis of polarity. The polar fraction was subjected to sugar analysis, and the nonpolar fraction to transesterification. GC-MS analysis experimental parameters are the same as reported in [1].

Protein extraction and digestion

Samples were prepared as reported in [2]. Briefly, the protein containing fractions were evaporated under vacuum and the pellet resuspended in 10 μ L Urea 6 M and incubated for 20 min at room temperature and then for 30 min in the sonicator. Samples were then 6-fold diluted with ammonium bicarbonate 10 mM pH 7.5 and enzymatic digestion carried out by the addition of 1 μ g of trypsin at 37°C for 16 hours. The supernatants were then recovered by centrifugation, filtered on 0.22 μ m PVDF membrane (Millipore), and peptides were desalted and concentrated on in-house made C18 extraction stage tips as described by Cappellini et al.[3]. Peptides were eluted with 20 μ L of a solution made up of 50% Acetonitrile, 0.1% Formic acid in Milli-Q water and analysed by LC-MS/MS.

Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS–PAGE)

Collagen extraction was performed as reported in *Hong et al 2017*, with slight modifications. 10 mg of each animal glue were under stirring for 24 h in a solution of acetic acid 0.5M (1:10 w/v). The solution was centrifuged at 10000g for 15 min at 4 °C. The supernatant that consists of acid-soluble collagen (ASC) was collected and salted out by dialysis (10 kDa cut off), in a solution of ammonium bicarbonate 100 mg/mL at 4 °C for 24h. pH was measured, verifying neutralization. Subsequently, 10 μ L of each acid soluble collagen (ACS) fraction was treated with 8 μ L sample buffer (65.8 mM Tris-HCl, pH 6.8, 2.1% SDS, 26.3% (w/v) glycerol, 0.01% bromophenol blue) containing 0.1 M DTT, and denatured, at 100 °C for 10 min. The samples were loaded onto a mono-dimensional SDS-PAGE (7%). To visualize the protein bands, gel was stained with Coomassie brilliant blue. Gel was then scanned with ChemiDoc MP imaging system Bio-Rad.

Instrumentation

LC-MS/MS

The animal glues provided by S. Orsola Benincasa restoration laboratory have been analyzed by Q Accurate-Mass Q-ToF LC/MS System, whereas animal glues provided by Museum del Prado by LTQ Orbitrap XL™ Hybrid Ion Trap-Orbitrap MS System.

LTQ Orbitrap XL™ Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany).

C-18 reverse phase capillary column 75 μm *10 cm (Thermo Fisher Scientific), was used with a flow rate of 300 nl/min, with a gradient from eluent A (0.2% formic acid in 2% acetonitrile) to eluent B (0.2% formic acid in 95% acetonitrile). The following gradient conditions were used: t=0 min, 5% solvent B; t=10 min, 5% solvent B; t=90 MIN, 50% solvent B; t=100 min, 80% solvent B; t=105 min, 100% solvent B; t=115 min, 100% solvent B; t=120 min; 5% solvent B. Peptides analysis was performed using data-dependent acquisition of one MS scan followed by CID fragmentation of the five most abundant ions. For the MS/MS experiment, we selected the three most abundant precursors and subjected them to sequential CID-MS/MS acquisitions. For the MS scans, the scan range was set to 400–1800 m/z at a resolution of 60,000, and the automatic gain control (AGC) target was set to 1×10^6 . For the MS/MS scans, the resolution was set to 15,000, the AGC target was set to 1×10^5 , the precursor isolation width was 2 Da, and the maximum injection time was set to 500 ms. The CID normalized collision energy was 35%; AGC target was set to 1×10^5 . Data were acquired by Xcalibur™ software (Thermo Fisher Scientific).

Q Accurate-Mass Q-ToF LC/MS System (Agilent Technologies, Palo Alto, CA, USA)

The samples were fractionated on a C18 reverse-phase capillary column (Agilent Technologies chip) at a flow rate of 400 nL/min, with a linear gradient of eluent B (0.1% formic acid in 95% acetonitrile) in A (0.1% formic acid in 2% aceto-nitrile) from 3% to 80% in 50 min. Peptide analysis was performed using data-dependent acquisition of one MS scan (mass range from 300 to 2000m/z) followed by MS/MS scans of the three most abundant ions in each MS scan. MS/MS spectra were measured automatically when the MS signal surpassed the thresh-old of 50,000 counts. Double and triple charged ions were preferably isolated and fragmented.

Data analysis

Protein Identification

MS/MS spectra were transformed in Mascot Generic files (.mgf) format and routinely used to query the SwissProt database 2015_04 (548208 sequences; 195282524 residues), with Chordata as the taxonomy restriction for protein identification. A licensed version of MASCOT software (www.matrixscience.com) version 2.4.0 was used. Standard parameters in the searches were trypsin as the enzyme; 3, as allowed number of missed cleavages; 10 ppm MS tolerance and 0.6 Da MS/MS tolerance; peptide charge from 2+ to 3+. In all the database searches, no fixed chemical modification was inserted, but possible oxidation of methionine residues, deamidation at asparagines and glutamines, hydroxylation on lysine and proline were considered as variable modifications. To reduce the search space, ultimate searches were carried out using a homemade database, which we named COLLE (60 sequences; 88859 residues), that collects the sequences of collagen type I and III for all the common domesticates generally used for animal glues. Only proteins presenting two or more peptides were considered as positively identified. Protein scores are derived from ions scores as a non- probabilistic basis for ranking protein hits. ([http:// www.matrixscience.com/ help/ interpretation_help.html](http://www.matrixscience.com/help/interpretation_help.html)) Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed peptide match is a random event. Individual ion score threshold provided by MASCOT software was used to evaluate the quality of matches in MS/MS data.

Evaluation of chemical modifications.

The evaluation of deamidation was carried out by MaxQuant software. Parameters common amongst all runs are as follows: tryptic search with up to two missed cleavages; minimum peptide length was set to 7; No fixed modification was set, while oxidation of methionine, hydroxylation of proline and hydroxylation of lysine were set as variable modifications, with up to a maximum of 5 modifications per peptide. Protein identifications were supported by a false discovery rate (FDR) of 0.01 applied (same FDR for dependent peptides when applied) and manually filtered by at least 2 different non overlapping peptides above 40 ion score threshold. Contaminant proteins were assessed using the contamination.fasta provided by MQ, which includes common laboratory contaminants (see MaxQuant Downloads -contaminants.fasta, can be found under http://www.coxdocs.org/doku.php?id=maxquant:start_downloads.htm, n.d) [4]. These protein hits were excluded from further analysis. After each run, the evidence file.txt of each animal glue was firstly cross-validated for its peptides and proteins according to the protein identification that had already been performed with the use of Mascot. Afterwards, the “evidence” files were used for the evaluation of deamidation (N,Q) level both in the total sample (global deamidation) but also for each of the containing

proteins in each animal glue. with a software that is public available(<https://github.com/dblyon/deamidation>). The visualization of all deamidation plots was performed with the use of R studio.

The evaluation of the deamidation (Asn, Gln) and oxidation (M) levels along the amino acidic sequence was performed by manually inspection of MS/MS data. The positions of Asn and Gln that were detected as unmodified were characterized as X, the positions that were found only deamidated as D, the position that were detected both as deamidated and unmodified as DX and, finally, the positions that were not detected at all, nor unmodified nor deamidated, as NF. Similarly, for the evaluation of oxidation, the detected as oxidized positions were characterized as OX and those that have been detected as unmodified, as X and the non-detected as NF.

Backbone cleavage evaluation was carried out by setting the same parameters as above and semitrypsin as enzyme; ion score cut off ≥ 25 for unmodified and modified peptides. The assessment of the occurrence of backbone cleavage was carried out by counting the PSMs (Peptide Spectrum Matches) in the individual samples, focusing the attention to Type I and Type III collagen chains only.

Results

Protein identification

Table S2: Proteins identified in the *Rabbit Glue (Postwar)* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Oryctolagus cuniculus</i>	P02456	2887	57	47	6
COL1A2	<i>Oryctolagus cuniculus</i>	Q28668	2847	55	47	17
COL3A1	<i>Oryctolagus cuniculus</i>	G1T8J0	1689	33	32	15

Table S3: Proteins identified in the *Rabbit Glue (Giosi)* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein name	Taxonomy	Protein identifier	Protein Score	Sequence coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Bos taurus</i>	P02453	2698	48	48	2
COL1A1	<i>Sus scrofa</i>	A0A287A1S6	2492	48	12	5
COL1A2	<i>Sus scrofa</i>	A0A286ZWS8	2140	45	36	12
COL1A2	<i>Bos taurus</i>	P02459	1802	39	18	7
COL3A1	<i>Bos taurus</i>	P02458	1119	28	22	7
COL3A1	<i>Sus scrofa</i>	F1RYI8	865	23	6	4

Table S4: Proteins identified in the *Rabbit Glue (Totten 60s)* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein name	Taxonomy	Protein identifier	Protein Score	Sequence coverage (%)	No. of peptides	No. of Unique Peptides
COL1A1	<i>Bos Taurus</i>	P02453	2704	53	47	1
COL1A2	<i>Bos Taurus</i>	P02459	2568	54	42	7
COL3A1	<i>Bos Taurus</i>	P04258	1669	34	27	10

Table S5: Proteins identified in the *Rabbit Glue (CTS)* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Bos Taurus</i>	P02453	2345	44	43	1
COL1A2	<i>Bos Taurus</i>	P02459	1781	36	29	6
COL3A1	<i>Bos Taurus</i>	P04258	622	16	13	10

Table S6: Proteins identified in the *Rabbit Glue (Pearls)* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein name	Taxonomy	Protein identifier	Protein Score	Sequence coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Bos Taurus</i>	P02459	2497	48	47	2
COL1A2	<i>Bos Taurus</i>	P02453	1647	35	30	4
COL3A1	<i>Bos Taurus</i>	P04258	925	28	22	6

Table S7: Proteins identified in the *Strong glue (Pearls)* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein name	Taxonomy	Protein identifier	Protein Score	Sequence coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A2	<i>Bos Taurus</i>	P02459	2555	51	44	9
COL1A1	<i>Bos Taurus</i>	P02453	2507	44	41	1
COL1A1	<i>Sus scrofa</i>	A0A287A1S6	2367	42	16	5
COL1A2	<i>Sus scrofa</i>	A0A286ZWS8	2292	43	11	9

Table S8: Proteins identified in the Strong glue (Old manufacturing) sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications. Bovine casein was identified against SwissProt database (561911 sequences; 202173710 residues) using Chordata as taxonomy (85899 sequences) and considering methionine oxidation and phosphorylation of serine and threonine as variable modifications.

Protein name	Taxonomy	Protein identifier	Protein Score	Sequence coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A2	<i>Bos Taurus</i>	P02459	2660	55	45	7
COL1A1	<i>Bos Taurus</i>	P02453	932	45	42	5
COL1A1	<i>Sus Scrofa</i>	A0A287A1S6	2250	42	7	4
COL1A2	<i>Sus Scrofa</i>	A0A286ZWS8	2064	40	14	8
COL1A1	<i>Equus asinus</i>	B9VR88	2372	43	4	3
COL1A2	<i>Equus asinus</i>	B9VR89	2027	45	12	8
CASA1	<i>Bos Taurus</i>	P02662	552	19	3	3

Table S9: Proteins identified in the *Fish Glue (Flakes)* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein identifier	Protein Score	Sequence coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Sus scrofa</i>	A0A287A1S6	3107	54	52	8
COL1A2	<i>Sus scrofa</i>	A0A286ZWS8	3026	55	47	19
COL3A1	<i>Sus scrofa</i>	F1RYI8	1345	33	29	15

Table S10: Proteins identified in the *Sturgeon fish glue (pearls)* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Scyliorhinus canicula</i>	D0PQF7	586	13	11	8
COL1A1	<i>Sus scrofa</i>	A0A287A1S6	425	11	6	3
COL1A2	<i>Sus scrofa</i>	A0A286ZWS8	205	8	5	4

Table S11: Proteins identified in the *Strong Glue 1* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A2	<i>BosTaurus</i>	P02459	2841	60	58	8
COL1A1	<i>BosTaurus</i>	P02453	2735	60	61	1
COL1A2	<i>Sus scrofa</i>	A0A286ZWS8	1917	47	44	5
COL1A2	<i>Equus asinus</i>	B9VR89	1656	43	38	7
COL1A1	<i>Equus asinus</i>	B9VR88	2263	51	52	1

Table S12: Proteins identified in the *Strong Glue 2* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A2	<i>Ovis Aries</i>	M4JBX7	2310	59	52	5
COL1A1	<i>Capra hircus</i> or <i>Ovis Aries</i>	A0A452FHU9	2308	53	53	0
COL1A1	<i>Oryctolagus Cuniculus</i>	P02456	2053	60	49	4
COL3A1	<i>Ovis Aries</i>	W5Q4S0	1254	31	32	7
COL1A2	<i>Oryctolagus Cuniculus</i>	Q28668	1240	41	31	10
COL3A1	<i>Oryctolagus Cuniculus</i>	G1T8J0	866	23	22	7

Table S13: Proteins identified in the *Rabbit glue 3* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A2	<i>Oryctolagus Cuniculus</i>	Q28668	2508	59	51	17
COL1A1	<i>Capra hircus or Ovis Aries</i>	A0A452FHU9	2386	56	54	2
COL1A1	<i>BosTaurus</i>	P02453	2297	54	48	0
COL1A1	<i>Oryctolagus Cuniculus</i>	P02456	2231	65	47	5
COL1A2	<i>BosTaurus</i>	P02459	2131	53	45	8
COL1A2	<i>Ovis aries</i>	M4JBX7	1984	50	42	3
COL3A1	<i>Oryctolagus Cuniculus</i>	G1T8J0	1221	32	98	7
COL3A1	<i>Bos Taurus</i>	P02458	1123	28	26	3

Table S14: Proteins identified in the *Fish glue 4* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>BosTaurus</i>	P02453	2393	53	58	2
COL1A2	<i>BosTaurus</i>	P02459	2537	59	57	6
COL3A1	<i>Bos Taurus</i>	P02458	843	31	25	6

Table S15: Proteins identified in the *Fish glue 5* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Bos taurus</i>	P02453	2561	57	53	1
COL1A2	<i>Bos taurus</i>	P02459	2474	60	51	10
COL3A1	<i>Bos taurus</i>	P02458	1303	36	31	8

Table S16: Proteins identified in the *Rabbit glue 6* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A2	<i>BosTaurus</i>	P02453	1463	46	31	5
COL1A1	<i>BosTaurus</i>	P02459	1430	41	34	0
COL1A2	<i>Oryctolagus Cuniculus</i>	NP_001182597.1	569	23	15	3
COL3A1	<i>BosTaurus</i>	P02458	485	16	11	2
COL3A1	<i>Oryctolagus Cuniculus</i>	G1T8J0	251	8	6	1

Table S17: Proteins identified in the *Rabbit glue 7* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A2	<i>Sus scrofa</i>	NP_001230584.1	2617	61%	53	15
COL1A1	<i>Sus scrofa</i>	A0A287A1S6	2524	54%	52	7
COL3A1	<i>Sus scrofa</i>	NP_001230226.1	1949	43%	46	22
COL1A1	<i>Oryctolagus cuniculus</i>	G1T4A5	1757	48	38	1

Table S18: Proteins identified in the *Strong Glue 8* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>BosTaurus</i>	P02453	1845	65	45	1
COL1A2	<i>BosTaurus</i>	P02459	2048	63	48	9
COL1A1	<i>Equus asinus</i>	NP_001310708.1	1624	57	40	4
COL1A2	<i>Equus asinus</i>	NP_001310709.1	1290	48	31	7
COL1A2	<i>Sus scrofa</i>	NP_001230584.1	1130	49	26	5

Table S19: Proteins identified in the *Strong Glue 9* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage %	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Bos taurus</i>	P02459	2337	53%	50	0
COL1A2	<i>Bos taurus</i>	P02453	2228	54%	48	7
COL1A1	<i>Sus scrofa</i>	AOA287A1S6	2242	50%	49	4
COL1A2	<i>Sus scrofa</i>	NP_001230584.1	1860	48%	43	11
COL1A1	<i>Capra hircus</i> <i>o Ovis aries</i>	W5P481	2255	53%	49	0
COL1A2	<i>Ovis aries</i>	M4JBX7	1918	50%	42	2
COL1A1	<i>Equus asinus</i>	NP_001310708.1	1930	44%	42	1
COL1A2	<i>Equus asinus</i>	NP_001310709.1	1476	39%	33	5

Table S20: Proteins identified in the *Rabbit Glue 10* sample. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications.

Protein Name	Taxonomy	Protein Identifier	Protein Score	Sequence Coverage (%)	No. of Peptides	No. of Unique Peptides
COL1A1	<i>Bos Taurus</i>	P02453	2498	56	52	1
COL1A2	<i>Bos Taurus</i>	P02459	2512	57	51	9
COL3A1	<i>Bos Taurus</i>	P02458	1654	39	38	9

Organic molecules identified by GC-MS

Table S20: Compounds identified as TMS by GC-MS in all the animal glue samples.

Retention time \pm 0.2 (min)	Name	Chemical category
15.92	Deoxyspergualin	Synthetic compound
15.99	1H-Indole-2,3-dione, 5-tert-butyl-3-(O-ethylxime)	Heterocyclic compounds
16.77	Estra-1,3,5(10)-trien-17 α -ol	Steroids
16.93	1H-Indene, 2-butyl-5-hexyloctahydro-	Heterocyclic compounds
20.97	Pregna-5,16-dien-20-one, 3-(acetyloxy)-16-methyl	Steroids
21.07	Bisnorabietic acid	Abietane diterpenoids
21.74	Ethyliso-allocholate	Steroids

22.53	Preg-4-en-3-one, 12,17-dihydroxy-20-nitrilo-	Steroids
22.55	Pregn-4-ene-3,20-dione, 16,17-epoxy-	Steroids
22.69	2-Secoandrosta-1,6-diene-17,19-diol, 2-cyano-4-methylene-, diacetate	Steroids
22.74	Abietic acid	Abietane diterpenoids
23.21	3Beta-hydroxy-5-cholen-24-oic acid	Steroids
23.25	Retinal	Polyenes
23.54	Pregna-4,6-diene-3,20-dione, 17-(acetyloxy)-6,16-dimethyl	Steroids
24.29	8-Androsten-3-ol, 17-(2-methylallyl)	Steroids
25.38	5-Chloro-6beta-nitro-5alpha-cholestan-3-one	Steroids
25.41	3,5-Bis[4-(1,1-dimethylethyl)phenyl]-2,3-dihydro-1H-indene-1-one	Heterocyclic compounds

Table S21: Organic molecules identified as TMS by GC-MS.

Retention time \pm 0.2 (min)	Name	Chemical category	Animal glue sample
16.99	Valerophenone	Aromatic compound	Rabbit (Postwar)
24.56	Squalene	Hydrocarbons	Fish (Flakes), Rabbit 6, Strong 8
26.71	Erucamide	Synthetic compound	Strong 1, Rabbit 2, Rabbit 3, Fish 5, Strong 9 and Rabbit 10
16.56	Stevioside	Sugar	Rabbit 7

Table S22: Organic molecules identified as FAME by GC-MS in the animal glue samples provided by the University Suor Orsola Benincasa.

Fatty acids	Hide glues						Strong glues		Fish glue
	Rabbit glue (Postwar)	Rabbit Glue Totin (60s)	Rabbit glue (CTS)	Rabbit glue (pearls)	Rabbit glue (Giosi)	Fish glue (flakes)	Strong glue (old manufacturing)	Strong glue (pearls)	Sturgeon fish glue (pearls)
Oleic acid		✓		✓			✓	✓	✓
Linoleic acid			✓					✓	
Palmitic acid	✓	✓	✓	✓	✓	✓	✓	✓	✓
Palmitoleic acid	✓	✓	✓						
Stearic acid	✓	✓	✓	✓	✓	✓	✓	✓	✓

Myristic acid	✓	✓	✓	✓			✓	✓	✓
Pentadecanoic acid	✓		✓						✓
Margaric acid	✓	✓	✓						
Vaccenic acid	✓	✓			✓				
Adipic acid	✓		✓	✓	✓	✓	✓	✓	
Valeric acid					✓				
Lauric acid								✓	

Table S23: Monosaccharides, identified as TMS by GC-MS in the animal glue samples provided by the University Suor Orsola Benincasa.

Mono-saccharides	Hide glues						Strong glues	
	Rabbit glue (Postwar)	Rabbit Glue Toten (60s)	Rabbit glue (CTS)	Rabbit glue (pearls)	Rabbit glue (Giosi)	Fish glue (flakes)	Strong glue (old manufacturing)	Strong glue (pearls)
Glucose		✓		✓			✓	
Mannose	✓	✓	✓	✓				✓
Arabinose							✓	✓
Galactose	✓	✓	✓	✓			✓	✓
Rhamnose	✓	✓						
Xilose	✓	✓		✓			✓	
N-acetylglucosamine	✓	✓					✓	
Ribitol	✓							
Mannitol						✓		✓

Table S24: Lipids identified as FAMES by GC-MS in the animal glue samples provided by the Museum del Prado in Madrid.

Fatty acids	Hide glues								Strong glues	
	Rabbit 2	Rabbit 3	Fish 4	Fish 5	Rabbit 6	Rabbit 7	Rabbit 10	Strong 9	Strong 1	Strong 8
Oleic acid	✓	✓			✓	✓	✓	✓	✓	✓
Linoleic acid										
Palmitic acid	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Palmitoleic acid	✓	✓					✓			
Stearic acid	✓	✓	✓	✓			✓	✓	✓	✓
Myristic acid	✓	✓					✓			
Pentadecanoic acid	✓									
Margaric acid	✓									
Vaccenic acid	✓									
Methyl elaidate	✓									

Table S25: Monosaccharides, identified as TMS by GC-MS in the animal glue samples provided by the Museum del Prado in Madrid.

Mono-saccharides	Hide glues								Strong glues	
	Rabbit 2	Rabbit 3	Fish 4	Fish 5	Rabbit 6	Rabbit 7	Strong 9	Rabbit 10	Strong 1	Strong 8
Glucose		✓	✓		✓		✓	✓		✓
Mannose	✓	✓	✓	✓	✓		✓		✓	
Arabinose	✓	✓	✓	✓	✓				✓	✓
Galactose	✓	✓	✓	✓	✓		✓	✓		
Rhamnose	✓		✓	✓	✓	✓			✓	✓
N- acetylglucosamine								✓		
Ribitol			✓		✓					
Mannitol			✓	✓		✓			✓	✓
Stevioside						✓				

Chemical modifications in animal glues

Deamidation

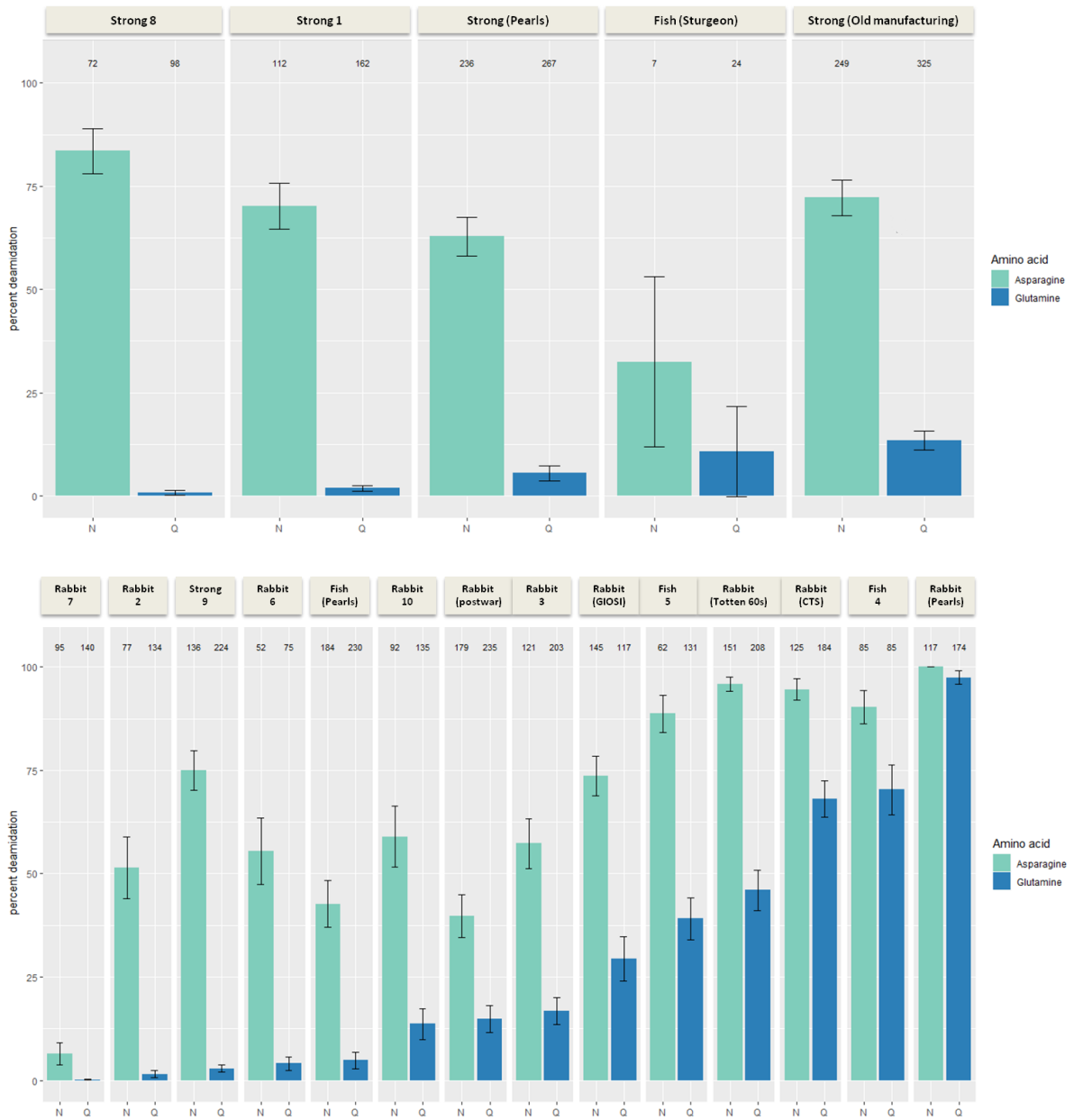


Figure S1: Overall percentage of deamidation for asparagines (N) and glutamines (Q) of strong (upper panel) and hide glues (lower panel). Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

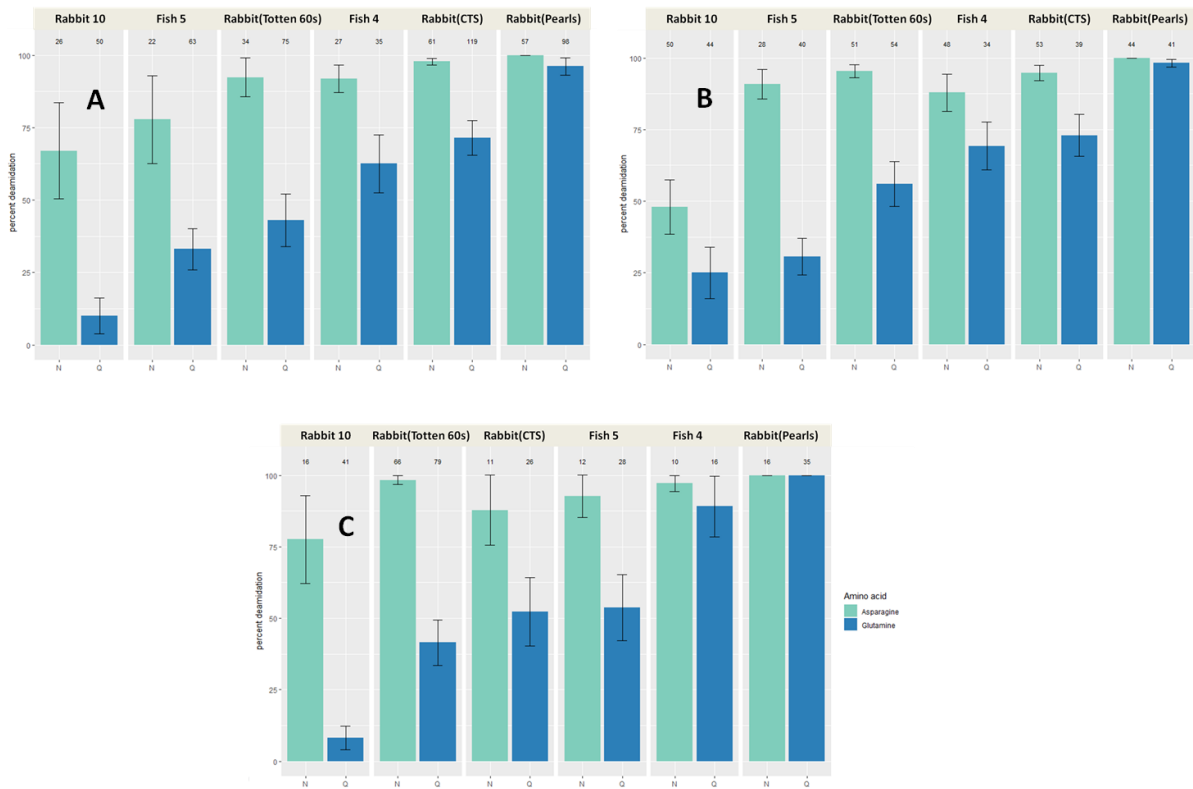


Figure S2: Overall percentage of deamidation for asparagines (N) and glutamines (Q) of the hide glues that contain only bovine collagen alpha-1(I) (A), collagen alpha-2(I) (B) and collagen alpha-1(III) (C). Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

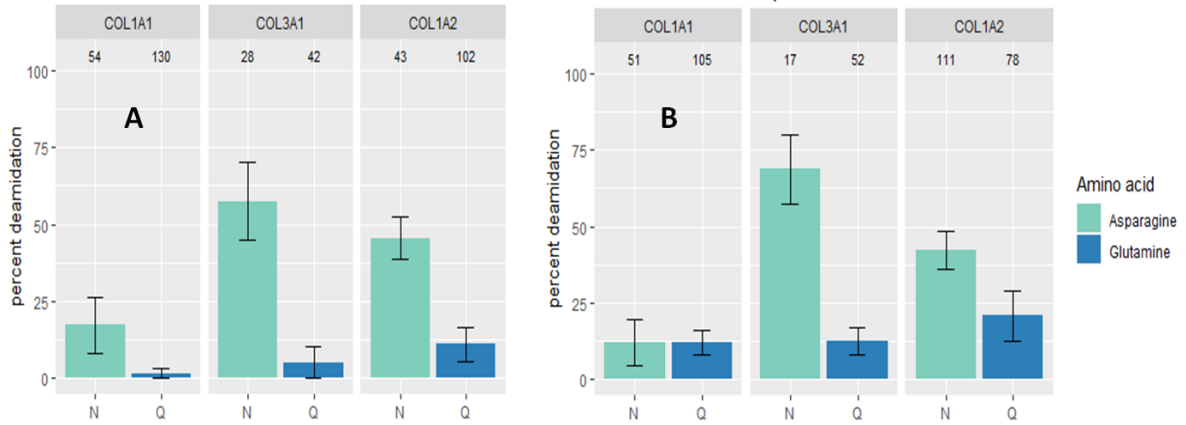


Figure S3: Overall percentage of deamidation for asparagines (N) and glutamines (Q) of the (A): Fish (flakes) hide glue that contain only porcine glue and (B): Rabbit(Postwar) hide glue that contain only rabbit glue. Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on for the several chains of collagen.

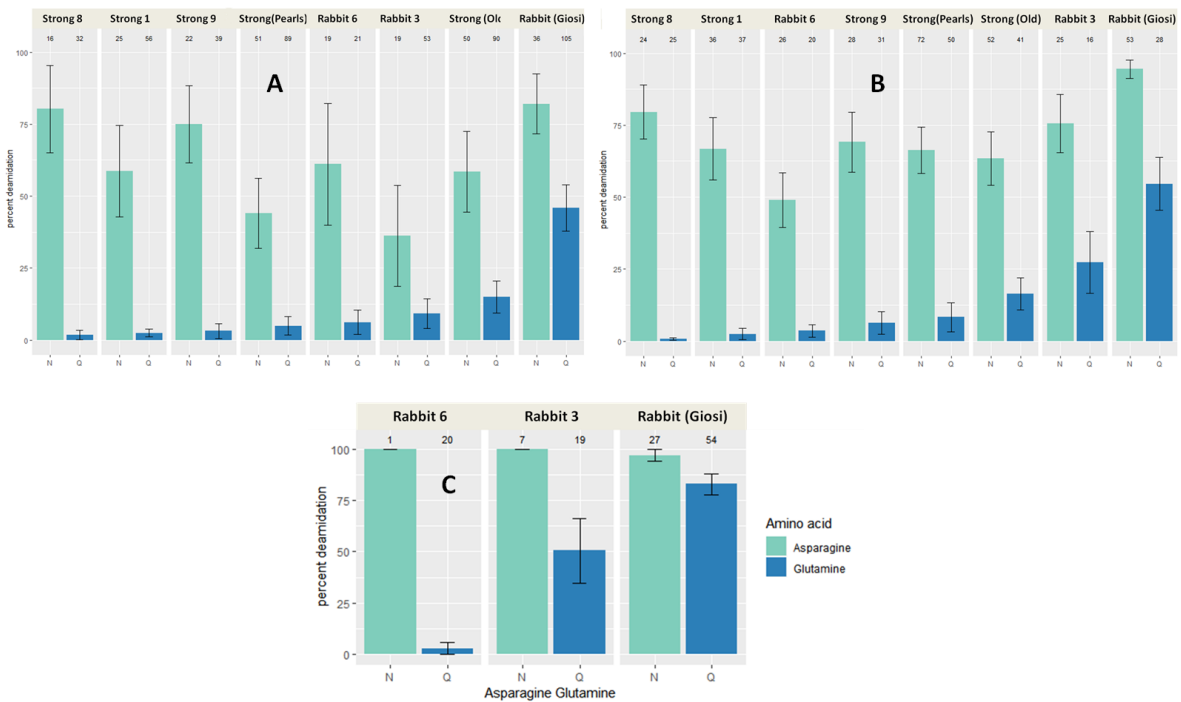


Figure S4: Overall percentage of deamidation for asparagines (N) and glutamines (Q) of the mixed collagen hide and strong glues that contain bovine collagen alpha-1(I) (A), collagen alpha-2(I) (B) and collagen alpha-1(III) (C). Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

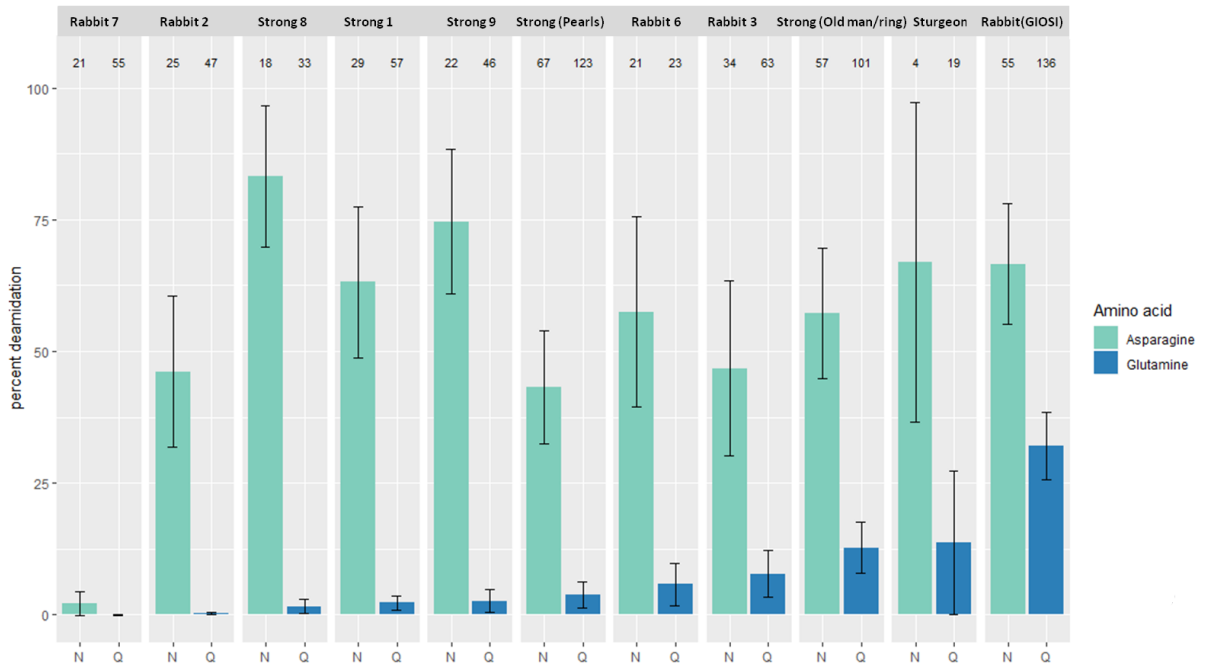


Figure S5: Overall percentage of deamidation for asparagines (N) and glutamines (Q) of the protein family: collagen alpha-1(I). Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

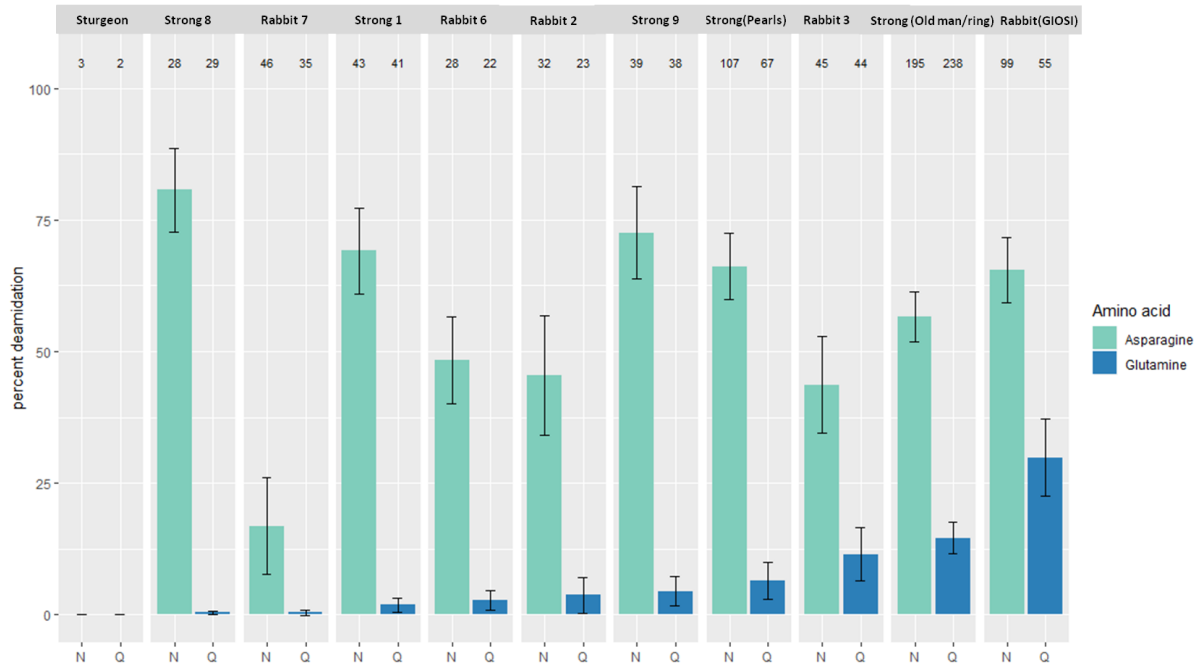


Figure S6: Overall percentage of deamidation for asparagines (N) and glutamines (Q) of the protein family: collagen alpha-1(II). Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

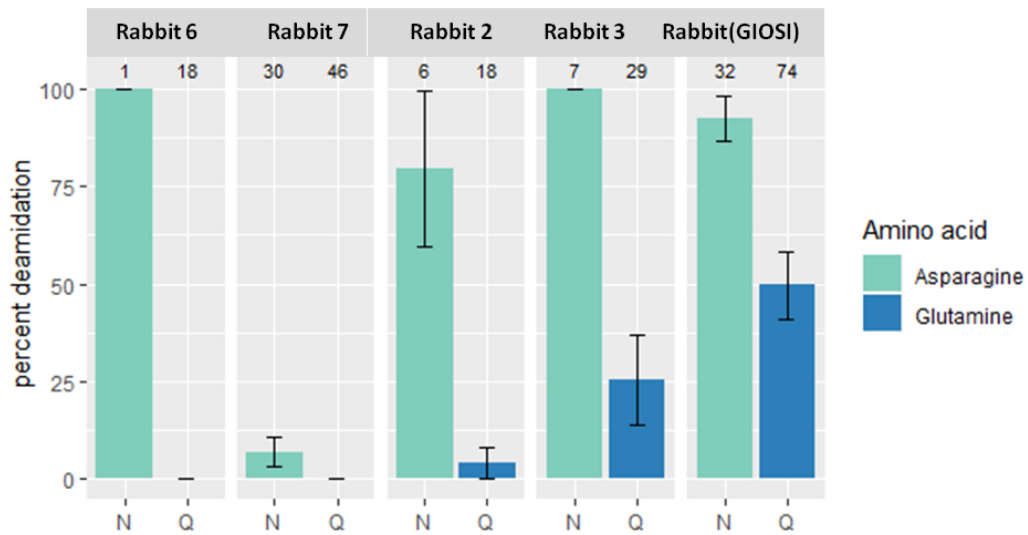


Figure S7: Overall percentage of deamidation for asparagines (N) and glutamines (Q) of the protein family: collagen alpha-1(III). Error bars represent standard deviation and numbers above each bar represent the number of peptides the data is based on.

References

- [1] P. Cennamo, M. Rosaria, B. Lumaga, G. Fatigati, A. Amoresano, A. Carpentieri, CONSERVATION SCIENCE A MULTIDISCIPLINARY ASSESSMENT TO INVESTIGATE A XXII, 11 (2020) 25–38.
- [2] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: from bones to paintings, *Microchem. J.* 126 (2015) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.
- [3] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A.R. da Fonseca, T.W. Stafford, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, M.T.P. Gilbert, J. V. Olsen, Proteomic Analysis of a Pleistocene Mammoth Femur Reveals More than One Hundred Ancient Bone Proteins, *J. Proteome Res.* 11 (2012) 917–926. <https://doi.org/10.1021/pr200721u>.
- [4] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372. <https://doi.org/10.1038/nbt.1511>.

4. Development of micro-invasive techniques for protein identification and characterization in the field of cultural heritage

Introduction

Any new information on the chemical composition of artistic objects or archaeological remains provides keys to decipher ancient and historic materials revealing new insights and/or serving cultural heritage conservation. Recent technical advances in analytical chemistry and, in particular, the astonishing improvements in mass spectrometry (MS), have allowed adapting methods to the study of ancient proteins in artworks and objects of cultural heritage [1]. It is now possible to detect proteins with extreme sensitivity (below femtomoles) and characterize them in detail [2–5]. Protein identification on artworks or archaeological samples has historically been achieved by removing minute amount of the object, and then performing protein digestion followed by classical LC–MS/MS bottom-up proteomics methods. Peptide signals measured in the LC–MS/MS experiment are then matched to virtual ones in the protein databases by the clever use of bioinformatic software such as Mascot (www.matrixscience.com). Although the advent of high performance mass spectrometers allows the identification and quantification of very low amount of substances, the classical protocols and strategies are considered invasive since micro sampling is required.

Even if non-invasive analyses are always recommended from cultural institutions, the level of details that these micro-invasive approaches can provide, ranging from the type of material to the organism origin of the material itself, makes the expense of minute amount of sample fully justified in several cases. However, a continuous search for reducing the invasiveness of the analysis is therefore desirable.

Most recently, as further improvements towards non-invasiveness of the analysis, softer protein extraction methods have been developed. An elegant protein extraction protocol from the parchment surface was proposed that uses an electrostatic charge generated by gentle rubbing of a PVC eraser on the membrane surface [6]. More recently, an ethyl-vinyl acetate film functionalized with strong cation/anion exchange and C8 resins was used for the extraction of proteins and small molecules from the surface of several types of supports [7-8].

Aim of study

This chapter is focused on the development of micro-invasive proteomic strategies in the analysis of archaeological objects and works of art. To meet this aim we took advantage of the adhesion ability of some fungal proteins, hydrophobins, to form a stable and homogeneous layer on a flexible transparency sheet, like those usually employed on overhead projectors, on which trypsin was immobilized. In particular, we used

the self-assembling protein Vmh2, a fungal hydrophobin to functionalize polymer-based sheets where trypsin was immobilized to directly digest *in-situ* proteins on painted surfaces without any minimal removal from the artefact. This method is an improvement towards non-invasiveness of the analysis of proteinaceous binders. Moreover, it is fully portable and can be used in the diagnosis of paintings, valuable materials and ancient archeological artifacts, without the need of micro sampling from the object. These results obtained were published in 2018 on Analytical Chemistry and further improvement and application on objects other than paintings will be published early in 2021.

I will describe the two papers and then a paragraph will be devoted to further developments in progress and perspectives.

Manuscripts:

Cicatiello P, Ntasi G, Rossi M, Marino G, Giardina P, Birolo L. Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings. 2018 Anal Chem. 90(17):10128-10133. doi: 10.1021/acs.analchem.8b01718.

Ntasi G., Kirby D.P., Stanzione I., Carpentieri A., Somma P., Cicatiello P., Marino G., Giardina P., Birolo L. A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects. 2021 J Proteomics. 231: 104039, <https://doi.org/10.1016/j.jprot.2020.104039>.

4.1 Manuscript 1: A minimally-invasive and portable method for the identification of proteins in ancient paintings

Paola Cicatiello[#], Georgia Ntasi[#], Manuela Rossi, Gennaro Marino, Paola Giardina, Leila Birolo

Abstract

As it has been previously shown, the class I hydrophobin Vmh2, from the basidiomycete fungus *Pleurotus ostreatus*, is able to spontaneously form bioactive layers [9] on several surfaces, such as Teflon [10], polystyrene [11], silicon [9,12] and graphene [13], forming extremely stable films. Also it has been already reported that hydrophobin layers are amphiphilic [14] and can efficiently bind functioning proteins allowing us to develop several biotools [11-12,15-17]. In particular a Vmh2 self-assembled layer was used as a coating of a MALDI steel sample-loading plate, either as a simple and effective desalting method prior to MALDI analysis [18] and as a lab-on-plate by immobilizing several enzymes commonly used in proteomic studies [19].

By taking advantage of the above experience we have included cellulose acetate sheets as a novel support surfaces for Vmh2 hydrophobin. We have then immobilized trypsin on the hydrophobin layer and demonstrate that the bioactive film can be used to obtain *in situ*, directly on the work of arts, sufficient peptides to unambiguously identify the proteinaceous binder by proteomic analyses. Optical microscopy analysis has shown that the set-up procedure does not visibly affect the work of art itself. To the best of our knowledge, the method described herein is among the least invasive protocols to identify the proteinaceous binders in works of art thus represents a further step towards the non-invasive molecular analysis of artefacts.

Novelty

Major advantages of this bioactive film are: i) it can be used without the need to transport the works of art; ii) there is no need for specific technical skills to collect the sample since functionalized sheets can be provided as dried films and moistened just before use, then simply put in contact with the work of art. The bioactive film was developed as a user-friendly device, since sampling could eventually be carried out by restorers; after use, the sheet can be stored and peptides recovered just before mass spectrometric analysis. This project was described by the European Union as one of the most innovative EU-funded projects with a "remarkable" level of market capability, ranking it among the most advanced technology development processes (<https://www.innoradar.eu/innovation/36157>).

Materials and Methods

Hydrophobin production

The protein Vmh2 was extracted from the mycelium of *Pleurotus ostreatus* as reported by Gravagnuolo et al., 2016 [20].

Immobilization of trypsin on Vmh2 coated polyester surface and *in-situ* hydrolysis

A sheet of cellulose acetate was cut under hood obtaining small pieces of 2x2 cm. Then, 0.05 mL of 0.01mg/mL of Vmh2 in 60% ethanol solution was spotted on the surfaces and incubated at 60°C until the complete evaporation of solution. The pieces were washed with 0.3 mL of 60% ethanol solution to remove the excess of protein and dried at 60°C. This procedure was repeated twice before to proceed with the enzyme immobilization. 0.05 mL of 0.1mg/mL of trypsin solution in 50 mM of ammonium hydrogen carbonate at pH 7.0, (AMBIC), was added on the coated surfaces for 5 minutes at room temperature. After the incubation, the surfaces were washed with 0.3 mL of 50 mM of AMBIC to remove the excess of trypsin and they were left to dry at room temperature. Before the use, a solution of AMBIC was sprayed on the functionalized surfaces, to moisten the surface and restore the optimal working environment of the trypsin.

Hydrolysis of bovine serum albumin (BSA) on trypsin functionalized surface: 0.05 mL of 0.1 mg/mL of BSA were spotted on the functionalized surface and the reaction was carried out for 5 minutes at room temperature. Then the supernatant was removed, and the surface was dried. 0.025 mL of AMBIC were used to extract the peptides.

Hydrolysis of pictorial tests and historical samples on trypsin functionalized surface: three different pictorial tests, containing either casein or egg, or animal glue were used. They were cut into small pieces of 0.5x0.5 cm and were put in direct contact with the trypsin functionalized surface for 10 minutes at room temperature. Subsequently, the surfaces were left to dry after the removal of the pieces from the samples, and eventually, 0.025mL of AMBIC were used to extract the peptides.

The peptides solution was dried under vacuum and dissolved in 0.1% of formic acid to be spotted on MALDI plate and analyzed by MALDI-TOF.

Water Contact Angle

Contact angle measurements were performed on a KSV Instruments LTD CAM 200 Optical Contact Angle Meter coupled with drop shape analysis software. Each contact angle was calculated as the average of two drops of 5 μ L, spotted on different points of the surface.

Description of samples.

Pictorial samples were prepared as reported in [21]. Historical samples were obtained from the restoration laboratory of S. Orsola Benincasa, provided by Dr. P. Somma, University Suor Orsola Benincasa, and the sample from the Monumental Cemetery in Pisa was kindly provided by Prof. M.P. Colombini from the University of Pisa.

Proteomic analysis

Positive Reflectron MALDI spectra were recorded on a 5800 MALDI ToF/ToF instrument (Sciex, Framingham, MA). The analyte solutions were mixed with α -Cyano-4-hydroxycinnamic acid (10 mg/mL in 70% acetonitrile and 30% of 50 mM citric acid) as the matrix. Typically, 1 μ L of matrix was applied to the metallic sample plate and 1 μ L of analyte was added. The mixture thus obtained was then dried at room temperature. Mass calibration was performed using external peptide standards purchased from Sciex. MS spectra were acquired from 400 to 5000 m/z, for 10000 laser shots total from randomly chosen spots per sample position. Laser intensity remained fixed for all the analyses. MS/MS analyses were performed using 1 kV collision energy with air as CID gas., and MS/MS spectra analysis was performed manually.

Raw data were analyzed using the computer software provided by the manufacturers and are reported as monoisotopic masses.

Microscopy analysis

Morphological and textural analyses of samples were carried out by means of Zeiss Axio Zoom V16 motorized microscope for large fields. The objective aperture of Axio Zoom V16 is big compared to stereomicroscopes and leads to resolution rates that are clearly better and are equipped with a system of direct drive of the zoom, with 99 % reproducibility and allows for maximum precision and accuracy for always correctly scaled images. The microscope configuration for this study is with objective Apo Z 1.5X/0.37 FWD 30mm. In particular, the images of details were acquired with: magnification of 168x, field of view of 1.4mm, resolution of 0.4 μ m and depth of field of 4 μ m.

The acquisition of images was carried out with AxioCam ICC5 (D) microscopes camera and with Zeiss Axiovision 4 software with Z Stack and Extended Focus modules.

Trypsin Activity Assay

Trypsin activity was measured by the method of Hummel [22] using p-toluene-sulfonyl-L-arginine methyl ester (TAME) as a substrate. One unit of enzyme activity was defined as the amount of the enzyme hydrolyzing 1 μ mol of TAME in a minute at 25 °C, pH 8.1.

Results and Discussion

Recently, the class I hydrophobin Vmh2 from was extensively exploited to successfully functionalize several materials to be used as tools for different biotechnological applications. In this study we tried to use flexible cellulose acetate sheets, such as those used for overhead projection, as hydrophobins support surface.

The actual self-assembling of Vmh2 on this polyester surface was verified by measuring the change of wettability of the surface from hydrophobic to hydrophilic upon Vmh2 protein deposition (Fig S1).

Then, the modified surface was used to immobilize trypsin following a procedure already described [2] . After several washings with 50mM AMBIC pH7.0 to remove the excess of unbound enzyme, the support was dried and ready to use as reported in materials and methods. The workflow in Fig.1 describe, in a cartoon format, the essential steps of the set-up protocol.

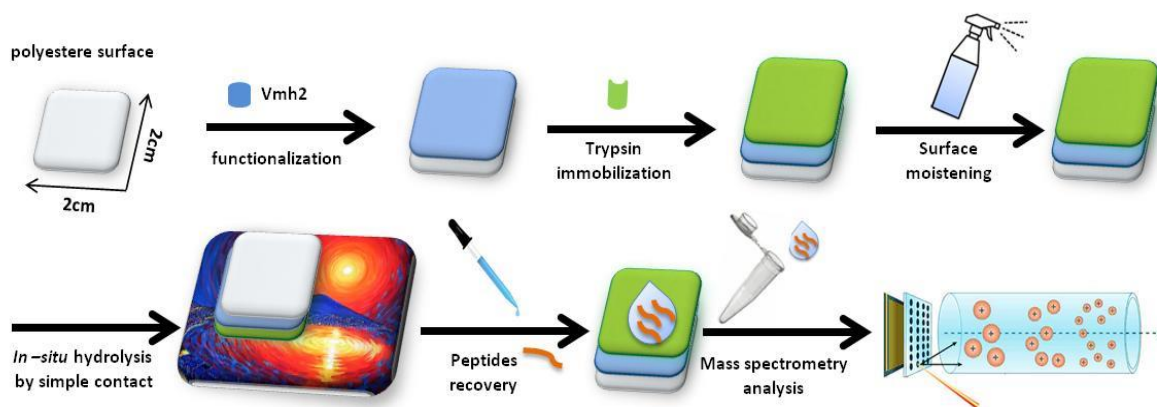


Figure 1: Cartoon showing the essential steps in the experimental workflow.

Development and validation of the diagnostic tool.

The functionalized film was in a first instance tested with a solution of bovine serum albumin (BSA), to verify the actual and active immobilization of trypsin. An aqueous solution of BSA, pH 7.0, was directly spotted on the functionalized surface, and incubated at room temperature for 10 min. The supernatant was recovered, and the peptides mixture analyzed by MALDI-TOF and compared to a control experiment where trypsin was directly added to an aliquot of the same BSA solution. 22% of BSA sequence coverage, compared with the 24% of the control experiment (Table S1), clearly demonstrates that polyester film was actively functionalized with VmH2 and trypsin.

Functionalized polyester was tested with pictorial samples, containing the most common proteinaceous binders, i.e. casein, animal glue and egg, mixed with some common pigments (such as azurite, copper-based pigment, and St. John's white, containing CaCO_3).

Three pictorial samples, containing either casein, egg or animal glue, were put in contact for 10 min at room temperature with the trypsin functionalized flexible sheet, followed by MALDI-TOF analysis of the supernatants. No sample pre-treatment was carried out in order to simulate as much as possible an *in situ* portable usage.

To assess the quality of the results, the performances of the trypsin functionalized film were visually compared to those obtained with the conventional procedure [2] that requires micro-invasive sampling of the object. The MALDI-TOF spectra are reported in comparison to those obtained with the classical procedure for digestion in heterogeneous phase with trypsin in solution (Figure 2), showing a good deal of signals identifying the proteins.

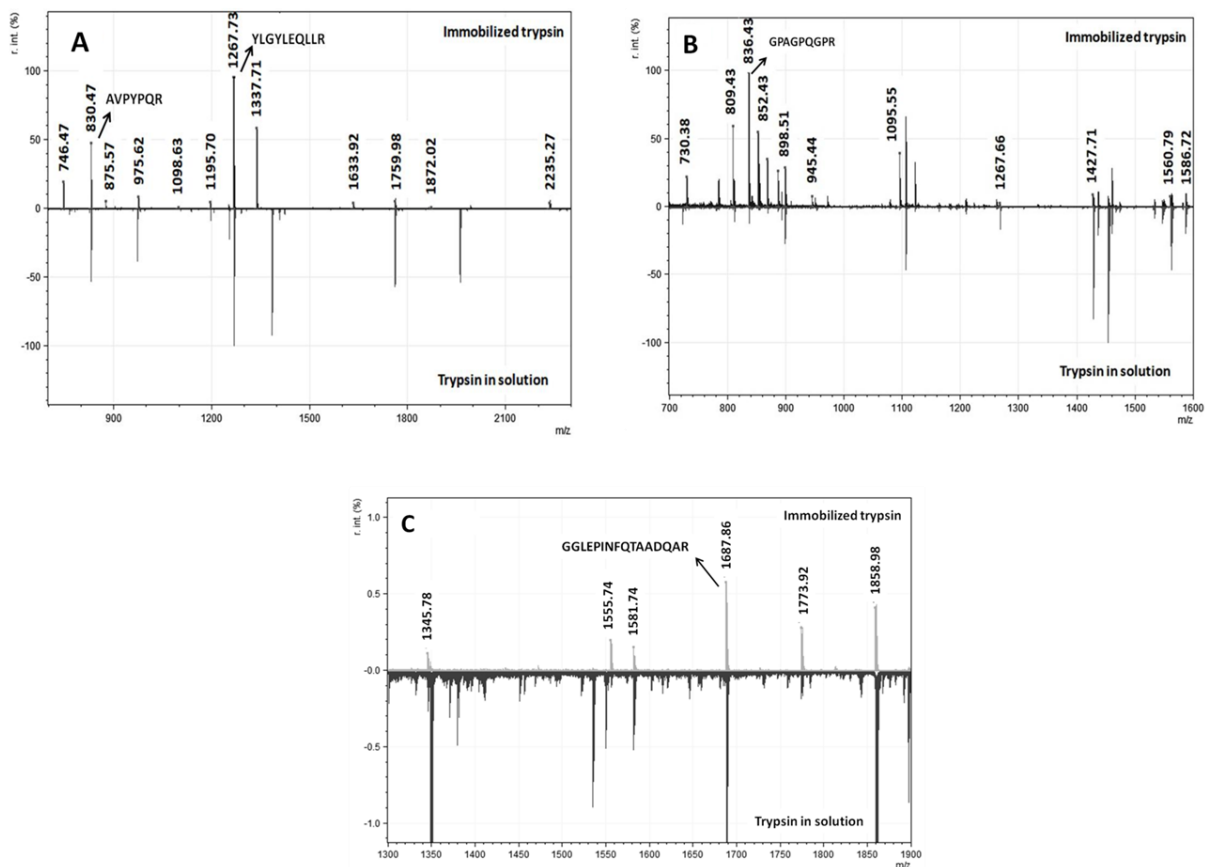


Figure 2: Comparison of the peptide mass fingerprint of pictorial samples with the trypsin functionalized film (top spectra) and the standard protocol² in heterogeneous phase with trypsin in solution (lower spectra). Casein-based binder (A), Animal glue-based binder (B) and egg-based binder (C). Labels refer to identified peptides and sequences are reported for peptides that have been fragmented as further confirmation. Details of the identifications are reported in Table S2 and Table S3.

Some selected peptides were subjected to MS/MS analysis and the product ion (MS/MS) spectra obtained on collision-induced dissociation were manually interpreted as a further confirmation (Figure S2-S5 as examples).

In all the cases, it was possible to confidently identify the binder. A good sequence coverage was obtained (Tables S2 & S3). In glue and casein containing samples, the sequence coverage of the proteins is even higher in comparison to that obtained in the classical heterogeneous digestion, while for the egg containing sample, only ovalbumin was identified, although with good confidence.

Analysis of historical samples.

Binder identification in ancient samples is hindered by aging, since proteins in paintings are prone to deteriorate and degrade. Several reactions occur, still not fully characterized, that can physically hinder the accessibility of proteases, such as aggregation and crosslink formation, while reactions on aminoacidic side chains introduce mass shifts that might complicate proteomic identification. We therefore tested the trypsin functionalized cellulose acetate sheets on eight different historical samples and compared the MALDI-TOF spectra to those of the painting tests reported above. Some selected peptides were subjected to MS/MS analysis and the product ion (MS/MS) spectra obtained on collision-induced dissociation were manually interpreted as a further confirmation. The identified proteins with the detected peptides are listed in table S4. Bovine collagen alpha-1(I) and collagen alpha-2(I) was reliably identified in all the samples. As an example, data for the sample named “Vecchio rifodero” are reported, in comparison with animal glue-based pictorial sample (Figure 3). The labelled signals match to peptides from type I bovine collagen. Manual interpretation of signals at m/z 836.46 and 1427.76 (Figure S6 and S7), confirmed their attribution to peptides 1084-GPAGPQGPR-1092 from bovine collagen 1(I) (P02453) and 572-GIPGEFGLPGPAGAR-586+2Hydroxylation(P) from bovine collagen 2(I) (P02465) respectively.

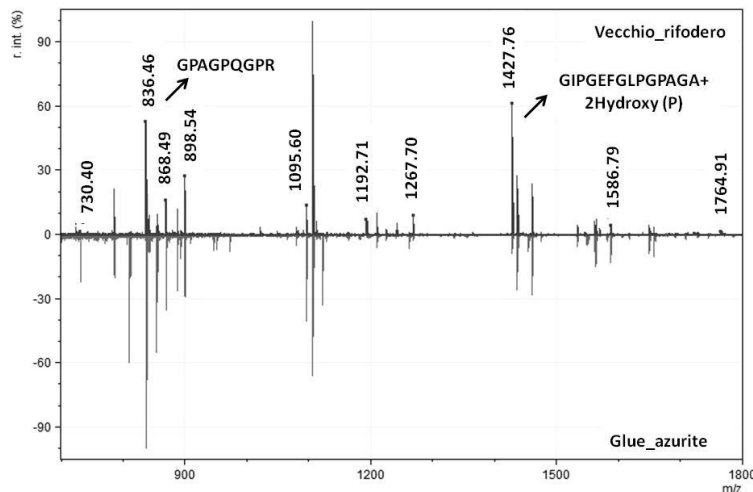


Figure 3: Comparison of the peptide mass fingerprint with the trypsin functionalized film of the historical sample “Vecchio rifodero” with the pictorial sample containing animal glue and azurite (lower spectra). Labels refer to identified peptides and sequences are reported for peptides that have been fragmented as further confirmation (Figures S6 and S7). Details of the identification are reported in Table S4.

Invasiveness of the developed methodology

Invasiveness of the diagnostic method is a very relevant issue. The ability to extract proteins/peptides and/or other molecules from works of art without damaging their surface is of vital relevance in the development of diagnostic tool. Therefore, samples were analyzed by optical microscopy to demonstrate that this approach do not alter the painting surface to make the methodology feasible and applicable to real-life diagnosis of cultural heritage objects. Figures in the supplementary (Figures S8 and S9) show the enlarged images of the surfaces of pictorial samples and historical samples before and after the analysis with the trypsin functionalized sheets. The acquired images show that the treatment does not alter at the microscopic level the surface of the sample, apart from the expected and practically negligible difference that arises because of painting support fibers become wet (figure S10). Moreover, in order to verify that no trypsin was released from the film onto the work of art surface, after the removal of the trypsin functionalized sheet, the surface of the pictorial sample was further washed. Within the limit of the assay sensitivity, no residual trypsin activity was observed to be present in the wash material. This result is in agreement with the observation that no trypsin activity was neither observed in a similar washing of the functionalized film, thus demonstrating that, once immobilized, trypsin is not spontaneously released in aqueous solution and ensuring that trypsin and the painting surface interact only when they are physically in contact and no trypsin is left behind when the physical contact is discontinued.

Conclusions

We demonstrated the possibility of using flexible biofunctionalized cellulose acetate sheets with trypsin to digest proteins and capture peptides from works of art with good efficiency and without affecting at microscopic level the surface of the material under investigation. This flexible tool is fully portable and thus amenable to be used in the diagnosis of paintings, valuable material and ancient archeological artifacts such as those stored in museums and private collections, without the need of micro sampling from the object. An advantage of the developed method is that, once prepared the bioactive film, it could be universally adopted by any scientist working on the work of art without risking of damaging such artefacts samples, and without the need to transport the works of art. The resulting samples can then be analyzed when required, since released peptides are captured onto the film surface, by a simple MALDI-TOF analysis or even by more resolute and informative LC-MSMS analyses. The quality of the identifications is comparable to that obtained with the standard protocol that uses trypsin in solution on samples taken from the work of art.

Supporting material

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b01718.

Author Contributions

G.N, L.B, P.G, P.C, M.R, I.S, G.M. designed research; G.N, P.C, M.R, I.S. carried out the experimental research; G.N, L.B, P.G, P.C, M.R, I.S, G.M. analysed data; G.N, L.B, P.G, P.C, M.R, I.S, G.M wrote the paper.

Acknowledgments

Authors want to thank Dr. Patrizia Somma from Laboratorio di Restauro Tele, Università Suor Orsola Benincasa, Naples, for providing the historical samples for the tests, and Prof. M.Perla Colombini, Dipartimento di Chimica e Chimica Industriale, Università degli Studi di Pisa, for the sample Cimitero Munumentale. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area), and by a grant from the University Federico II "Progetto di Ateneo IENA (Immobilization of ENzymes on hydrophobin-functionalized NANomaterials)

References

- (1) Dallongeville, S.; Garnier, N.; Rolando, C.; Tokarski, C. *Chem. Rev.* **2015**, *116*, 2–79.
- (2) Leo, G.; Cartechini, L.; Pucci, P.; Sgamellotti, A.; Marino, G.; Birolo, L. *Anal. Bioanal. Chem.* **2009**, *395* (7), 2269–2280.
- (3) Leo, G.; Bonaduce, I.; Andreotti, A.; Marino, G.; Pucci, P.; Colombini, M. P.; Birolo, L. *Anal. Chem.* **2011**, *83* (6), 2056–2064.
- (4) Villa, P.; Pollarolo, L.; Degano, I.; Birolo, L.; Pasero, M.; Biagioni, C.; Douka, K.; Vinciguerra, R.; Lucejko, J. J.; Wadley, L. *PLoS One* **2015**, *10* (6), e0131273.
- (5) Lluveras-Tenorio, A.; Vinciguerra, R.; Galano, E.; Blaensdorf, C.; Emmerling, E.; Colombini, M. P.; Birolo, L.; Bonaduce, I. *PLoS One* **2017**, *12* (4).
- (6) Fiddymont, S.; Holsinger, B.; Ruzzier, C.; Devine, A.; Binois, A.; Albarella, U.; Fischer, R.; Nichols, E.; Curtis, A.; Cheese, E.; Teasdale, M. D.; Checkley-Scott, C.; Milner, S. J.; Rudy, K. M.; Johnson, E. J.; Vnouček, J.; Garrison, M.; McGrory, S.; Bradley, D. G.; Collins, M. J. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (49), 15066–15071.
- (7) Manfredi, M.; Barberis, E.; Gosetti, F.; Conte, E.; Gatti, G.; Mattu, C.; Robotti, E.; Zilberstein, G.; Koman, I.; Zilberstein, S.; Marengo, E.; Righetti, P. G. *Anal. Chem.* **2017**, *89* (6), 3310–3317.
- (8) Zilberstein, G.; Maor, U.; Baskin, E.; D'Amato, A.; Righetti, P. G. *J. Proteomics* **2017**, *152*, 102–108.
- (9) De Stefano, L.; Rea, I.; Armenante, A.; Giardina, P.; Giocondo, M.; Rendina, I. *Langmuir* **2007**, *23* (15), 7920–7922.
- (10) Askolin, S.; Linder, M.; Scholtmeijer, K.; Tenkanen, M.; Penttilä, M.; de Vocht, M. L.; Wösten, H. A. B. *Biomacromolecules* **2006**, *7* (4), 1295–1301.

- (11) Wang, Z.; Lienemann, M.; Qiau, M.; Linder, M. B. *Langmuir* **2010**, *26* (11), 8491–8496.
- (12) De Stefano, L.; Rea, I.; De Tommasi, E.; Rendina, I.; Rotiroti, L.; Giocondo, M.; Longobardi, S.; Armenante, A.; Giardina, P. *Eur. Phys. J. E* **2009**, *30*, 181–185.
- (13) Gravagnuolo, A. M.; Morales-Narváez, E.; Matos, C. R. S.; Longobardi, S.; Giardina, P.; Merkoçi, A. *Adv. Funct. Mater.* **2015**, *25* (38), 6084–6092.
- (14) Linder, M. B. *Curr. Opin. Colloid Interface Sci.* **2009**, *14*, 356–363.
- (15) Sun, T.; Qing, G.; Su, B.; Jiang, L. *Chem. Soc. Rev.* **2011**, *40* (40), 2909–2921.
- (16) Longobardi, S.; Gravagnuolo, A. M.; Funari, R.; Della Ventura, B.; Pane, F.; Galano, E.; Amoresano, A.; Marino, G.; Giardina, P. *Anal. Bioanal. Chem.* **2015**, *407* (2), 487–496.
- (17) Zhao, Z.-X.; Qiao, M.-Q.; Yin, F.; Shao, B.; Wu, B.-Y.; Wang, Y.-Y.; Wang, X.-S.; Qin, X.; Li, S.; Yu, L.; Chen, Q. *Biosens. Bioelectron.* **2007**, *22*, 3021–3027.
- (18) Longobardi, S.; Gravagnuolo, A. M.; Rea, I.; De Stefano, L.; Marino, G.; Giardina, P. *Anal. Biochem.* **2014**, *449*, 9–16.
- (19) Longobardi, S.; Gravagnuolo, A. M.; Funari, R.; Della Ventura, B.; Pane, F.; Galano, E.; Amoresano, A.; Marino, G.; Giardina, P. *Anal. Bioanal. Chem.* **2015**, *407* (2), 487–496.
- (20) Gravagnuolo, A. M.; Longobardi, S.; Luchini, A.; Appavou, M. S.; De Stefano, L.; Notomista, E.; Paduano, L.; Giardina, P. *Biomacromolecules* **2016**, *17* (3), 954–964.
- (21) Vinciguerra, R.; Galano, E.; Vallone, F.; Greco, G.; Vergara, A.; Bonaduce, I.; Marino, G.; Pucci, P.; Amoresano, A.; Birolò, L. *Anal. Chem.* **2015**, *87*, 10178–10182.
- (22) B.C.W. Hummel, A MODIFIED SPECTROPHOTOMETRIC DETERMINATION OF CHYMOTRYPSIN, TRYPSIN, AND THROMBIN, *Can. J. Biochem. Physiol.* **37** (1959) 1393–1399. <https://doi.org/10.1139/o59-157>.

4.2 Manuscript 2: A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects.

Georgia Ntasi, Daniel P. Kirby, Ilaria Stanzione, Andrea Carpentieri , Patrizia Somma, Paola Cicatiello, Gennaro Marino , Paola Giardina, Leila Birolo.

Introduction

Proteins from ancient objects can provide key information and contribute to study the context of objects and artists and can provide invaluable information for designing restoration intervention [1]. Most, classical procedures and strategies include digestion protocols in solution and in heterogeneous phase that are considered to be too invasive even when only micro sampling is required [2-4].

In the previous work, of this chapter a novel method towards the non-invasive analysis of proteinaceous materials from artworks was developed, based on a protease functionalized sheet of cellulose acetate to directly digest *in-situ* proteins on painted surfaces without even minimal sample removal from the artefact [5]. The peculiarity of this functionalized film is that, differently from similar existing film/resins [6–9], this is the only one, to the best of our knowledge, that directly allows the digestion of the proteins and, thanks to the adhesive properties of Vmh2, extracts the peptides from the object, whereas the other films remove intact proteins from the artefacts, and the proteins have to be digested afterwards before MS analysis and identification.

After the proof of concept has been shown with pictorial replicas and real paintings [5], we made further improvements by testing:

- I. The effectiveness of the functionalized film approach to identify proteins on a wider range of different historical objects and supports
- II. The possibility of using it for characterizing their conservation state since this sampling is amenable to analysis by LC-MSMS and raw data can be searched successfully for chemical modifications;
- III. The noncovalent immobilization of other enzymes commonly used in protein as well as in small molecules identification/characterization, such as other, glycosidases, being aware that a combination of N-glycosidase F and trypsin improves the identification of egg containing binders [12].

Abstract

Identification and characterization of ancient proteins still require technical developments towards non-invasiveness, sensitivity, versatility and ease of use of the analyses. Enzyme functionalized films were demonstrated to work efficiently on the surface of different objects ranging from fixative-coated paper to

canvas to the coating on an albumen silver print, as well as the much harder surfaces of ivory objects and the proteinaceous binders on a wooden Egyptian coffin. Digested peptides are efficiently captured on the functionalized surface and are also amenable to subsequent analysis to characterize the conservation state of proteins. Moreover, we have combined the trypsin functionalized film with a PNGaseF functionalized film, thus integrating a deglycosylation pretreatment to improve detection of glycosylated coated proteins.

Significance

The enzyme functionalized films provide versatile, user friendly and modular tools that can be widely exploited in the world of diagnosis of cultural heritage objects, with sampling that could be carried out by restorers *on-site* and later analysed with standard MS techniques.

Materials and Methods

Samples for Peptide Mass Fingerprinting (PMF). Paper samples (Barcham & Green Langley Handmade paper) were brush coated with isinglass, albumen and casein fixatives prepared according to historic methods. The salt print photograph was from a study collection and known to have an albumen coating. Ivory samples from walrus, mammoth and elephant were not prepared in any way except to clean the surface with an alcohol swab prior to analysis. Pictures of the objects being sampled are reported in the supplementary materials Fig. S1-S5.

Historical samples.

Historical Canvas. The canvas sample was taken from an oil painting of the Virgin and Child, Saints and Cherubins (an artwork confiscated from Neapolitan mafia) stored in the storehouse of the Cultural Heritage Superintendence of Naples. The painting is attributed to the school of Solimena and can be dated back to 1700. A picture of the sample is reported in the supplementary materials Fig. S6. *Yellow coffin.* The coffin analyzed belongs to the Drosso-Picchianti Collection, MANN in. Naples. The artefact is an inner anthropoid coffin and belongs to a specific type known as “yellow coffin” because of a rich decoration painted on a yellow priming layer. According to its iconography and inscriptions, the coffin is datable towards the end of the XXI and the beginning of the XXII Dynasty (959-889 BC ca.) (Niwiński, Nicola, and Egizio 2004; AA.VV. *Guida alla collezione egizia del MANN* 2016; Cantilena and Rubino 1989). A picture of the coffin can be found in [10].

Hydrophobin Production. The protein Vmh2 was extracted from the mycelium of *Pleurotus ostreatus* as reported by Gravagnuolo et al. [11].

Enzyme immobilization. Trypsin was immobilized on Vmh2 coated cellulose acetate surface as previously reported [5]. PNGaseF was immobilized on Vmh2 coated cellulose acetate surface following a very similar procedure. Briefly, 200 μL of 500 U/mL of PNGaseF solution in 10mM of sodium phosphate pH 7.5 were deposited for 5 min on the Vmh2 coated surface(0.5cm \times 0.5cm). After incubation, the surface was washed with 500 μL of 10mM of sodium phosphate and left to dry at room temperature. As for trypsin activation, a solution of 50mM of ammonium bicarbonate pH 7.0 (AMBIC) was sprayed on the PNGaseF biofunctionalized surface to restore the optimal environment for the enzymatic activity.

***in situ* Hydrolyses for Peptide Mass Fingerprinting (PMF).** Large film pieces, $\sim 20\text{mm}^2$, were spray-moistened with 50mM AMBIC and applied to the surface being analyzed for 10 min. For irregular surfaces, the films were held in place with tweezers to maintain as much surface contact as possible. Films were removed and air dried, and peptides were solubilized from the surface with 25 μL 50mM AMBIC. When smaller film pieces were used, films were cut from the original 20mm^2 pieces into $\sim 10\text{mm}^2$ and $\sim 5\text{mm}^2$ pieces. Rather than spray moistening the films, 5 μL AMBIC (10mm^2 films) or 2 μL AMBIC (5mm^2 films) was pipetted onto the surface/object being sampled, and the film placed on top for 10 min. Films were then placed into Eppendorf tubes with 25 μL AMBIC, vortexed to extract peptides, dried under air and the films removed. The solubilized peptides in Eppendorf tubes were taken to dryness under a stream of air and resolubilized with 10 μL 0.1% TFA (trifluoroacetic acid). MALDI samples were prepared by the addition of 3 μL of the peptides in 0.1% TFA to 20 μL of matrix solution (40% acetonitrile, 0.1% TFA, saturated alpha-Cyano-4-hydroxycinnamic acid) and then spotted onto the MALDI plate.

***in situ* Hydrolyses for LC-MSMS analysis.** When requested, pretreatment of samples with PNGaseF functionalized surface was carried out, similarly to the procedure for tryptic digestion, by putting the sample in direct contact with the PNGaseF functionalized surface for 10 min at room temperature. Then, samples underwent tryptic digestion with the trypsin functionalized film. Digestion was carried out as reported in [5] and peptides recovered in 25-50 μL of AMBIC. For irregular surfaces, the films were held in place with tweezers to maintain as much surface contact as possible. PNGaseF and trypsin hydrolysis in solution were carried out as reported in [12].

Mass spectrometric analyses.

MALDI-TOF analyses were carried out on a 5800 MALDI TOF/TOF instrument (Sciex, Framingham, MA) as reported in [5]. Data analysis was done using the mMass freeware program [13].

For LC-MS/MS analysis, the recovered peptides were filtered on 0.22 μm PVDF membrane (Millipore), and peptides were desalted and concentrated on in-house made C18 extraction stage tips [14]. To increase peptides recovery, the unbound fraction from the stage tips was loaded on a reverse-phase C18 Zip Tip pipet

tip (Millipore). Peptides were eluted with 20 μ L of a solution made of 50% acetonitrile, 50% formic acid 0.1% in Milli-Q water and combined with the eluted fraction from stage tips, dried under vacuum, and finally dissolved in 0.1% of formic acid.

LC-MS/MS analyses were carried out on a 6520 Accurate-Mass Q-TOF LC/MS System (Agilent Technologies, Palo Alto, CA) equipped with a 1200 HPLC system and a chip cube (Agilent Technologies) and on a LTQ Orbitrap-XL (ThermoScientific, Bremen, Germany) as reported in [15], and raw data analyses as reported in [12]. Each LC-MS/MS analysis was preceded and followed by blank runs to avoid carryover contamination. MS/MS spectra were transformed in Mascot Generic files (.mgf) format and routinely used to query the SwissProt database 2015_04 (548208 sequences; 195282524 residues), with Chordata as the taxonomy restriction for protein identification. A licensed version of MASCOT software (www.matrixscience.com) version 2.4.0 was used. Standard parameters in the searches were trypsin as the enzyme; 3, as allowed number of missed cleavages; 10 ppm MS tolerance and 0.6 Da MS/MS tolerance; peptide charge from 2+ to 3+. In all the database searches, no fixed chemical modification was inserted, but possible oxidation of methionine residues and deamidation at asparagines and glutamines were considered as variable modifications. When collagen proteins were identified, further identification runs were carried out with the insertion of hydroxylation on lysine and proline as variable modifications, since more confident identifications are commonly obtained for these proteins by taking into consideration their extensive posttranslational modifications. Moreover, to reduce the search space, these runs were carried out using a homemade database, which we named COLLE (60 sequences; 88859 residues), that collects all the common domesticates used for animal glue. Other parameter changes for the specific database searches are indicated in the Table captions. Only proteins presenting two or more peptides were considered as positively identified. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. (http://www.matrixscience.com/help/interpretation_help.html) Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed peptide match is a random event. Individual ion score threshold provided by MASCOT software to evaluate the quality of matches in MS/MS data was used for confidence threshold in protein identification.

Results and discussion

Application of the micro-invasive trypsin functionalized film to different types of artistic/historical samples.

The trypsin functionalized films were evaluated for detection and identification of proteinaceous materials on objects from cultural heritage other than paintings, including fixative-coated paper (albumen, casein and isinglass), albumen coated salt print photographs, and ivory from elephant, mammoth and walrus. Any alteration of the sampled surface was visually noted since surface alteration is a major consideration when sampling artworks and cultural objects.

A first set of samples was analyzed using 20mm² pieces of film and proteins were identified by visual inspection of MALDI spectra for marker ions. The method is applicable to simple protein mixtures, such as the proteins found in egg tempera, for example, and was proven successful in several cases [16–18] as well as in the first development of this methodology [5]. Marker ions for proteins commonly found in artworks have been published [5,16] and are used here to detect and identify albumen, casein and fish-based materials (See Table S-1, Supplementary Materials).

Intense spectra were obtained for all the objects and surfaces sampled in this first set. No visible alteration of the surfaces of the ivory objects was observed. With the fixative coated papers, the area under the film became slightly swollen, and this relaxed/disappeared as the paper dried. For the albumen coated salt print, the surface under the film was heavily distorted, as anticipated, because albumen coatings on prints are known to be very moisture sensitive. As an example of the quality of the MALDI data obtained, Fig. 1 is the spectrum from the casein-coated paper with casein markers indicated [16,18]. The other samples gave similar results. Keratin was noted in the spectra of many of the samples. This contaminant is often encountered with surface sampling. It generally originates from dust/dirt accumulation, typically epithelial keratin, and does not interfere with data interpretation.

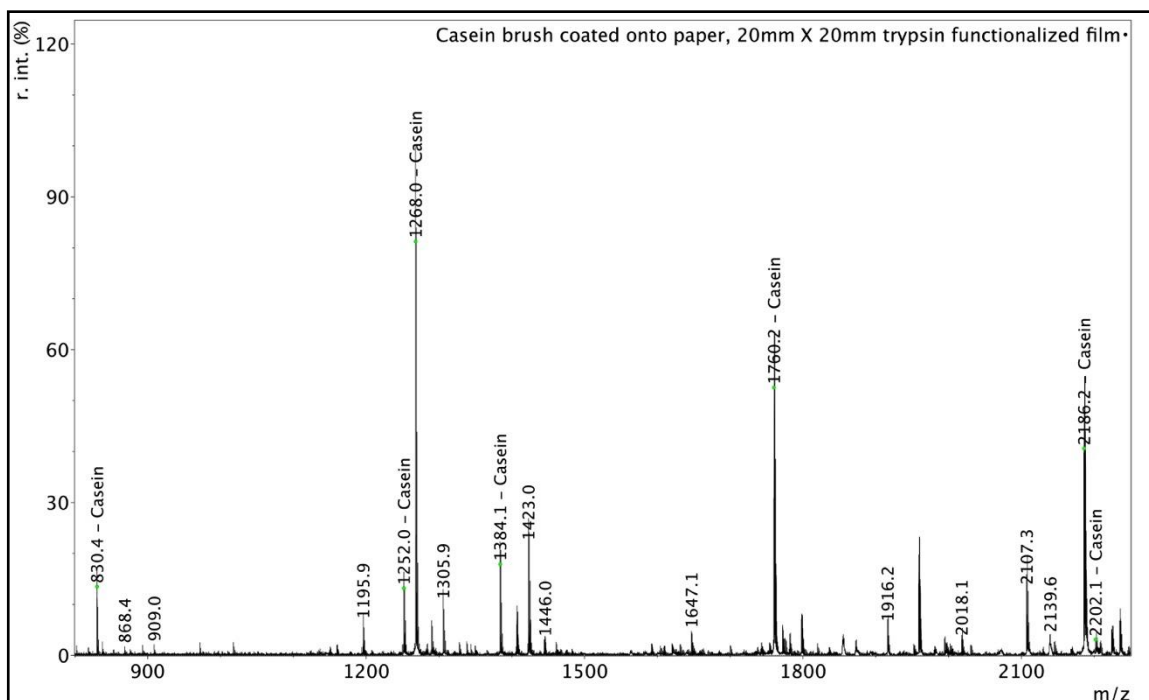


Figure 1: MALDI spectrum from casein coated paper sampled with a 20mm² trypsin functionalized film. Major casein marker ions are indicated.

Given the high detection sensitivity observed with the first set of samples, a second set was analyzed with smaller film pieces (10mm² or 5mm²) and reduced quantities of AMBIC (5μL for 10mm² films and 2μL for 5mm² films, respectively). The AMBIC solution was pipetted directly onto the surface/object being sampled, and the film was placed on top for 10 min. As noted above, with the smaller film formats, the film was then placed directly into the Eppendorf tube thus improving the overall recovery of produced peptides. The results for 10mm² and 5mm² films were very similar, and the quality of the spectra at these reduced sizes indicates that further size reduction will be possible, further minimizing alteration of the surfaces being sampled. Figure 2 compares the results from 10mm² and 5mm² film samples of the casein-fixed paper surface. The inset compares the relative signal-to-noise ratio (S/N) calculated by the mMass program observed for two of the weaker ions with the different film sizes.

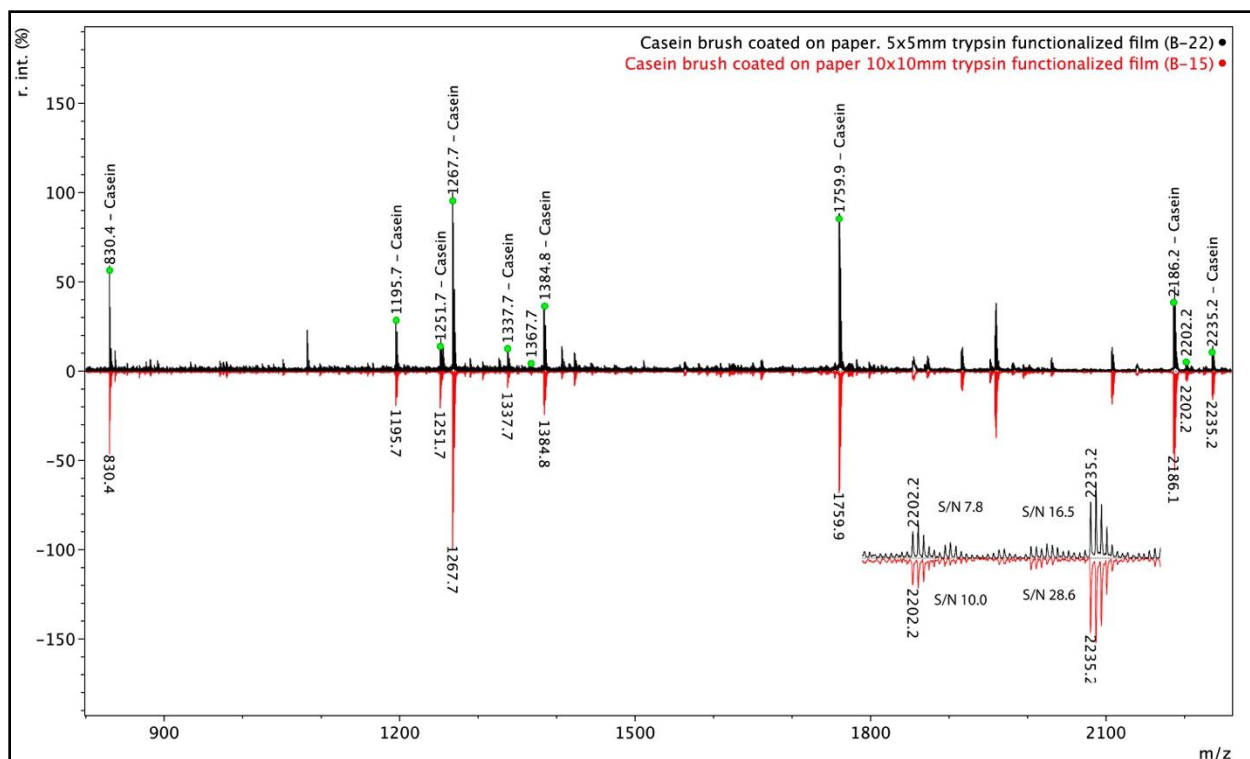


Figure 2: MALDI spectrum from the casein-fixed paper. Upper, black, 5mm² film. Lower, red, 10mm² film. The inset shows the 2202 Da and 2235 Da ions to compare the relative S/N observed with the larger (lower) and smaller (upper) film sizes.

Figure S7-S9 are MALDI spectra from elephant ivory, albumen coated salt print, and isinglass-fixed paper, respectively. PMF is used routinely in archaeology and cultural heritage to identify mammalian sources of collagen-based [19] and keratin-based materials [20]. In the case of skin or hide, gut, sinew, bone and ivory, collagen is the major constitutive protein, and for each mammalian source, the amino acid sequence of collagen is unique. Marker ions in the MALDI spectra from known reference materials [21] are compared with those from unknown samples for identification and are used here to identify ivory from elephant, mammoth and walrus.

It is worth noting the high quality of the spectra obtained, considering that no cleanup or concentration, for example using Zip Tips, was done before analyzing the samples. The peptide mixtures were simply collected from the trypsin functionalized film, dried, resolubilized, mixed with matrix and spotted on the MALDI plate. Given the high sensitivity observed with the smallest film pieces (5mm²), it seems very likely that even smaller formats will be possible, further limiting the impact on sensitive surfaces and allowing sampling of smaller features.

The impact of sampling on sensitive surfaces is extremely relevant in the field of cultural heritage. PMF, one of the modern techniques for protein analysis recently introduced into conservation science, is advantageous because of its sensitivity and specificity, speed of analysis, and relatively low cost. The minimal samples required for analysis are making it possible to study art and objects where the consequences of removing even micro-amounts of material can be prohibitive. Improved methods for sampling sensitive objects are an important consideration for conservation, thus the interest in investigating the use of trypsin-functionalized films.

The trypsin-functionalized film can be used for the characterization of ancient proteins.

The trypsin functionalized film has been shown to be amenable for protein identification in several artistic objects, thus showing up as an extremely interesting and powerful tool for all stakeholders in cultural heritage interested in understanding the composition of works of art.

However, the story of ancient proteins does not stop at the identification level. It is now more and more common that, beside understanding the composing material, chemical analyses must face the challenge of understanding at the molecular level the conservation state of the work of art. This is at the basis of a wide range of questions, from authentication to restoration interventions. Can the functionalized film be used in this perspective? That would mean that the peptide mixtures recovered from the film can be analysed in search for molecular signatures of degradations, i.e. chemical modifications such as oxidations, deamidations or spontaneous hydrolysis. To meet this aim, a sample from a historical canvas dated back to 1700 was digested with the trypsin functionalized flexible sheet and the recovered supernatant analysed by LC-MSMS. The results were compared to those obtained with a control experiment where trypsin was directly added to an aliquot of the same sample, in the conventional procedure in heterogeneous phase [22], that requires microinvasive sampling of the object. A summary of the identification results is shown in Table 1 and the details of the identifications are reported in the supplementary results (Table S2-S3).

Table 1: Identification of Proteins in Historical Canvas by LC-MSMS. Samples were subjected to *in situ* digestion with the trypsin functionalized film and peptide mixtures were analyzed by LC-MSMS. Results were compared to those obtained with the standard protocol in heterogeneous phase [22]. Raw data were searched by Mascot MS/MS Ion search using the homemade COLLE database. Deamidation on Gln and Asn, hydroxylation at Pro and Lys, oxidation on Met, were considered as variable modifications. Identification details are reported in Table S2-S3.

		Digestion with the Trypsin functionalized film				Overnight digestion in heterogenous phase			
Sample	Protein (Uniprot entry)	Score	Protein Coverage (PSC) (%)	No. of sequences	No. of Unique sequences	Score	Protein Coverage (PSC) (%)	No. of sequences	No. of Unique sequences
Historical canvas	Collagen 1(I) (W5P481)	1223	27	26	4	2621	53	53	2
	Collagen 2(I) (W5NTT7)	757	22	19	4	2847	43	36	5
	Collagen 1(III) (W5Q4S0)	597	15	14	5	1212	29	30	5

Not unexpectedly, identification was on average better with the digestion in heterogeneous phase, most likely because of the much longer incubation time (o/n digestion versus 10 min, roughly grinded samples to expose larger surface, constant 37°C incubation, etc.).

Despite the lower yield in peptides, data collected from the samples recovered from the trypsin-functionalized film are amenable to be used for characterizing the degradation state of the proteins inside the samples. MS/MS data were used to search common chemical modifications classically linked to molecular damage pattern, such as deamidation at asparagine (Asn) and glutamine (Gln) residues [23,24], spontaneous hydrolysis[25,26], and glycation at lysine (Lys) and arginine (Arg) [24,27].

As it can be seen from the figure 3 the levels of collagen deamidation that can be calculated are comparable with the two protocols, that means that, despite the lower yield in peptides with the film, there is no bias in extracting deamidated rather than unmodified peptides in the samples.

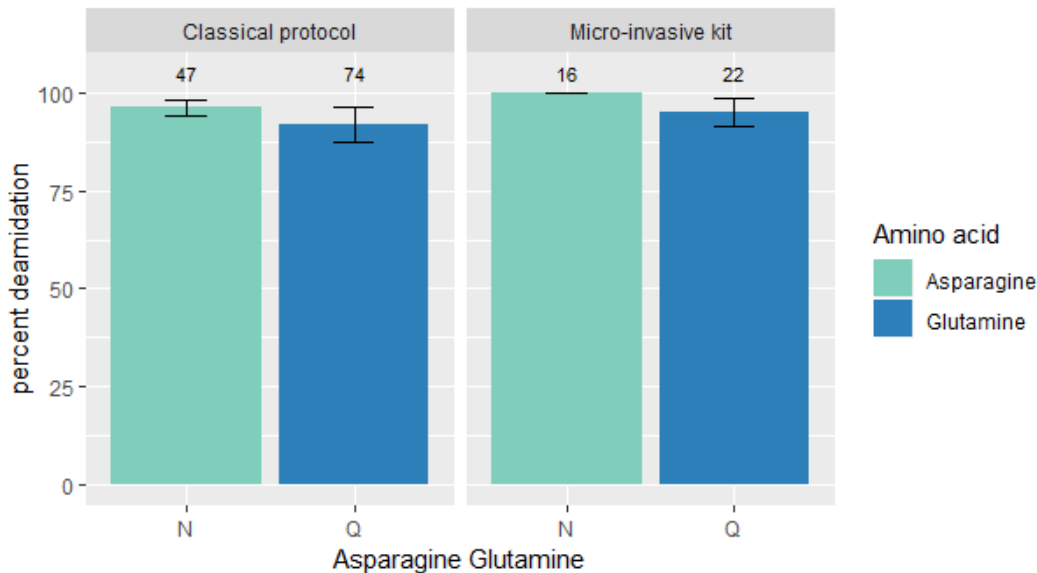


Figure 3: Overall percentage of deamidation for asparagine (N) and Glutamine (Q) residues for the proteins identified on the surface of the historical canvas detected with the trypsin functionalized film in comparison with the standard approach in solution. Error bars represent standard deviation and numbers above the bars represent the number of peptides containing Asn and Gln on which the data was based on.

This result is extremely important since deamidation is actually considered an indication of “authentic age” [2,23] of the sample, although we embrace the sense by Schroeter and colleagues [28] that deamidation should be rather viewed as a global indicator of their preservational quality, since deamidation rates and levels are greatly affected by numerous chemical and environmental factors.

We also searched for spontaneous hydrolysis and modifications such as glycation, both non-enzymatic modifications well known to occur upon aging/deterioration. Table 2 reports the results of the search for semitryptic peptides, which are the hallmark for partially hydrolyzed proteins[26,27].

Table 2 :Characterization of spontaneous hydrolysis in proteins in a historical canvas by LC-MSMS. LC-MSMS raw data from *in situ* digestion with the trypsin functionalized film were searched by Mascot MS/MS Ion search using COLLE database considering semitrypsin as the enzyme, deamidation on Gln and Asn, oxidation on Met, hydroxylation at Pro and Lys as variable modifications. Identification details are reported in Table S4-S5.

		Digestion with the Trypsin functionalized film				Overnight digestion in heterogenous phase			
Sample	Protein (Uniprot entry)	Score	PSC (%)	No. of peptides	No. of Semitryptic peptides	Score	PSC (%)	No. of peptides	No. of Semitryptic peptides
Historical canvas	Collagen 1(I) (W5P481)	3970	48	113	90	3852	55	120	55
	Collagen 2(I) (W5NTT7)	2326	43	75	57	2507	52	74	34
	Collagen 1(III) (W5Q4S0)	1286	40	55	35	1468	35	51	18

It is worth mentioning that, when semitryptic peptides are considered for protein score computation, the results became almost comparable. However, if we go into the details, the trypsin functionalized film seems to extract a larger extent of “damaged” peptides, since semitryptic peptides exceed the 80% of the detected peptides with the film versus the 50% of the peptides observed with the standard protocol in heterogeneous phase. A likely explanation might lie in the fact that trypsin in solution could penetrate more deeply into the layer, whereas when immobilized on the film it can attack only the external, and therefore more degraded, surface of the object. This aspect will deserve further investigations.

Even more interesting is the observation that almost all the glycation products, arising from exposure of proteins to sugars, were detected on semitryptic peptides. Tables S6-S7 illustrate the peptides were advanced glycation end-products (AGEs) products of lysine and arginine were found. A wide set of AGEs was detected in the sample from trypsin functionalized film: glycation of arginine with the formation of glyoxal and methylglyoxal adduct (mass shift of + 39.99 and +54.01 Da respectively), as well as formation of carboxymethyl and carboxyethyl lysine (with mass shift of +72.01 and +58.01 Da, respectively).

Glycation products are consistent with the presence of sugars as identified by GC-MS analysis (data not shown).

Conclusively, we suggest that the trypsin functionalized film can be used to gain information on chemical modifications of proteins inside an artistic object and therefore on its degradation state and not only to merely identify the protein component.

Development and Implementation of the approach.

The success of proteomics experiments in cultural heritage objects lies in the efficiency of either protein extraction from solid matrixes or proteolytic digestion of substrates incorporated inside the matrix itself. In this respect, we already observed that glycosylation of egg proteins created a significant molecular hindrance which hampered proteases to efficiently interact with the proteinaceous substrates, thus greatly decreasing proteolytic efficiency and therefore reducing the confidence of protein identification [12]. In these cases, in the standard approach in solution, we bypassed the problem of low confidence and reliable identification by introducing a deglycosylation step before treatment with trypsin. What if we want to apply the trypsin functionalized film? Can we introduce similarly a treatment with PNGaseF to get rid of the sugar layer that, as we suggested, seems to create a molecular hindrance that prevents the accessibility of proteases? We reasoned to adopt a similar strategy, by immobilizing also PNGaseF on a separate acetate of cellulose sheet to trim out the glycosidic decoration before proceeding with the trypsin functionalized film. The PNGaseF functionalized film was in a first instance tested on a pictorial egg-based binder that contained azurite as pigment. Briefly, the sample was placed for ten minutes in contact with the PNGaseF functionalized film and subsequently for 10 minutes in contact with the trypsin functionalized film. As a control, the same sample was also analysed only with the trypsin functionalized film. MALDI-TOF analysis of the recovered supernatants was then recorded as a pre-screening and the figure S10 demonstrates a richer spectrum in case of use of PNGase functionalized film. Therefore, even in this more elaborated configuration, it is possible to successfully analyse samples with a simple, rapid, sensitive and specific approach such as Peptide Mass Fingerprinting, still amenable to a museum laboratory, and that can be practiced successfully even by non-specialists.

Table 3 reports the LC-MS/MS results, confirming the effectiveness of the deglycosylation pretreatment step with immobilized PNGaseF on functionalized films, since more proteins were identified when the sample was pretreated with PNGaseF and all more confidently. Details are reported in Table S8-9

Table 3: Proteins Identified in the Paint Replica Containing Azurite and Albumen by LC-MS/MS. Aliquots of paint replica were *in situ* digested with the trypsin functionalized film with or without pretreatment with PNGaseF functionalized film and peptide mixtures were analysed by LC-MS/MS. Raw data were searched by Mascot MS/MS Ion search using Chordata as the taxonomic restriction in the SwissProt protein database, with deamidation on Gln and Asn, and oxidation on Met, as variable modifications. ND: not detected. Details of the identification are given in Table S8 and S9.

		PNGaseF functionalized film + Trypsin functionalized film				Trypsin functionalized film			
Sample	Protein (Uniprot entry)	Score	PSC (%)	No. of peptides	No. of Unique peptides	Score	PSC (%)	No. of peptides	No. of Unique peptides
Pictorial- egg based sample	Ovalbumin (P01012)	446	47	15	9	252	30	9	8
	Ovotransferrin (P02789)	186	13	9	8	117	7	6	5
	Ovalbumin-related protein Y (P01014)	168	13	7	6	ND	ND	ND	ND
	Elongation factor 1- alpha 1 (Q90895)	74	7	4	4	57	4	2	2
	Ubiquitin (P79781)	49	19	2	2	ND	ND	ND	ND
	Ovalbumin (P01012)	446	47	15	9	252	30	9	8

The combination of PNGaseF and trypsin immobilized on cellulose acetate sheets was exploited on a wood micro-sample, collected from the decoration of XXII dynasty wooden coffin (yellow coffin), where egg proteins had been recently identified with the standard protocol in heterogeneous phase, and GCMS analysis had also detected the copresence of gum Arabic (reference: sample Sarcophago A in the paper by Melchiorre *et al* [10]). As a further control, the same sample was also digested only with the trypsin functionalized film. The recovered peptides were subjected to LC-MS/MS analysis. Results are shown in Table 4 and details are reported in Tables S10-11. First of all, we can conclude that the trypsin functionalized film is also effective on wooden surface, thus demonstrating again its versatility of use. Identifications are in fairly good agreement with the already published data [10]. When the combined PNGaseF plus trypsin protocol was applied, the trypsin functionalized film produced more confident identifications since, as we reported in [10], lysozyme was the only protein detected. Moreover, we ascribed lysozyme to *Colinus virginianus*, although the detected peptides could not actually discriminate between the little bird and *Gallus gallus*. In the experiment with the trypsin functionalized film instead, an unambiguous chicken origin could be depicted from the numerous proteins that were identified. Moreover, a clear identification of several collagen chains confirms that a mixture of egg and animal glue was present, accordingly to what had already been observed in the different layers of the coffin [10].

Table 4: Proteins Identified in the wooden micro sample collected from the decoration of XXII dynasty wood Sarcophagus (yellow coffin) with the use of PNGaseF and trypsin functionalized films. Aliquots of wooden micro sample were *in situ* digested with the trypsin functionalized film with or without pretreatment with PNGaseF functionalized film and peptide mixtures were analysed by LC-MS/MS. Raw data were searched by

Mascot MS/MS Ion search using Chordata as the taxonomic restriction in the SwissProt protein database, with deamidation on Gln and Asn, and oxidation on Met, as variable modifications. Since collagen was detected, further searches were carried out using COLLE database considering deamidation on Gln and Asn, oxidation on Met, hydroxylation at Pro and Lys as variable modifications. Identification details are reported in Table S10 and S11.

Wood-micro sample	Protein (Uniprot entry)	Score	Sequences	Unique Sequences	Protein sequence coverage (PSC) (%)
PNGase + Trypsin functionalized films	Ovalbumin (P01012)	26	4	2	15
	Ubiquitin (P79781)	45	2	2	20
	ATP synthase subunit beta (Q5ZLC5)	34	2	2	5
	Vitellogenin 2 (P02845)	20	4	4	2
	Collagen 1(I) (<i>Bos taurus</i>) (P02453)	101	5	0	9
	Collagen 2 (I) (<i>Bos taurus</i>) (P02465)	78	5	0	9
	Collagen 2 (I) (<i>Gallus gallus</i>) (P02460)	37	2	1	2
	Collagen 1 (III) (<i>Gallus gallus</i>) (P12105)	32	2	1	3
	Collagen 1 (I) (<i>Gallus gallus</i>) (P02457)	15	2	2	4
	Trypsin functionalized films	Ovalbumin (P01012)	69	3	3
Collagen 2 (I) (<i>Bos taurus</i>) (P02465)		63	4	0	7

Please, note that raw data were searched using in a first instance the Swissprot database with Chordata as taxonomic restriction, and when collagen was detected, since hydroxylation on prolines and lysines was considered as variable modifications, we restricted the search space to the common domesticates used for standard animal glue preparation by using a homemade database.

Conclusions

Trypsin can work very efficiently when immobilized on a cellulose acetate sheet, and digested peptides are efficiently captured on the functionalized surface. These are the major advantages of the biofunctionalized film, which we have demonstrated can work on the surfaces of different objects such as albumen and casein used as fixatives on paper, the much harder surface of ivory objects or the proteinaceous binders on a wooden Egyptian coffin. The obtained hydrolysate is definitely comparable to that which is obtained from a standard proteomic approach, and a simple MALDI-TOF analysis can allow protein identification and even species identification according to [19]. The hydrolysate can be used not only for mere identification purposes but also for characterizing the conservation state of proteins. Moreover, the functionalized film is amenable to further implementation and modularity: herein we have combined the trypsin biofunctionalized film with a PNGaseF functionalized film thus integrating the deglycosylation pretreatment to improve detection of glycosylated coated proteins. In conclusion, we demonstrated that the enzyme functionalized film is actually a versatile, user-friendly and modular tool that can be widely exploited in the world of diagnosis of cultural heritage objects.

Supporting material

<https://pubmed.ncbi.nlm.nih.gov/33147491/>

Author Contributions

G.N, L.B., P.G., D.K. G.M. designed research; G.N., D.K., I.S., P.C. carried out the experimental research; G.N., D.K., I.S., A.C., P.S., P.C., P.G., L.B. analysed data; G.N., D.K., I.S., A.C., P.C., P.G., G.M., L.B wrote the paper.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area), and by POR, Parco Archeologico Urbano di Napoli (PAUN). The authors want to thank Prof. Giancarlo Fatigati, Department of Humanities,

References

- [1] F. Colombini, Maria Perla and Modugno, ed., *Organic Mass Spectrometry in Art and Archeology*, John Wiley & Sons, 2009.
- [2] S. Dallongeville, N. Garnier, C. Rolando, C. Tokarski, *Proteins in Art, Archaeology, and Paleontology: From Detection to Identification*, *Chem. Rev.* 116 (2015) 2–79. <https://doi.org/10.1021/acs.chemrev.5b00037>.

- [3] J. Hendy, F. Welker, B. Demarchi, C. Speller, C. Warinner, M.J. Collins, A guide to ancient protein studies, *Nat. Ecol. Evol.* 2 (2018) 791–799. <https://doi.org/10.1038/s41559-018-0510-x>.
- [4] T.P. Cleland, E.R. Schroeter, A Comparison of Common Mass Spectrometry Approaches for Paleoproteomics, (2018). <https://doi.org/10.1021/acs.jproteome.7b00703>.
- [5] P. Cicatiello, G. Ntasi, M. Rossi, G. Marino, P. Giardina, L. Birolo, Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings, *Anal. Chem.* 90 (2018) 10128–10133. <https://doi.org/10.1021/acs.analchem.8b01718>.
- [6] S. Fiddyment, B. Holsinger, C. Ruzzier, A. Devine, A. Binois, U. Albarella, R. Fischer, E. Nichols, A. Curtis, E. Cheese, M.D. Teasdale, C. Checkley-Scott, S.J. Milner, K.M. Rudy, E.J. Johnson, J. Vnouček, M. Garrison, S. McGrory, D.G. Bradley, M.J. Collins, Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting., *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) 15066–71. <https://doi.org/10.1073/pnas.1512264112>.
- [7] M. Manfredi, E. Barberis, F. Gosetti, E. Conte, G. Gatti, C. Mattu, E. Robotti, G. Zilberstein, I. Koman, S. Zilberstein, E. Marengo, P.G. Righetti, Method for Noninvasive Analysis of Proteins and Small Molecules from Ancient Objects, *Anal. Chem.* 89 (2017) 3310–3317. <https://doi.org/10.1021/acs.analchem.6b03722>.
- [8] G. Zilberstein, U. Maor, E. Baskin, A. D’Amato, P.G. Righetti, Unearthing Bulgakov’s trace proteome from the Master i Margarita manuscript, *J. Proteomics.* 152 (2017) 102–108. <https://doi.org/10.1016/j.jprot.2016.10.019>.
- [9] E. Barberis, S. Baiocco, E. Conte, F. Gosetti, A. Rava, G. Zilberstein, P.G. Righetti, E. Marengo, M. Manfredi, Towards the non-invasive proteomic analysis of cultural heritage objects, *Microchem. J.* 139 (2018) 450–457. <https://doi.org/10.1016/j.microc.2018.03.033>.
- [10] C.A. Melchiorre Chiara, Dello Iorio Laura, Ntasi Georgia, Birolo Leila, Trojisi Giorgio, Cennamo Paola, Barone Lumaga M. Rosaria, Fatigati Giancarlo, Amoresano Angela, A multidisciplinary assessment to investigate a XXII Dynasty wooden coffin, *Int. J. Conserv. Sci.* 11 (2020) 25–38.
- [11] A.M. Gravagnuolo, S. Longobardi, A. Luchini, M.S. Appavou, L. De Stefano, E. Notomista, L. Paduano, P. Giardina, Class i Hydrophobin Vmh2 Adopts Atypical Mechanisms to Self-Assemble into Functional Amyloid Fibrils, *Biomacromolecules.* 17 (2016) 954–964. <https://doi.org/10.1021/acs.biomac.5b01632>.
- [12] R. Vinciguerra, E. Galano, F. Vallone, G. Greco, A. Vergara, I. Bonaduce, G. Marino, P. Pucci, A. Amoresano, L. Birolo, Deglycosylation Step to Improve the Identification of Egg Proteins in Art Samples, *Anal. Chem.* 87 (2015) 10178–10182. <https://doi.org/10.1021/acs.analchem.5b02423>.
- [13] M. Strohmalm, M. Hassman, B. Kosata, M. Kodicek, mMass data miner: an open source alternative for mass spectrometric data analysis., *Rapid Commun. Mass Spectrom.* 22 (2008) 905–908. <https://doi.org/10.1002/rcm.3444>.
- [14] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A.R. Da Fonseca, T.W. Stafford, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, M.T.P. Gilbert, J. V. Olsen, Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, *J. Proteome Res.* 11 (2012) 917–926. <https://doi.org/10.1021/pr200721u>.
- [15] R. Linn, I. Bonaduce, G. Ntasi, L. Birolo, A. Yasur-Landau, E.H. Cline, A. Nevin, A. Lluveras-Tenorio, Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings, *Angew. Chemie Int. Ed.* 57 (2018) 13257–13260. <https://doi.org/10.1002/anie.201806520>.
- [16] D.P. Kirby, N. Khandekar, J. Arslanoglu, K. Sutherland, Protein identification in artworks by peptide mass fingerprinting, in: *Prepr. ICOM Comm. Conserv. 16th Trienn. Conf. Lisbon, 2011*.
- [17] D.P. Kirby, M. Buckley, E. Promise, S.A. Trauger, T.R. Holdcraft, Identification of collagen-based materials in cultural heritage, *Analyst.* 138 (2013) 4849–4858. <https://doi.org/10.1039/c3an00925d>.
- [18] S. Kuckova, R. Hynek, M. Kodicek, MALDI-MS Applied to the Analysis of Protein Paint Binders, in: F. Colombini, Maria Perla and Modugno (Ed.), *Org. Mass Spectrom. Art Archeol.*, John Wiley & Sons, 2009: pp. 181–203.
- [19] M. Buckley, M. Collins, J. Thomas-Oates, J.C. Wilson, Species identification by analysis of bone

collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* 23 (2009) 3843–3854. <https://doi.org/10.1002/rcm.4316>.

[20] C. Solazzo, M. Wadsley, J.M. Dyer, S. Clerens, M.J. Collins, J. Plowman, Characterisation of novel α -keratin peptide markers for species identification in keratinous tissues using mass spectrometry, *Rapid Commun. Mass Spectrom.* (2013) 2685–2698. <https://doi.org/10.1002/rcm.6730>.

[21] M. Buckley, M.J. Collins, M.I. Biocentre, Collagen survival and its use for species identification in Holocene-lower Pleistocene bone fragments from British archaeological and paleontological sites, *Antiqua.* 1 (2011) 1–7. <https://doi.org/10.4081/antiqua.2011.e1>.

[22] G. Leo, L. Cartechini, P. Pucci, A. Sgamellotti, G. Marino, L. Birolo, Proteomic strategies for the identification of proteinaceous binders in paintings, *Anal. Bioanal. Chem.* 395 (2009) 2269–2280.

[23] G. Leo, I. Bonaduce, A. Andreotti, G. Marino, P. Pucci, M.P. Colombini, L. Birolo, Deamidation at asparagine and glutamine As a major modification upon deterioration/aging of proteinaceous binders in mural paintings, *Anal. Chem.* 83 (2011) 2056–2064. <https://doi.org/10.1021/ac1027275>.

[24] M. Mackie, P. R  ther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Peggie, H. Howard, L. Harrison, L.J. Jensen, J. V. Olsen, E. Cappellini, Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting, *Angew. Chemie - Int. Ed.* 57 (2018) 7369–7374. <https://doi.org/10.1002/anie.201713020>.

[25] S. Orsini, A. Yadav, M. Dilillo, L.A. McDonnell, I. Bonaduce, Characterisation of degraded proteins in paintings using bottom-up proteomic approaches: New strategies for protein digestion and analysis of data, *Anal. Chem.* 90 (2018) 6403–6408. <https://doi.org/10.1021/acs.analchem.8b00281>.

[26] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: from bones to paintings, *Microchem. J.* 126 (2015) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.

[27] S. Orsini, A. Yadav, M. Dilillo, L.A. McDonnell, I. Bonaduce, Characterization of Degraded Proteins in Paintings Using Bottom-Up Proteomic Approaches: New Strategies for Protein Digestion and Analysis of Data, *Anal. Chem.* 90 (2018) 6403–6408. <https://doi.org/10.1021/acs.analchem.8b00281>.

[28] E.R. Schroeter, T.P. Cleland, Glutamine deamidation: An indicator of antiquity, or preservational quality?, *Rapid Commun. Mass Spectrom.* 30 (2016) 251–255. <https://doi.org/10.1002/rcm.7445>.

4.3 Development and future perspective for a micro-invasive tool for the analysis of proteins in historical objects.

Introduction

In the already published results, the one-step use of trypsin functionalization film or the two-steps use of trypsin and PNGase functionalized films demonstrated successful, making possible the extraction and analysis of proteins in different ancient objects [1, 2]. Furthermore, they resulted to be most promising also for the characterization of proteins' conservation state, widening their exploitation in the field of cultural heritage.

The success of the analytical methodology is based on several factors. First of all, since trypsin is effectively immobilized by hydrophobins on the selected substrate, protein enzymatic hydrolysis is very efficient and, moreover, the enzyme is not released on the surface of the object. This latter aspect the artwork is crucial to avoid any possible damage because of the presence of the protease on the object over the time. Furthermore, it's important that with our approach, there is no need to use any denaturing agent, that might interfere with the analysis beside being an hazard for the object itself.

In our first approach, we used cellulose acetate as inert support for the immobilization of hydrophobin and trypsin. However, cellulose acetate is a relatively stiff material, and in cases where the surface of the ancient object or work of art is not smooth, impairing a perfect adhesion of the two surfaces (kit and object of study). As a consequence, we tried to improve the analytical methodology by immobilizing trypsin on different surfaces that were more flexible than cellulose acetate. We kept the same philosophy as described in the previous works, selecting among cheap and commercially available materials, in order to stay with easily accessible and easy to handle. Moreover, the inert support material should be a good compromise, being flexible but maintaining a shape. We eventually selected polyvinyl chloride (PVC), polyethylene glycol (PEG), and nitrile butadiene rubber (NBR, nitrile gloves), and experiments were performed in pictorial and historical samples in comparison with the cellulose acetate film. The efficiency of the functionalized film was roughly evaluated on the basis of the obtained protein sequence coverages. The evaluation of the object conservation state was also estimated in comparison to determine which of the inert supporting materials fits better.

Materials and methods

Pictorial & historical samples

A pictorial sample that contains animal glue and milk as proteinaceous binders and azzurite was used. The sample was kindly provided by Prof. M. P. Colombini, University of Pisa (Italy) and has been prepared as reported in [3]. "Vecchio Rifodero" is an unknown painting lining from the XIX century, kindly provided by Dr. P. Somma from Laboratorio di Restauro Tele, University Suor Orsola Benincasa, Naples (Italy).

Hydrophobin Production.

The protein Vmh2 was extracted from the mycelium of *Pleurotus ostreatus* as reported by Gravagnuolo et al [4].

Enzyme immobilization & enzymatic hydrolysis.

Trypsin was immobilized on Vmh2 coated cellulose acetate surface as previously reported [2]. In the cases of polyvinyl and poly ethylene, the experiments were performed as previously reported. In the case of nitrile butadiene rubber, the analyst wore the glove where trypsin had been immobilized over another sterile glove and touched the sample for 10 minutes with the modified finger. After ten minutes, the recovered peptides were extracted from the glove surface and analyzed by LC-MS/MS, with a procedure similar to that set up for the other functionalized films [1, 2].

Mass spectrometric analyses.

For LC-MS/MS analysis, the recovered peptides were filtered on 0.22 μ m PVDF membrane (Millipore), and peptides were desalted and concentrated on in-house made C18 extraction stage tips [5]. To increase peptides recovery, the unbound fraction from the stage tips was loaded on a reverse-phase C18 Zip Tip pipet tip (Millipore). Peptides were eluted with 20 μ L of a solution made of 50% acetonitrile, 50% formic acid 0.1% in Milli-Q water and combined with the eluted fraction from stage tips, dried under vacuum, and finally dissolved in 0.1% of formic acid. LC-MS/MS analyses were carried out on a 6520 Accurate-Mass Q-TOF LC/MS System (Agilent Technologies, Palo Alto, CA) equipped with a 1200 HPLC system and a chip cube (Agilent Technologies) and on a LTQ Orbitrap-XL (ThermoScientific, Bremen, Germany) as reported in [6], and raw data analyses as reported in [1]. Each LC-MS/MS analysis was preceded and followed by blank runs to avoid carryover contamination. MS/MS spectra were transformed in Mascot Generic files (.mgf) format and routinely used to query the SwissProt database 2015_04 (548208 sequences; 195282524 residues), with Chordata as the taxonomy restriction for protein identification. A licensed version of MASCOT software

(www.matrixscience.com) version 2.4.0 was used. Standard parameters in the searches were trypsin as the enzyme; 3, as allowed number of missed cleavages; 10 ppm MS tolerance and 0.6 Da MS/MS tolerance; peptide charge from 2+ to 3+. In all the database searches, no fixed chemical modification was inserted, but possible oxidation of methionine residues and deamidation at asparagines and glutamines were considered as variable modifications. When collagen proteins were identified, further identification runs were carried out with the insertion of hydroxylation on lysine and proline as variable modifications, since more confident identifications are commonly obtained for these proteins by taking into consideration their extensive posttranslational modifications. Moreover, to reduce the search space, these runs were carried out using a homemade database, which we named COLLE (60 sequences; 88859 residues), that collects all the common domesticates used for animal glue. Other parameter changes for the specific database searches are indicated in the Table captions. Only proteins presenting two or more peptides were considered as positively identified. Protein scores are derived from ions scores as a non- probabilistic basis for ranking protein hits. (http://www.matrixscience.com/help/interpretation_help.html) Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed peptide match is a random event. Individual ion score threshold provided by MASCOT software to evaluate the quality of matches in MS/MS data was used for confidence threshold in protein identification. Spectra with individual ion score <15 were not considered.

Calculation of deamidation levels (N,Q).

Samples were re-analyzed , using using the MaxQuant (MQ) software [7] with a handmade database (1674 sequences, 861333 residues) that combines the database COLLE with casein protein sequences downloaded from the Uniprot platform (<https://www.uniprot.org/>). No fixed chemical modification was inserted but hydroxylation of proline and lysine, deamidation of glutamine and asparagine, oxidation of methionine, were considered as variable modifications; ion score cut off was set ≥ 15 for unmodified and modified peptides. Deamidation rates for Asparagine (N) and Glutamine (Q) for individual samples was semi-quantitatively evaluated as variable modifications from MaxQuant's "evidence.txt" file with a code that is freely available on GitHub (<https://github.com/dblyon/deamidation>). Semi quantitative evaluation is based on calculation of all modification states of all identified peptides that contain N and Q [8].

Results and Discussion

As we have already demonstrated [1, 2], immobilized trypsin can work very efficiently in hydrophobins functionalized surfaces such as cellulose acetate. In a further development, we aim to immobilize trypsin in more flexible polymer materials such as polyethylene glycol, polyvinyl and nitrile butadiene rubber (nitrile

gloves), in order to produce a film that would better adapt to irregular surfaces. Protein sequence coverage was used as a qualitative criterion of film efficiency, using as benchmark the cellulose acetate film. Table 1 reports the proteins identified in the pictorial sample upon digestion with the four aforementioned functionalized films. Collagen and casein proteins were found in all cases but with very different protein sequence coverages. All the new functionalized films performed better than cellulose acetate, confirming that a more flexible film facilitate protein digestion, possibly because of an increased adhesion to the sample surface. Among the new inert supports, polyvinyl chloride (PVC) and nitrile butadiene rubber (NBR or nitrile gloves) gave the best results. Interestingly, the use of nitrile gloves provided the best kappa and beta casein identifications. Notably, quite hydrophobic and long peptides were identified (See table S3). Thereof, the choice of the inert support can play a key role in the exploitation of the film.

It is quite known that hydrophobins once immobilized are self-assembled and can turn a hydrophobic surface to hydrophilic and vice versa [4,9,10]. This is accomplished by an extensive formation of bonds between hydrophobins and the hydrophilic group of the surface, although the whole mechanism has not been fully understood yet and appears to be quite complex [4,9,10]. Surface wettability, calculated by measuring the water angle via the Young equation, is used to verify the immobilization of hydrophobin. An angle value higher than 90 degrees demonstrates the molecule immobilization, due to the change of polymer hydrophobicity. Most importantly, the nature of the inert support determines whether hydrophobin will expose their hydrophilic or hydrophobic side to the surface. In case of the herein tested polymers, they expose an hydrophilic surface, thus leading to bind other proteins such as trypsin via hydrophilic interactions. To have a deeper insight into how hydrophobicity of the polymer-based support changed after the Vmh2 immobilization, we performed a simple experiment. In particular, a photo of a spotted water drop was captured by a 38MP camera before and after the Vmh2 immobilization in the four new developed functionalized films. Figure 1 shows the 90 degrees of the water drop before Vmh2 immobilization, demonstrating the poor wetting of the supports and thus their hydrophobicity. It also shows the increase of the water drop angle after the immobilization of Vmh2. The bigger is the difference before and after Vmh2, the more the hydrophilicity of the supports has been increased, the more efficient the immobilization of trypsin.

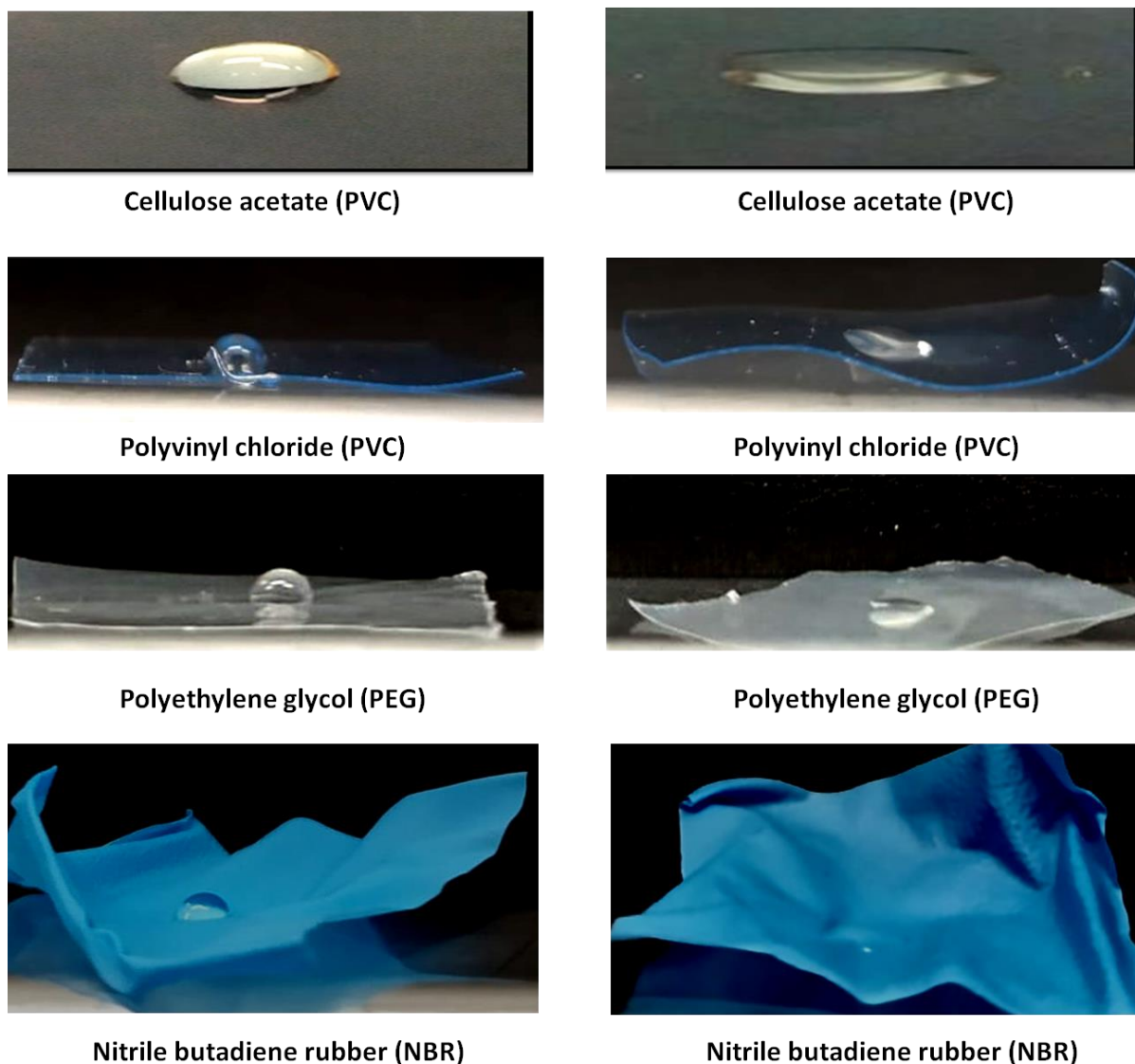


Figure 1: Water drop angle capture before (left) and after of Vmh2 immobilization (right).

We already demonstrated in case of the cellulose acetate that LC-MS/MS data can be used to search for chemical modifications that can be considered as an hallmark to characterize the conservation state of proteins beside a simple protein identification [11]. For instance, calculation of the deamidation (N,Q) levels can provide important information regarding protein degradation in ancient material. Here, we used the same bioinformatic tool [7] to calculate the deamidation levels of both the pictorial sample and the historical canvas with each polymer-based functionalized film to investigate if there is any difference in the results. Figures 2 and 3 show the deamidation values that were calculated in each case, with the error bars that truly reflect the deamidation variation in each sample. Interestingly, deamidation values are quite similar within the single sample, but slightly smaller in case of nitrile butadiene rubber polymer material (nitrile gloves) and

the polyvinyl chloride substrate. Taking into account that the number of glutamines covered is bigger (118 in case of nitrile gloves and 113 in case of the polyvinyl chloride (PEG)), the error bars are smaller, making the statistics more robust and reliable.

The same conclusion can be drawn for the results obtained on the historical canvas, demonstrating that the use of these new inert supports might also be advantageous for the analysis of deamidation, and possibly for all the degradation processes. Subsequently, we investigated if the extraction of more peptides with nitrile butadiene rubber (nitrile gloves) and the polyvinyl chloride (PVC) supports could contribute to the study of other chemical modifications. As we have already demonstrated, glycation products are frequently observed in the surfaces of the historical paintings, possibly arising from the covalent binding of reducing sugars of gums to proteins in the complex mixture of the painting media [1].

For this reason, a wide set of AGEs were set as variable modifications in the basic MS Ions search program in Mascot software: glycation of arginine with the formation of glyoxal and methylglyoxal adduct (mass shift of + 39.99 and +54.01 Da respectively), as well as formation of carboxymethyl and carboxyethyl lysine (with mass shift of +72.01 and +58.01 Da, respectively). Table 3 and table S9 illustrate a comparative study regarding the identification of advanced glycation end-products (AGEs) of lysine and arginine in the historical canvas "Vecchio Rifodero" with the use of the four different inert supports. Based on the results, the use of nitrile butadiene rubber polymer material (nitrile gloves), provided the better AGEs detection, with glycation peptides with the highest scores.

In conclusion, we suggest that the nitrile butadiene rubber (NBR or nitrile gloves) is the best inert support that can be used also to gain information on chemical modifications of proteins inside an artistic object, thus allowing to gain insights on its degradation state and not only to merely identify the protein component. However, the kits have to be tested also in other samples in order to further validate the results herein presented.

As future prospective, therefore and moreover, an actual exploitation of these inert supports on samples other than paintings is also envisaged, as well as the combination of N-glycosidase F and trypsin in the use of nitrile gloves has to be tested in the case of egg containing samples [1]. Also, the study of the mechanism of interaction between hydrophobins and trypsin.

Most importantly, an actual exploitation of the different functionalized films in a series of study cases is foreseen in the future.

Table 1: Identification of Proteins in pictorial sample contained animal glue and casein by LC-MSMS. Samples were subjected to *in situ* digestion with the different developed trypsin functionalized films. Raw data were searched by Mascot MS/MS Ion search using the homemade COLLE database in case of collagen and Swissprot with Chordata as taxonomic restriction in case of casein proteins. Deamidation on Gln and Asn, hydroxylation at Pro and Lys, oxidation on Met, were considered as variable modifications. Identification details are reported in Tables S1-S4.PSC; Protein sequence coverage

Samples	Proteins (Accession number)	Polyvinyl chloride (PVC)		Polyethylene glycol (PEG)		Cellulose acetate		Nitrile butadiene rubber (Nitrile gloves)	
		Number of identified peptides	PSC %	Number of identified peptides	PSC %	Number of identified peptides	PSC %	Number of identified peptides	PSC %
Animal glue + casein	Collagen alpha1(I) (P02453)	49	53%	32	40%	39	40%	47	51%
	Collagen alpha2(I) (P02465)	46	51%	30	39%	37	46%	41	49%
	Collagen alpha 1(III) (P02458)	4	5%	4	6%	-	-	4	10%
	Alpha-S1 casein (P02662)	21	74%	11	44%	10	35%	16	75%
	Alpha-S2 casein (P02663)	13	45%	9	32%	4	18%	12	42%
	Beta-casein (P02666)	9	20%	2	5%	3	8	10	43%
	Kappa-casein (P02668)	10	50%	5	26%	3	18	6	74%

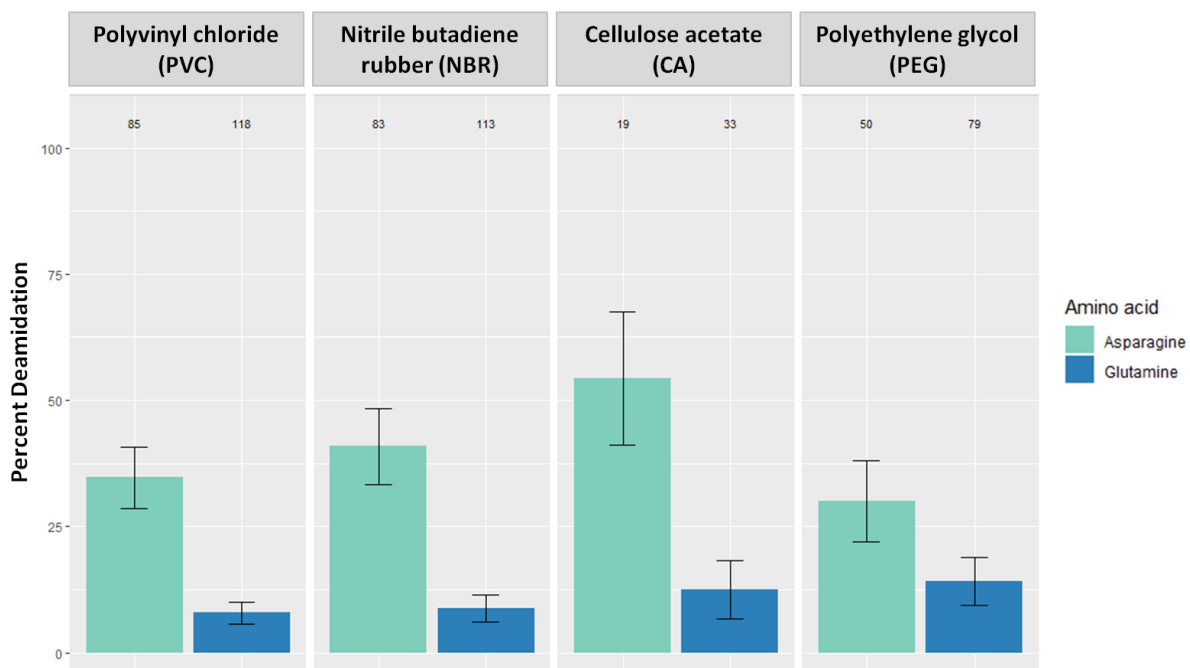


Figure 2: Overall percentage of deamidation for asparagine (N) and Glutamine (Q) residues for the proteins identified on the surface of pictorial sample contained animal glue and casein detected with the several trypsin functionalized films. Error bars represent standard deviation and numbers above the bars represent the number of peptides containing Asn and Gln on which the data was based on.

Table 2: Identification of Proteins in the historical canvas "Vecchio Rifodero" by LC-MSMS. Samples were subjected to *in situ* digestion with the several developed trypsin functionalized films. Raw data were searched by Mascot MS/MS Ion search using the homemade COLLE database. Deamidation on Gln and Asn, hydroxylation at Pro and Lys, oxidation on Met, were considered as variable modifications. Identification details are reported in Tables S5-S8. .PSC; Protein sequence coverage.

Samples	Proteins (Accession number)	Polyvinyl chloride (PVC)		Polyethylene glycol (PEG)		Cellulose acetate		Nitrile butadiene rubber (Nitrile gloves)	
		Number of identified peptides	PSC %	Number of identified peptides	PSC %	Number of identified peptides	PSC %	Number of identified peptides	PSC %
Historical canvas "Vecchio Rifodero"	Collagen alpha-1(I) (P02453)	30	34%	23	32%	14	20%	41	46%
	Collagen alpha-2(I) (P02465)	27	40%	24	37%	17	26%	28	38%
	Collagen alpha-1(III) (P02458)	7	9%	10	12%	6	3%	16	18%

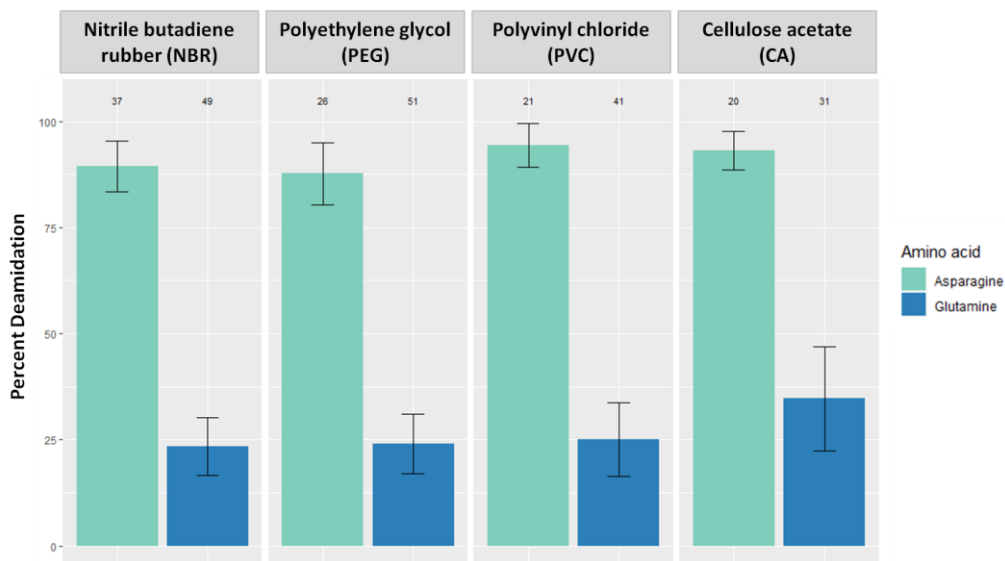


Figure 3: Overall percentage of deamidation for asparagine (N) and Glutamine (Q) residues for the proteins identified on the surface of the historical canvas "Vecchio Rifodero" detected with the several trypsin functionalized films. Error bars represent standard deviation and numbers above the bars represent the number of peptides containing Asn and Gln on which the data was based on.

Table 3: Glycation products in the historical canvas "Vecchio rifodero" digestion with the four different functionalized films. LC-MSMS raw data from *in situ* digestion with trypsin functionalized film were searched by Mascot MS/MS Ion search using COLLE database considering semitrypsin as the enzyme, deamidation on Gln and Asn, hydroxylation at Pro, carboxyethyl(K), carboxymethyl(K), Glyoxal(R) and methylglyoxal (R) as variable modifications. Individual ion score threshold provided by MASCOT software (>10) to evaluate the quality of matches in MS/MS data was used for confidence threshold in protein identification. The type of each modification is reported in the table S9.

Protein name (Entry)	Sum of the identified peptides			
	Cellulose acetate	Polyethylene glycol (PEG)	Polyvinyl chloride (PVC)	Nitrile butadiene rubber (Nitrile gloves)
Collagen, type I, alpha 1 [Bos taurus] (AAI05185.1)	3	1	4	6
Collagen, type I, alpha 2 [Bos taurus] (AAI49096.1)	1	1	2	3
Collagen, type III, alpha 1 [Bos taurus] (AAI23470.1)	2	0	3	4

Supplementary material

Tables S1-S4: Protein identification of the pictorial sample containing animal glue and milk after *in situ* digestion with the four developed trypsin functionalized films and then analyzed by LC-MSMS. MS/MS spectra were searched by Mascot MS/MS Ion search using the homemade COLLE database in case of collagen and Chordata as the taxonomic restriction in the SwissProt protein database in case of casein. Deamidation of Gln and Asn, oxidation on Met, pyro- Glu formation at Gln at the N-terminus were set as variable modifications were no fixed modifications were inserted.

Table S1: Protein identification of the pictorial sample containing animal glue and milk after *in situ* digestion with the use of the polyvinyl chloride (PVC) as substrate.

Identified protein (Uniprot entry)	<i>m/z</i>	Peptide score	Peptide
COL1A1 (P02453)	314.676	22	R.GLPGER.G
	392.223	42	R.GAAGLPGPK.G + Hydroxylation (P) (P)
	418.726	28	R.GPAGPQGPR.G
	426.22	35	R.GFSGLDGAK.G
	449.76	34	R.GVVGLPGQR.G + Hydroxylation (P) (P)
	360.53	24	R.GRPGAPGAGAR.G + Hydroxylation (P) (P)
	544.775	59	R.GFPGADGVAGPK.G + Hydroxylation (P) (P)
	553.79	37	R.GVQGGPPGAGPR.G + Deamidated (NQ); Hydroxylation (P) (P)
	565.80	35	R.GLPGTAGLPGMK.G + 2 Hydroxylation (P) (P)
	589.78	58	R.GQAGVMGFPGPK.G + Deamidated (NQ); Hydroxylation (P) (P); Oxidation (M)
	596.82	26	R.GVPGPPGAVGPAGK.D + 2 Hydroxylation (P) (P)
	414.562	31	R.GVVGLPGQRGER.G + Deamidated (NQ); Hydroxylation (P) (P)
	629.805	63	K.GLTGSPGSPGPDGK.T + 2 Hydroxylation (P) (P)
	667.33	66	R.GPSGPQQGSPGPPGPK.G + Deamidated (NQ); Hydroxylation (P) (P)
	672.830	31	R.GFPGLPGSPGEPGK.Q + 3 Hydroxylation (P) (P)
	718.35	60	R.GEPGPAGLPGPPGER.G + 3 Hydroxylation (P) (P)
	730.35	62	R.GSAGPPGATGFPGAAGR.V + 2 Hydroxylation (P) (P)
	748.89	21	K.STGISVPGPMGSPGPR.G
	529.594	44	K.GANGAPGIAGAPGFPGAR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	803.92	30	R.GLTGPIGPPGAGAPGDK.G + 3 Hydroxylation (P) (P)
	552.612	59	R.GFPGADGVAGPKGPAGER.G + Hydroxylation (P) (P)
	932.445	29	K.GEPGPTGIQGGPPGAGEEGK.R + 2 Hydroxylation (P) (P)
	654.99	16	R.GEPGPAGLPGPPGERGGPGSR.G + 4 Hydroxylation (P) (P)
	659.34	61	K.SGDRGETGPAGPAGPIGPVGAR.G
	674.00	52	K.GEPGPTGIQGGPPGAGEEGK.R + 2 Hydroxylation (P) (P)
	686.34	35	K.TGPPGPAGQDGRGPPGPPGAR.G + 4 Hydroxylation (P) (P)
	691.68	41	K.GAPGADGPAGAPGTPGPGQIAGQR.G + Hydroxylation (P) (P)
	546.518	15	R.GGPGSRGFPGADGVAGPKGPAGER.G + 2 Hydroxylation (P) (P)
	733.668	33	K.GDAGAPGAPGSQGAPGLQGMPGER.G + 4 Hydroxylation (P) (P)
	739.362	20	R.GETGPAGRPEVGGPPGPPGAGEK.G + 3 Hydroxylation (P)

			(P)
	569.53	29	R.GQAGVMGFPGPKGAAGEPGKAGER.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	769.722	17	K.GDAGPPGAPGAPGPPGPIGNVGVGAPGPK.G + 3 Hydroxylation (P) (P)
	832.037	55	K.GDAGPAGPKGEPGSPGENGAPGQMGR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	838.409	39	K.GDRGETGPAGPPGAPGAPGAPGVPAGK.S + 4 Hydroxylation (P) (P)
	850.41	41	R.GNDGATGAAGPPGPTGPAGPPGFPGAVGAK.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	638.058	30	K.GDTGAKGEPGPTGIQPPGAGEEGR.G + 2 Hydroxylation (P) (P)
	901.755	50	R.GAPDRGEPGPPGAPGAPGADGQPGAK.G + 4 Hydroxylation (P) (P)
	714.1154	60	R.GLTGPIGPPGAPGAPGDKGEAGSPGAPPTGAR.G + 2 Hydroxylation (P) (P)
	955.4680	25	R.GVPGPPGAVGPAGKDGEAGAQQPPGAPGAGER.G + 2 Hydroxylation (P) (P)
	1062.820	28	R.GSEGPPQVVRGEPGPPGAPGAAAGPAGNPGADGQPGAK.G + 3 Deamidated (NQ); 4 Hydroxylation (P) (P)
	807.353	36	R.GANGAPGNDGAKGDAGAPGAPGSQAPGLQGMPPER.G + 2 Deamidated (NQ); 5 Hydroxylation (P) (P)
	836.630	30	R.GFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGR.G + 2 Deamidated (NQ); 3 Hydroxylation (P) (P); Oxidation (M)
	907.20	22	R.GRPGAPGAPGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAK.G + Deamidated (NQ); 5 Hydroxylation (P) (P)
COL1A2 (P02465)	379.70	34	R.GLPGADGR.A + Hydroxylation (P) (P)
	393.22	17	R.GATGPAGVR.G
	421.23	47	R.GVVGPPQGAR.G + Deamidated (NQ)
	434.739	34	R.VGAPGPAGAR.G + Hydroxylation (P) (P)
	434.739	32	R.GPSGPQGIR.G
	451.732	54	R.AGVMGPAGSR.G
	361.20	20	R.GRVGAPGPAGAR.G + Hydroxylation (P) (P)
	601.30	35	R.GEPGNIGFPGPK.G + 2 Hydroxylation (P) (P)
	605.336	43	R.IGQPGAVGPAGIR.G + Deamidated (NQ); Hydroxylation (P) (P)
	611.81	17	R.GFPGTPGLPGFK.G + 3 Hydroxylation (P) (P)
	408.546	41	R.GPAGPSGPAGKDGR.I
	631.322	47	R.GEAGPAGPAGPAGPR.G
	634.34	64	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)
	644.818	52	R.GFPGSPGNIGPAGK.E + Deamidated (NQ); 2 Hydroxylation (P) (P)
	714.372	51	R.GIPGEFGLPGPAGAR.G + 2 Hydroxylation (P) (P)
	737.344	75	R.GDGGPPGATGFPGAAGR.T + 2 Hydroxylation (P) (P)
	511.602	26	R.GEPGPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)

	769.869	32	R.GAPGAIGAPGPAGANGDR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	388.462	16	K.EGLRGPRGDQGPVGR.S
	521.619	44	K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) (P)
	790.890	74	R.GPPGESGAAGPTGPIGSR.G + Hydroxylation (P) (P)
	557.637	28	R.GAAGIPGGKGEKGETGLR.G + Hydroxylation (P) (P)
	595.275	27	R.GPNGDSGRPGEPGLMGPR.G + Deamidated (NQ); Hydroxylation (P) (P); Oxidation (M)
	610.644	22	R.TGPPGPSGISGPPGPPGAGK.E + 3 Hydroxylation (P) (P)
	641.650	36	R.GERPPGESGAAGPTGPIGSR.G + Hydroxylation (P) (P)
	500.011	58	K.HGNRGEPPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)
	686.041	33	K.EGPVGLPGIDGRPGPIGPAGAR.G + Hydroxylation (P) (P)
	1066.069	79	R.GLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	745.370	31	R.GSTGEIGPAGPPGPPGLRGNPGSR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	754.386	40	R.GYPGNAGPVGAAGAPGPQGPVGPVVK.H + 2 Hydroxylation (P) (P)
	798.726	38	R.GEVGPAGPNGFAGPAGAAGQPGAKGER.G + Deamidated (NQ); Hydroxylation (P) (P)
	811.083	31	R.GEVGLPGLSGVPVPPGNPANGLPK.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	861.412	52	K.GENGPVGPVGAAGPSGPNPPGAGSR.G + Deamidated (NQ); Hydroxylation (P) (P)
	863.76	37	R.GSDGSVGPVGPAGPIGSAGPPGFPAGPK.G + 3 Hydroxylation (P) (P)
	696.092	62	R.GAPGAIGAPGPAGANGDRGEAGPAGPAGPR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	931.452	30	K.GEQGPAGPPGFQGLPGPAGTAGEAGKPGER.G + 3 Hydroxylation (P) (P)
	955.468	30	K.GPKGENGPVGPVGAAGPSGPNPPGAGSR.G + Deamidated (NQ); Hydroxylation (P) (P)
	749.111	62	R.GSPGERGEVGPAGPNGFAGPAGAAGQPGAKGER.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	1011.841	33	K.GPSGEPGTAGPPGTPGQGLLAPGFLGLPGSR.G + 5 Hydroxylation (P) (P)
	1076.841	27	R.GPSGPPGPDGNKGEVGVGAPGTAGPSGSLPGER.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	1125.899	22	K.GPSGEPGTAGPPGTPGQGLLAPGFLGLPGSR.G + 5 Hydroxylation (P) (P)
1127.554	35	R.SGETGASGPPGFVGEKGPSGEPGTAGPPGTPGQGLLAPGFLGLPGSR.G + 6 Hydroxylation (P) (P)	
COL3A1 (P02458)	371.20	31	R.GRPGLGAAGAR.G + 2 Hydroxylation (P) (P)
	564.779	40	R.GLAGPPGMPGAR.G + 2 Hydroxylation (P) (P); Oxidation (M)
	637.304	21	R.GSPGGPGAAGFPGGR.G + 2 Hydroxylation (P) (P)
	565.297	27	R.GVAGEPGRDGLPGGPGLR.G + 2 Hydroxylation (P) (P)

Identified protein (Uniprot entry)	m/z	Peptide score	Peptide
Alpha-S1-casein Bos taurus (P02662)	382.687	17	K.TTMPLW.- + Oxidation (M)
	316.181	22	K.EKVNELSK.D
	634.853	47	R.YLGYLEQLLR.L + Deamidated (NQ)
	669.347	40	K.HIQKEDVPSER.Y
	692.875	43	R.FFVAPFPEVFGK.E
	503.635	21	R.YLGYLEQLLRK.K
	547.965	41	R.FFVAPFPEVFGKEK.V
	587.325	63	K.HQGLPQEVLENLLR.F
	651.328	45	K.YKVPQLEIVPNSAEER.L + Phospho (ST)
	500.778	38	K.KYKVPQLEIVPNSAEER.L
	694.031	31	K.EDVPSERYLGYLEQLLR.L
	559.570	36	K.HPIKHQGLPQEVLENLLR.F
	565.034	20	K.VPQLEIVPNSAEERLHSMK.E + Phospho (ST)
	581.067	34	R.LKKYKVPQLEIVPNSAEER.L + Phospho (ST)
	1166.576	78	K.EPMIGVNQELAYFYPELFR.Q + Oxidation (M)
	637.827	25	K.YKVPQLEIVPNSAEERLHSMK.E + Phospho (ST)
	647.353	36	K.HIQKEDVPSERYLGYLEQLLR.L
	669.851	46	K.KYKVPQLEIVPNSAEERLHSMK.E + Phospho (ST)
	1069.879	65	K.EGIHAQQKEPMIGVNQELAYFYPELFR.Q
	637.491	31	R.LHSMKEGIHAQQKEPMIGVNQELAYFYPELFR.Q + Oxidation (M)
1572.742	43	R.QFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSENSEK.T	
Alpha-S2-casein Bos taurus (P02663)	373.733	25	K.VIPYVR.Y
	488.305	34	K.TKVIPYVR.Y
	490.288	36	K.FALPQYLK.T
	598.347	32	R.NAVPITPTLNR.E
	451.265	29	K.RNAVPTPTLNR.E
	684.357	51	K.ALNEINQFYQK.F
	466.934	20	R.YQKFALPQYLK.T
	733.815	59	K.TVDMESTEVFTK.K + Phospho (ST)
	545.307	21	K.LTEEEKNRLNFLK.K
	466.518	17	K.TKLTEEEKNRLNFLK.K
	717.137	39	K.FPQYLQYLYQGPIVLPWDQVQR.N
	1055.042	32	K.ALNEINQFYQKFPQYLQYLYQGPIVLPWDQVQR.N + 3 Deamidated (NQ)
Kappa-casein Bos taurus (P02668)	322.160	26	R.FFSDK.I
	319.183	27	R.FFSDKIAK.Y
	626.364	20	K.YIPIQYVLSR.Y
	402.970	32	R.HPHPHLSFMAIPPK.K
	660.708	63	R.SPAQILQWQVLSNTVPAK.S

	730.078	34	R.FFSDKIAKYIPIQYVLSR.Y
	809.407	31	R.SPAQILQWQVLSNTVPAKSCQAQPTTMAR.H + Phospho (ST)
	1003.279	50	R.YPSYGLNYYQQKPVALINNQFLPYPYYAKPAAVR.S
	1311.454	26	K.YIPIQYVLSRPSYGLNYYQQKPVALINNQFLPYPYYAKPAAVR.S
	883.899	19	R.FFSDKIAKYIPIQYVLSRPSYGLNYYQQKPVALINNQFLPYPYYAKPAAVR.S + 2 Deamidated (NQ)
Beta-casein Bos taurus (P02666)	371.730	18	R.GPFPIIV.-
	382.687	35	K.EMPFPK.Y + Oxidation (M)
	390.755	32	K.VLPVPQK.A
	415.732	21	K.AVPYPQR.D
	531.317	23	K.VLPVPQKAVPYPQR.D
	1093.596	68	R.DMPIQAFLLYQEPVLPVLR.G
	1463.306	63	R.DMPIQAFLLYQEPVLPVLRGPFPIIV.- + Oxidation (M)
	1005.212	37	K.AVPYPQRDMPIQAFLLYQEPVLPVLR.G + Hydroxylation (P) (P)
	1241.024	29	K.AVPYPQRDMPIQAFLLYQEPVLPVLRGPFPIIV.-

Table S2: Protein identification of the pictorial sample containing animal glue and milk after *in situ* digestion with the use of the polyethylene glycol polymer material as substrate.

Identified protein (Uniprot entry)	m/z	Peptide score	Peptide
COL1A2 (P02465)	597.338	45	R.IGQPGAVGPAGIR.G + Deamidated (NQ)
	601.297	45	R.GEPGNIGFPGPK.G + 2 Hydroxylation (P) (P)
	603.817	33	R.GFPGTPGLPGFK.G + 2 Hydroxylation (P) (P)
	634.346	53	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)
	644.815	78	R.GFPGSPGNIGPAGK.E + Deamidated (NQ); 2 Hydroxylation (P) (P)
	714.373	72	R.GIPGEFGLPGPAGAR.G + 2 Hydroxylation (P) (P)
	729.343	26	R.GDGGPPGATGFPGAAGR.T + Hydroxylation (P) (P)
	766.901	40	R.GEPGPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)
	769.869	29	R.GAPGAIGAPGPAGANGDR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	521.618	44	K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) (P)
	808.420	66	K.GELGPVGNPGPAGPAGPR.G + Hydroxylation (P) (P)
	602.318	43	K.RGSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	500.258	48	K.HGNRGEPPGAVGPAGAVGPR.G + Deamidated (NQ); Hydroxylation (P) (P)
	684.668	45	R.GEVGPAGPNGFAGPAGAAGQPGAK.G + Deamidated (NQ); Hydroxylation (P) (P)
	686.040	26	K.EGPVGLPGIDGRPGPIGPAGAR.G + Hydroxylation (P) (P)
1066.068	88	R.GLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P)	

			(P)
	745.370	33	R.GSTGEIGPAGPPGPPGLRGNGPSR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	754.387	49	R.GYPGNAGPVGAAGAPGPQGPVGPVK.H + 2 Hydroxylation (P) (P)
	798.726	43	R.GEVGPAGPNGFAGPAGAAGQPGAKGER.G + Deamidated (NQ); Hydroxylation (P) (P)
	811.079	48	R.GEVGLPGLSGPVGPPGNPGANGLPGAK.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	861.411	53	K.GENGPVGTGPVGAAGPSGPNPPGAGSR.G + Deamidated (NQ); Hydroxylation (P) (P)
	863.760	31	R.GSDGSVGPVGPAGPIGSAGPPGFPGAPGPK.G + 3 Hydroxylation (P) (P)
	696.093	52	R.GAPGAIGAPGPAGANGDRGEAGPAGPAGPR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	931.450	29	K.GEQGPAGPPGFQGLPGPAGTAGEAGKPGER.G + 3 Hydroxylation (P) (P)
	1011.838	40	K.GPSGEPGTAGPPGTPGQQLLGPAGFLGLPSR.G + 5 Hydroxylation (P) (P)
	1076.518	28	R.GPSGPPGPDGNKGEVGVGAPGTAGPSGSLPGER.G + 4 Hydroxylation (P) (P)
	1127.556	21	R.SGETGASGPPGFVGEKGPSGEPGTAGPPGTPGQQLLGPAGFLGLPSR.G + 6 Hydroxylation (P) (P)
COL1A1 (P02453)	442.255	25	R.GVVGLPGQR.G + Deamidated (NQ)
	544.777	53	R.GFPGADGVAGPK.G + Hydroxylation (P) (P)
	553.293	26	R.GVQGPAGPR.G + Hydroxylation (P) (P)
	565.799	39	R.GLPGTAGLPGMK.G + Hydroxylation (P) (P); Oxidation (M)
	589.784	59	R.GQAGVMGFPGPK.G + Deamidated (NQ); Hydroxylation (P) (P); Oxidation (M)
	672.830	38	R.GFPGLPGSGEPGK.Q + 3 Hydroxylation (P) (P)
	718.351	43	R.GEPGPAGLPGPPGER.G + 3 Hydroxylation (P) (P)
	487.238	16	R.GSAGPPGATGFPGAAGR.V + 2 Hydroxylation (P) (P)
	748.884	17	K.STGISVPGPMGSPGR.G
	780.916	28	R.GETGPAGPAGPIGPVGAR.G
	793.395	54	K.GANGAPGIAGAPGFPGAR.G + 3 Hydroxylation (P) (P)
	552.613	38	R.GFPGADGVAGPKGPAGER.G + Hydroxylation (P) (P)
	606.297	53	R.GPPGPMGPPGLAGPPGESGR.E + Hydroxylation (P) (P); Oxidation (M)
	932.446	44	K.GEPGPTGIQPPGPAGEEGK.R + 2 Hydroxylation (P) (P)
	659.340	63	K.SGDRGETGPAGPAGPIGPVGAR.G
	673.995	56	K.GEPGPTGIQPPGPAGEEGKR.G + 2 Hydroxylation (P) (P)
	697.010	45	K.GAPGADGPAGAPGTPGQGIAGQR.G + 2 Hydroxylation (P) (P)
	723.693	24	R.GETGPAGPPGAPGAPGAPGPVGPAGK.S + 3 Hydroxylation (P) (P)
	1099.983	43	K.GDAGAPGAPGSQGAPGLQGMPGER.G + 4 Hydroxylation

			(P) (P)
	752.367	21	R.GFPGLPGSPGEPGKQGPSGASGER.G + 2 Hydroxylation (P) (P)
	770.049	27	K.GDAGPPGPAGPAGPPGPIGNVVGAPGPK.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	819.085	31	R.GPPGSAGSPGKDGLNGLPGPIGPPGPR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	838.400	39	K.GDRGETGPAGPPGAPGAPGAPGVPAGK.S + 4 Hydroxylation (P) (P)
	850.409	52	R.GNDGATGAAGPPGPTGPAGPPGFPAGVAVGAK.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	901.757	54	R.GAPGDRGEPGPPGAGFAGPPGADGQPGAK.G + 4 Hydroxylation (P) (P)
	951.816	49	R.GLTGPIGPPGAPGAPGDKGEAGSPGAPGPTGAR.G + 2 Hydroxylation (P) (P)
	960.807	48	R.GVPGPPGAVGPAGKDGEAGAQQPPGAPGAGER.G + 3 Hydroxylation (P) (P)
	1034.137	26	R.GLPGPPGAPGPPGQGFQPPGEPGEPGASGPMGPR.G + 5 Hydroxylation (P) (P); Oxidation (M)
	1062.818	28	R.GSEGPQGVREPPGPPGAGAAGPAGNPGADGQPGAK.G + 3 Deamidated (NQ); 4 Hydroxylation (P) (P)
	1134.525	21	R.GNDGATGAAGPPGPTGPAGPPGFPAGVAVGAKGEGGPPGPR.G + 2 Deamidated (NQ); 4 Hydroxylation (P) (P)
	876.159	29	R.GAPGDRGEPGPPGAGFAGPPGADGQPGAKGEPGDAGAK.G + 5 Hydroxylation (P) (P)
	COL3A1 (P02458)	371.207	15
556.784		34	R.GLAGPPGMPPGAR.G + Hydroxylation (P) (P); Oxidation (M)
694.714		43	R.GAPGPPGPPGAPGPLGIAGLTGAR.G + Hydroxylation (P) (P)
1297.011		16	R.GPPGPPGNGNPPGSSGAPGKDPPGPPGNGAPGSPGISGPKGDSGPPGER.G + 12 Hydroxylation (P) (P); 4 Phospho (ST)

Identified protein (Uniprot entry)	m/z	Peptide score	Peptide
Alpha-S1-casein Bos taurus (P02662)	634.362	59	R.YLGYLEQLLR.L
	692.875	28	R.FFVAPFPEVFGK.E
	503.637	19	R.YLGYLEQLLR.LK.K
	547.966	44	R.FFVAPFPEVFGKEK.V
	830.908	40	K.VPQLEIVPNSAEER.L + Phospho (ST)
	587.325	54	K.HQGLPQEVLENLLR.F
	976.488	60	K.YKVPQLEIVPNSAEER.L + Phospho (ST)
	694.025	42	K.KYKVPQLEIVPNSAEER.L + Phospho (ST)
	559.569	45	K.HPIKHQGLPQEVLENLLR.F

	578.819	24	R.FFVAPFPEVFGKEKVNELSK.D
	806.656	45	K.EGIHAQQKEPMIGVNVQELAYFYPELFR.Q + Hydroxylation (P) (P)
Alpha-S2-casein Bos taurus (P02663)	317.702	21	R.LNFLK.K
	325.873	31	K.TKVIPYVR.Y
	490.287	25	K.FALPQYLK.T
	366.877	27	K.AMKPWIQPK.T
	399.234	30	R.NAVPITPTLNR.E
	451.266	33	K.RNAVPTPTLNR.E
	684.356	37	K.ALNEINQFYQK.F
	733.815	59	K.TVDMESTEVFTK.K + Phospho (ST)
	545.305	16	K.LTEEEKNRLNFLK.K
Kappa-casein Bos taurus (P02668)	319.181	24	R.FFSDKIAK.Y
	626.362	40	K.YIPIQYVLSR.Y
	402.969	40	R.HPHPHLSFMAIPPK.K
	348.195	19	R.HPHPHLSFMAIPPK.N
	660.707	62	R.SPAQILQWQVLSNTVPAK.S
Beta-casein Bos taurus (P02666)	371.730	24	R.GPFPIIV.-
	374.692	29	K.EMPFPK.Y

Table S3: Protein identification of the pictorial sample containing animal glue and milk after *in situ* digestion with the use of the nitrile butadiene rubber (NBR or nitrile gloves) as substrate.

Identified Protein (Uniprot entry)	<i>m/z</i>	Peptide score	Peptide
COL1A1 (P02453)	392.222	39	R.GAAGLPGPK.G + Hydroxylation (P) (P)
	426.221	52	R.GFSGLDGAK.G
	449.759	34	R.GVVGLPGQR.G + Hydroxylation (P) (P)
	544.77	50	R.GFPGADGVAGPK.G + Hydroxylation (P) (P)
	553.29	38	R.GVQGGPPGAGPR.G + Hydroxylation (P) (P)
	565.800	33	R.GLPGTAGLPGMK.G + Hydroxylation (P) (P); Oxidation (M)
	581.291	70	R.GQAGVMGFPGPK.G + Hydroxylation (P) (P)
	596.816	31	R.GVPGPPGAVGPAGK.D + 2 Hydroxylation (P) (P)
	414.56	22	R.GVVGLPGQRGER.G + Hydroxylation (P) (P); Deamidated (NQ)

629.803	52	K.GLTGSPGSPGPDGK.T + 2 Hydroxylation (P) (P)
664.829	38	R.GFPGLPGSPGEPGK.Q + 2 Hydroxylation (P) (P)
666.83	62	R.GPSGPQGPSPPGPK.G + Hydroxylation (P) (P)
718.344	51	R.GEPGPAGLPGPPGER.G + 3 Hydroxylation (P) (P)
730.350	60	R.GSAGPPGATGFPGAAGR.V + 2 Hydroxylation (P) (P)
748.882	24	K.STGISVPGPMGSPGPR.G
793.885	60	K.GANGAPGIAGAPGFPGAR.G + 3 Hydroxylation (P) (P); Deamidated (NQ)
803.91	36	R.GLTGPIGPPGPAGAPGDK.G + 3 Hydroxylation (P) (P)
552.608	38	R.GFPGADGVAGPKGPAGER.G + Hydroxylation (P) (P)
611.625	29	R.GPPGPMGPPGLAGPPGESGR.E + 2 Hydroxylation (P) (P); Oxidation (M)
932.94	44	K.GEPGPTGIQPPGPAGEEGK.R + 2 Hydroxylation (P) (P); Deamidated (NQ)
659.339	59	K.SGDRGETGPAGPAGPIGPGAR.G
674.329	29	K.GEPGPTGIQPPGPAGEEGKR.G + 2 Hydroxylation (P) (P); Deamidated (NQ)
686.333	40	K.TGPPGPAGQDGRPGPPGPPGAR.G + 4 Hydroxylation (P) (P)
697.005	33	K.GAPGADGPAGAPGTPGQGIAGQR.G + 2 Hydroxylation (P) (P)
718.36	25	R.GETGPAGPPGAPGAPGAPGVPAGK.S + 2 Hydroxylation (P) (P)
1099.990	56	K.GDAGAPGAPGSQGAPGLQGMPGER.G + 4 Hydroxylation (P) (P)
739.3584	22	R.GETGPAGRPEVGPMPGPPGAGEK.G + 3 Hydroxylation (P) (P)
752.363	28	R.GFPGLPGSPGEPGKQGPSGASGER.G + 2 Hydroxylation (P) (P)
569.276	25	R.GQAGVMGFPGPKGAAGEPGKAGER.G + 3 Hydroxylation (P) (P)
770.049	36	K.GDAGPPGPAGPAGPPGPIGNVAPGPK.G + 3 Hydroxylation (P) (P); Deamidated (NQ)
829.74	43	R.GPPGSAGSPGKDGLNGLPGPIGPPGPR.G + 5 Hydroxylation (P) (P); Deamidated (NQ)
833.073	36	K.GDRGETGPAGPPGAPGAPGAPGVPAGK.S + 3 Hydroxylation (P) (P)
837.370	34	K.GDAGPAGPKGEPGSPGENGAPGQMGPR.G + 4 Hydroxylation (P) (P); Deamidated (NQ)
850.08	55	R.GNDGATGAAGPPGPTGPAGPPGFPGAVGAK.G + 3 Hydroxylation (P) (P)
901.752	52	R.GAPGDRGEPGPPGAGFAGPPGADGQPGAK.G + 4 Hydroxylation (P) (P)
903.085	74	R.GFSLQGPPGPPGSPGEGQGPSGASGPAGPR.G + 3 Hydroxylation (P) (P); 2 Deamidated (NQ)
925.763	20	K.QGPSGASGERGPPGPMGPPGLAGPPGESGR.E + 4 Hydroxylation (P) (P)
957.15	50	R.GLTGPIGPPGPAGAPGDKGEAGPSGPAGPTGAR.G + 3 Hydroxylation (P) (P)
960.806	68	R.GVPGPPGAVGPAGKDGEAGAQQGPPGPAGPAGER.G + 3 Hydroxylation (P) (P)

	775.863	27	R.GLPGPPGAPGPQGFQPPGEPGEPGASGPMGPR.G + 5 Hydroxylation (P) (P); Oxidation (M)
	1062.822	24	R.GSEGPQGVRRGEPGPPGAGAAGPAGNPGADGQPGAK.G + 4 Hydroxylation (P) (P); 3 Deamidated (NQ)
	1076.147	34	R.GANGAPGNDGAKGDAGAPGAPGSQGAPLQGMPPER.G + 5 Hydroxylation (P) (P); 2 Deamidated (NQ)
	836.630	31	R.GFSGLDGAKGDAGPAGPKGEPGSPGENGAPQMGPR.G + 3 Hydroxylation (P) (P); 2 Deamidated (NQ); Oxidation (M)
	1134.52	24	R.GNDGATGAAGPPGPTGPAGPPGFGAVGAKGEGGPOGPR.G + 4 Hydroxylation (P) (P); 2 Deamidated (NQ)
	876.1584	26	R.GAPGDRGEPGPPGAGFAGPPGADGQPGAKGEPGDAGAK.G + 5 Hydroxylation (P) (P)
COL1A2 (P02465)	434.73	28	R.GPSGPQGIR.G
	459.725	17	R.AGVMGPAGSR.G + Oxidation (M)
	542.787	24	R.GLVGEPGPAGSK.G + Hydroxylation (P) (P)
	601.30	48	R.GEPGNIGFPGPK.G + 2 Hydroxylation (P) (P)
	603.81	47	R.GFPGTPGLPGFK.G + 2 Hydroxylation (P) (P)
	605.33	44	R.IGQPGAVGPAGIR.G + Hydroxylation (P) (P); Deamidated (NQ)
	631.32	56	R.GEAGPAGPAGPAGPR.G
	634.344	52	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)
	644.322	39	R.GFPGSPGNIGPAGK.E + 2 Hydroxylation (P) (P)
	714.370	55	R.GIPGEFGLPGPAGAR.G + 2 Hydroxylation (P) (P)
	737.343	33	R.GDGGPPGATGFPGAAGR.T + 2 Hydroxylation (P) (P)
	766.90	53	R.GEPGPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)
	769.87	29	R.GAPGAIGAPGPAGANGDR.G + 2 Hydroxylation (P) (P); Deamidated (NQ)
	521.619	45	K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) (P)
	790.890	58	R.GPPGESGAAGPTGPIGSR.G + Hydroxylation (P) (P)
	808.412	64	K.GELGPVGNPGPAGPAGPR.G + Hydroxylation (P) (P)
	418.477	27	R.GAAGIPGGKGEKGETGLR.G + Hydroxylation (P) (P)
	600.605	34	R.GPNGDSRPGEPGLMGPR.G + 2 Hydroxylation (P) (P); Deamidated (NQ); Oxidation (M)
	602.319	41	K.RGSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	684.667	53	R.GEVGPAGPNGFAGPAGAAGQPGAK.G + Hydroxylation (P) (P); Deamidated (NQ)
	686.038	36	K.EGPVGLPGIDGRPGPIGPAGAR.G + Hydroxylation (P) (P)
	1066.062	82	R.GLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	745.368	21	R.GSTGEIGPAGPPGPPGLRGNPGSR.G + 3 Hydroxylation (P) (P); Deamidated (NQ)
	754.712	37	R.GYPGNAGPVGAAGAPGQGPVGPVVK.H + 2 Hydroxylation (P) (P); Deamidated (NQ)
	798.724	38	R.GEVGPAGPNGFAGPAGAAGQPGAKGER.G + Hydroxylation (P) (P); Deamidated (NQ)
	811.08	46	R.GEVGLPGLSGPVGPPGNPGANGLPGAK.G + 4 Hydroxylation (P) (P); Deamidated (NQ)
	858.43	42	R.GSDGSVGPVGPAGPIGSAGPPGFPAGPK.G + 2 Hydroxylation (P) (P)

	861.41	72	K.GENGPVGTGTPVGAAGPSGPNPAPPAGSR.G + Hydroxylation (P) (P); Deamidated (NQ)
	696.089	50	R.GAPGAIGAPGPAGANGDRGEAGPAGPAGPR.G + 2 Hydroxylation (P) (P); Deamidated (NQ)
	931.451	29	K.GEQGPAGPPGFQGLPGPAGTAGEAGKPGER.G + 3 Hydroxylation (P) (P)
	1006.461	34	R.GPPGASGAPGPGQGFQGGPPGEPGEGTGPAGAR.G + 5 Hydroxylation (P) (P)
	1012.169	38	K.GPSGEPGTAGPPGTPGQGLLGGAPFLGLPGSR.G + 5 Hydroxylation (P) (P); Deamidated (NQ)
	807.639	33	R.GPSGPPGPDGNKGEPGVGAPGTAGPSGSPGLPGER.G + 4 Hydroxylation (P) (P)
	1331.011	37	R.GEVGLPGLSGVPPGNPANGLPAGAKGAAGLPVAGAPGLPGR.G + 8 Hydroxylation (P) (P); Deamidated (NQ)
	1070.05	16	R.GLPGERGRVGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGA PGPK.G + 5 Hydroxylation (P) (P)
	1127.796	34	R.SGETGASGPPGFVGEKGPSGEPGTAGPPGTPGQGLLGGAPFLGL PGSR.G + 6 Hydroxylation (P) (P); Deamidated (NQ)
	1213.088	28	R.SGETGASGPPGFVGEKGPSGEPGTAGPPGTPGQGLLGGAPFLGL PGSRGER.G + Hydroxylation (P) (P); Phospho (ST)
COL3A1 (P02458)	694.711	41	R.GAPGPQPPGAPGLGIAGLTGAR.G + Hydroxylation (P) (P)
	604.563	21	R.GPVGPSGPPGKDGASGHPGIPPPGPR.G + Hydroxylation (P) (P)
	989.820	24	R.GPTGPIPPGAPQPGDKGESGAPGVPIAGPR.G + 4 Hydroxylation (P) (P); Deamidated (NQ)
	1553.421	18	R.GSDGQPPGPPGTAGFPSPGAKGEVGPAGSPSSGAPGQRGE PGPQGHAGAPPPGPPGNSGSPGK.G + 7 Hydroxylation (P) (P); Deamidated (NQ); 2 Phospho (ST)

Identified protein (Uniprot entry)	m/z	Peptide score	Peptide
Alpha-S1-casein Bos taurus (P02662)	382.686	17	K.TT <u>M</u> PLW.- + Oxidation (M)
	634.358	52	R.YLGYLEQLLR.L
	692.8711	45	R.FFVAPFPEVFGK.E
	547.961	45	R.FFVAPFPEVFGKEK.V
	554.271	36	K.VPQLEIVPNS <u>A</u> EER.L + Phospho (ST)
	587.320	54	K.HQGLPQEVLENLLR.F
	964.351	45	K.DIG <u>S</u> E <u>S</u> TEDQAMEDIK.Q + 2 Phospho (ST)
	976.482	35	K.YKVPQLEIVPNS <u>A</u> EER.L + Phospho (ST)
	500.778	39	K.KYKVPQLEIVPNSAEER.L
	559.568	44	K.HPIKHQGLPQEVLENLLR.F
	578.811	43	R.FFVAPFPEVFGKEKVNELSK.D
1158.578	88	K.EPMIGVNQELAYFPELFR.Q	

	637.825	22	K.YKVPQLEIVPN S AEERLHSMK.E + Phospho (ST)
	802.657	51	K.EGIHAQQKEPMIGVNVQELAYFYPELFR.Q
	764.988	24	R.LHSMKEGIHAQQKEPMIGVNVQELAYFYPELFR.Q + Deamidated (NQ); Oxidation (M)
	1179.796	51	R.QFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSENSEK.T
Beta-casein Bos taurus (P02666)	371.729	32	R.GPFPIIV.-
	382.686	21	K.EMPFPK.Y + Oxidation (M)
	390.752	19	K.VLPVPQK.A
	415.730	33	K.AVPYPQR.D
	531.146	45	K.VLPVPQKAVPYPQR.D
	1093.590	97	R.DMPIQAFLLYQEPVLPVVR.G
	1463.308	63	R.DMPIQAFLLYQEPVLPVVRGPFPIIV.- + Oxidation (M)
	1005.205	37	K.AVPYPQRDMPIQAFLLYQEPVLPVVR.G + Hydroxylation (P) (P)
	1246.345	37	K.AVPYPQRDMPIQAFLLYQEPVLPVVRGPFPIIV.- + Hydroxylation (P) (P)
	1333.960	23	K.IHPFAQTQSLVYFPFGPIPNQLPQNIPLTQTPVVVPPFLQPEVMG VSK.V + Oxidation (M)
Alpha-S2-casein Bos taurus (P02663)	317.702	25	R.LNFLK.K
	373.731	17	K.VIPYVR.Y
	490.779	33	K.FALPQYLK.T + Deamidated (NQ)
	372.205	17	K.AMKPWIQPK.T + Oxidation (M)
	399.232	27	R.NAVPITPTLNR.E
	451.265	43	K.RNAVITPTLNR.E
	684.355	63	K.ALNEINQFYQK.F
	733.812	34	K.TVDMESTEVFTK.K + Phospho (ST)
	532.243	16	K.TVDMESTEVFTK.K + Phospho (ST)
	532.243	19	K.KTVDMESTEVFTK.K + Phospho (ST)
545.303	17	K.LTEEEKNRLNFLK.K	
Kappa-casein Bos taurus (P02668)	319.182	23	R.FFSDKIAK.Y
	626.358	26	K.YIPIQYVLSR.Y
	402.968	31	R.HPHPHLSFMAIPPK.K
	990.554	80	R.SPAQILQWQVLSNTVPAK.S
	1003.273	52	R.YPSYGLNYYQQKPVALINNQFLPYPYAKPAAVR.S
	1505.472	24	K.NQDKTEIPTINTIASGEPTSTPTTEAVESTVATLEDSPEVIESPPEINT VQVTSTAV

Table S4: Protein identification of the pictorial sample containing animal glue and milk after *in situ* digestion with the use of cellulose acetate as substrate.

Identified Protein (Uniprot entry)	<i>m/z</i>	Peptide score	Peptide
COL1A1 (P02453)	339.665	16	R.GFPGER.G + Hydroxylation (P) (P)
	392.225	46	R.GAAGLPGPK.G + Hydroxylation (P) (P)
	419.216	26	R.GPAGPQGPR.G + Deamidated (NQ)
	426.218	42	R.GFSGLDGAK.G
	450.252	40	R.GVVGLPGQR.G + Deamidated (NQ); Hydroxylation (P) (P)
	544.775	55	R.GFPGADGVAGPK.G + Hydroxylation (P) (P)
	365.856	27	R.GRPGAPGAGAR.G + 2 Hydroxylation (P) (P)
	553.788	41	R.GVQGPAGPR.G + Deamidated (NQ); Hydroxylation (P) (P)
	376.535	41	R.GAAGLPGPKGDR.G + 2 Hydroxylation (P) (P)
	589.776	58	R.GQAGVMGFPGPK.G + Deamidated (NQ); Hydroxylation (P) (P); Oxidation (M)
	596.816	29	R.GVPGPPGAVGPAGK.D + 2 Hydroxylation (P) (P)
	414.234	20	R.GVVGLPGQRGER.G + Hydroxylation (P) (P)
	629.804	73	K.GLTGSPGSPGPDGK.T + 2 Hydroxylation (P) (P)
	664.832	38	R.GFPGLPGSGEPGK.Q + 2 Hydroxylation (P) (P)
	666.834	53	R.GPSGPQGPSGPPGPK.G + Hydroxylation (P) (P)
	718.346	42	R.GEPGAPLPGPPGER.G + 3 Hydroxylation (P) (P)
	730.353	65	R.GSAGPPGATGFPGAAGR.V + 2 Hydroxylation (P) (P)
	780.916	60	R.GETGPAGPAGPIGPVGAR.G
	793.389	44	K.GANGAPGIAGAPGFPGAR.G + 3 Hydroxylation (P) (P)
	552.610	23	R.GFPGADGVAGPKGPAGER.G + Hydroxylation (P) (P)
	605.299	39	R.VGPPGPSNAGPPGPPGAGK.E + Deamidated (NQ); 3 Hydroxylation (P) (P)
	611.627	30	R.GPPGPMGPPGLAGPPGESGR.E + 2 Hydroxylation (P) (P); Oxidation (M)
	932.939	48	K.GEPGPTGIQGGPPGAGEEGK.R + Deamidated (NQ); 2 Hydroxylation (P) (P)
	494.758	18	K.SGDRGETGPAGPAGPIGPVGAR.G
	674.327	25	K.GEPGPTGIQGGPPGAGEEGKR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	686.663	43	K.TGPPGAPQDGRGPPGPPGAR.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	733.663	32	K.GDAGAPGAPGSQAPGLQGMPGER.G + 4 Hydroxylation (P) (P)
	739.357	16	R.GETGPAGRPGEVPPGPPGAGEK.G + 3 Hydroxylation (P) (P)
752.362	18	R.GFPGLPGSGEPGKQGPSGASGER.G + 2 Hydroxylation (P) (P)	

	829.749	56	R.GPPGSAGSPGKDGLNGLPGPIGPPGPR.G + Deamidated (NQ); 5 Hydroxylation (P) (P)
	837.699	46	K.GDAGPAGPKGEPGSPGENGAPGQMGPR.G + 2 Deamidated (NQ); 4 Hydroxylation (P) (P)
	844.741	55	R.GNDGATGAAGPPGPTGPAGPPGFPAGVAVGAK.G + 2 Hydroxylation (P) (P)
	901.755	58	R.GAPGDRGEPGPPGPAFAGPPGADGQPGAK.G + 4 Hydroxylation (P) (P)
	951.807	39	R.GLTGPIGPPGAPAGDKGEAGPSGPAGPTGAR.G + 2 Hydroxylation (P) (P)
	960.810	32	R.GVPGPPGAVGPAGKDGEAGAQGPPGAPAGER.G + 3 Hydroxylation (P) (P)
COL1A2 (P02465)	379.697	32	R.GLPGADGR.A + Hydroxylation (P) (P)
	393.22	36	R.GATGPAGVR.G
	421.233	31	R.GVVGPQGAR.G + Deamidated (NQ)
	434.739	56	R.VGAPGPAGAR.G + Hydroxylation (P) (P)
	448.238	34	R.GPAGPSGPAGK.D
	451.731	52	R.AGVMGPAGSR.G
	542.789	30	R.GLVGEPGPAGSK.G + Hydroxylation (P) (P)
	601.298	22	R.GEPGNIGFPGPK.G + 2 Hydroxylation (P) (P)
	605.339	39	R.IGQPGAVGPAGIR.G + Deamidated (NQ); Hydroxylation (P) (P)
	611.808	37	R.GFPGTGPLPGFK.G + 3 Hydroxylation (P) (P)
	408.545	55	R.GPAGPSGPAGKDGR.I
	631.323	54	R.GEAGPAGPAGPAGPR.G
	423.23	62	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)
	644.318	60	R.GFPGSPGNIGPAGK.E + 2 Hydroxylation (P) (P)
	714.368	46	R.GIPGEFGLPGPAGAR.G + 2 Hydroxylation (P) (P)
	737.344	64	R.GDGGPPGATGFPGAAGR.T + 2 Hydroxylation (P) (P)
	511.601	44	R.GEPGPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)
	769.860	26	R.GAPGAIGAPGPAGANGDR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	521.616	35	K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) (P)
	790.891	51	R.GPPGESGAAGPTGPIGSR.G + Hydroxylation (P) (P)
	808.418	51	K.GELGPVGNPGPAGPAGPR.G + Hydroxylation (P) (P)
	824.920	40	R.GSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	418.475	25	R.GAAGIPGGKGEKGETGLR.G + Hydroxylation (P) (P)
	595.268	35	R.GPNGDSRPGEPGLMGPR.G + Deamidated (NQ); Hydroxylation (P) (P); Oxidation (M)
	602.317	37	K.RGSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	915.463	32	R.TGPPGPSGISGPPGPPGAGK.E + 3 Hydroxylation (P) (P)
	500.254	63	K.HGNRGEPPGAVGPAGAVGPR.G + Deamidated (NQ); Hydroxylation (P) (P)
	684.999	36	R.GEVGPAGPNGFAGPAGAAGQPGAK.G + 2 Deamidated (NQ); Hydroxylation (P) (P)

	686.035	43	K.EGPVGLPGIDGRPGPIGPAGAR.G + Hydroxylation (P) (P)
	1066.064	85	R.GLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	745.370	38	R.GSTGEIGPAGPPGPPGLRGNPGSR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	754.384	40	R.GYPGNAGPVGAAGAPGPQGPVGPVVK.H + 2 Hydroxylation (P) (P)
	811.075	31	R.GEVGLPGLSGPVGPPGNPGANGLPGA.K.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	861.747	44	K.GENGPVGPVGAAGPSGPNPPGAGSR.G + 2 Deamidated (NQ); Hydroxylation (P) (P)
	863.760	43	R.GSDGSVGPVGPAGPIGSAGPPGFPAGPK.G + 3 Hydroxylation (P) (P)
	696.089	43	R.GAPGAIGAPGAGANGDRGEAGPAGPAGPR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	1076.509	23	R.GPSGPPGPDGNKGEVVGAPGTAGPSGSLPGER.G + 4 Hydroxylation (P) (P)
	1235.573	16	R.GAPGPDGNGAQQGPPGLQGVQGGKGEQGPAGPPGFQGLPGPAGTAGEAGKPKGER.G + 4 Deamidated (NQ); 5 Hydroxylation (P) (P)

Identified protein (Uniprot entry)	m/z	Peptide score	Peptide
Alpha-S1-casein Bos taurus (P02662)	382.686	18	K.TTMLW.- + Oxidation (M)
	634.355	40	R.YLGYLEQLLR.L
	446.569	19	K.HIQKEDVPSER.Y
	692.870	23	R.FFVAPFPEVFGK.E
	547.962	28	R.FFVAPFPEVFGKEK.V
	830.899	26	K.VPQLEIVPNSAEER.L + Phospho (ST)
	587.321	50	K.HQGLPQEVLENLLR.F
	976.484	32	K.YKVPQLEIVPNSAEER.L + Phospho (ST)
	694.020	33	K.KYKVPQLEIVPNSAEER.L + Phospho (ST)
559.565	40	K.HPIKHQGLPQEVLENLLR.F	
Alpha-S2-casein Bos taurus (P02663)	490.284	31	K.FALPQYLK.T
	399.229	28	R.NAVPITPTLNR.E
	684.349	40	K.ALNEINQFYQK.F
	733.809	48	K.TVDMESTEVFTK.K + Phospho (ST)
Kappa-casein Bos taurus (P02668)	319.181	21	R.FFSDKIAK.Y
	626.359	22	K.YIPIQYVLSR.Y
	660.703	56	R.SPAQILQWQVLSNTVPAK.S
Beta-casein Bos taurus (P02666)	374.690	27	K.EMPFPK.Y
	390.755	21	K.VLPVPQK.A
	415.731	32	K.AVPYPQR.D

Tables S5-S8: Protein identification of the historical sample "Vecchio Rifodero" after *in situ* digestion with the four developed trypsin functionalized films and then analyzed by LC-MSMS. MS/MS spectra were searched by Mascot MS/MS Ion search using the homemade COLLE database. Deamidation of Gln and Asn, oxidation on Met, pyro- Glu formation at Gln at the N-terminus were set as variable modifications were no fixed modifications were inserted.

Table S5: Protein identification of the historical sample "Vecchio Rifodero" *in situ* digestion with the use of nitrile butadiene rubber polymer material (NBR or nitrile gloves) as substrate.

Identified Protein (Uniprot Entry)	<i>m/z</i>	Peptide Score	Peptide
COL1A1(P02453)	392.224	40	R.GAAGLPGPK.G + Hydroxylation (P) (P)
	449.762	28	R.GVVGLPGQR.G + Hydroxylation (P) (P)
	544.777	67	R.GFPGADGVAGPK.G + Hydroxylation (P) (P)
	553.295	36	R.GVQGGPPGAGPR.G + Hydroxylation (P) (P)
	596.823	47	R.GVPGPPGAVGPAGK.D + 2 Hydroxylation (P) (P)
	597.782	47	R.GQAGVMGFPGPK.G + Deamidated (NQ); Oxidation (M); Hydroxylation (P) (P); Hydroxylation (K) (K)
	621.805	31	K.GLTGSPGSPGPDGK.T + Hydroxylation (P) (P)
	664.832	37	R.GFPGPLGPGSGEPGK.Q + 2 Hydroxylation (P) (P)
	666.836	46	R.GPSGPPGQPSGPPGPK.G + Hydroxylation (P) (P)
	718.348	31	R.GEPGPAGLPGPPGER.G + 3 Hydroxylation (P) (P)
	730.355	42	R.GSAGPPGATGFPGAAGR.V + 2 Hydroxylation (P) (P)
	780.913	62	R.GETGPAGPAGPIGPVGAR.G
	781.897	16	K.DGLNGLPGPIGPPGPR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	793.885	48	K.GANGAPGIAGAPGFPGAR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	552.612	37	R.GFPGADGVAGPKGPAGER.G + Hydroxylation (P) (P)
	845.899	75	K.DGEAGAQQPPGAGPAGER.G
	611.629	50	R.GPPGPMGPPGLAGPPGESGR.E + Oxidation (M); 2 Hydroxylation (P) (P)
	621.966	42	K.GEPGPTGIQPPGPAGEEGK.R + 2 Hydroxylation (P) (P)
	659.341	45	K.SGDRGETGPAGPAGPIGPVGAR.G
	674.33	51	K.GEPGPTGIQPPGPAGEEGKR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	686.665	37	K.TGPPGAGQDGRPGPPGPPGAR.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	692.01	34	K.GAPGADGPAGAPGTPGPQGIAGQR.G + Deamidated (NQ); Hydroxylation (P) (P)
	718.361	27	R.GETGPAGPPGAPGAPGAPGPVGPAGK.S + 2 Hydroxylation (P) (P)

	1107.995	36	K.GDAGAPGAPGSQGAPGLQGMPGER.G + Oxidation (M); 4 Hydroxylation (P) (P)	
	739.360	40	R.GETGPAGRPGEVGPVPPGPPAGEK.G + 2 Hydroxylation (P) (P); Hydroxylation (K) (K)	
	767.690	35	R.GEPGPPGAPGAAGPAGNPGADGQPGAK.G + Deamidated (NQ); 3 Hydroxylation (P) (P)	
	775.382	46	K.GDAGPPGAPGAPGPPGPIGNVGPAGPK.G + Deamidated (NQ); 4 Hydroxylation (P) (P)	
	829.745	53	R.GPPGSAGSPGKDGLNGLPGIPPPGPR.G + Deamidated (NQ); 5 Hydroxylation (P) (P)	
	837.701	61	K.GDAGPAGPKGEPGSPGENGAPGQMGR.G + 2 Deamidated (NQ); Oxidation (M); 2 Hydroxylation (P) (P); Hydroxylation (K) (K)	
	838.412	39	K.GDRGETGPAGPPGAPGAPGAPVGPAGK.S + 4 Hydroxylation (P) (P)	
	850.407	63	R.GNDGATGAAGPPGPTGPAGPPGFPAGVAK.G + Deamidated (NQ); 3 Hydroxylation (P) (P)	
	870.101	38	K.GDAGPPGAPGAPGPPGPIGNVGPAGPKGAR.G + Deamidated (NQ); 4 Hydroxylation (P) (P)	
	901.752	47	R.GAPGDRGEPGPPGAPGAPGADGQPGAK.G + 3 Hydroxylation (P) (P); Hydroxylation (K) (K)	
	902.756	72	R.GFSGGLQGGPPGPPGSPGEGQPSGASGPAGPR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)	
	920.433	27	K.QGSPGASGERGPPGPMGPPGLAGPPGESGR.E + Oxidation (M); 2 Hydroxylation (P) (P)	
	714.115	43	R.GLTGPIGPPGAPGAPGDKGEAGPSGAPGTGAR.G + 2 Hydroxylation (P) (P)	
	960.806	56	R.GVPGPPGAVGPAGKDGEGAQGGPPGAPGAGER.G + 3 Hydroxylation (P) (P)	
	1034.142	33	R.GLPGPPGAPGPPGQGFQGGPEGEPGASGPMGR.G + Oxidation (M); 5 Hydroxylation (P) (P)	
	1134.203	32	R.GNDGATGAAGPPGPTGPAGPPGFPAGVAKGEGGPPGPR.G + Deamidated (NQ); 4 Hydroxylation (P) (P)	
	COL1A2(P02465)	597.339	48	R.IGQPGAVGPAGIR.G + Deamidated (NQ)
		601.300	25	R.GEPGNIGFPGPK.G + 2 Hydroxylation (P) (P)
		611.811	19	R.GFPGTGPLPGFK.G + 2 Hydroxylation (P) (P); Hydroxylation (K) (K)
		634.344	44	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)
714.370		66	R.GIPGEFGLPGAGAR.G + 2 Hydroxylation (P) (P)	
729.349		27	R.GDGGPPGATGFPGAAGR.T + Hydroxylation (P) (P)	
766.900		36	R.GEPGPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)	
521.613		50	K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) (P)	
790.892		42	R.GPPGESGAAGPTGPIGSR.G + Hydroxylation (P) (P)	
808.906		37	K.GELGPVGNPAGPAGPR.G + Deamidated (NQ); Hydroxylation (P) (P)	
824.922		67	R.GSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)	
562.290		39	R.AGVMGPAGSRGATGPAGVR.G + Oxidation (M)	

	600.609	29	R.GPNGDSGRPGEPGLMGPR.G + Deamidated (NQ); Oxidation (M); 2 Hydroxylation (P) (P)
	641.652	43	R.GERGOPGESGAAGPTGPIGSR.G + Hydroxylation (P) (P)
	500.256	45	K.HGNRGEPPGAVGPAGAVGPR.G + Deamidated (NQ); Hydroxylation (P) (P)
	686.039	33	K.EGPVGLPGIDGRPGPIGPAGAR.G + Hydroxylation (P) (P)
	711.044	65	R.GLPVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	754.387	29	R.GYPGNAGPVGAAGAPGPQGPVGPVK.H + 2 Hydroxylation (P) (P)
	804.050	28	R.GEVGPAGPNFAGPAGAAGQPAGKER.G + Deamidated (NQ); Hydroxylation (P) (P); Hydroxylation (K) (K)
	816.736	47	R.GEVGLPGLSGVPVPPGNGANGLPGAK.G + 2 Deamidated (NQ); 4 Hydroxylation (P) (P); Hydroxylation (K) (K)
	825.090	26	R.GERGLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	858.428	22	R.GSDGSVGPVGPAGPIGSAGPPGFPAGPK.G + 2 Hydroxylation (P) (P)
	861.415	46	K.GENGPVGTGPVGAAGPSGPNPPGAGSR.G + Deamidated (NQ); Hydroxylation (P) (P)
	696.093	47	R.GAPGAIGAPGAGANGDRGEAGPAGPAGPR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	931.447	34	K.GEQGPAGPPGFQGLPGPAGTAGEAGKPKER.G + 2 Hydroxylation (P) (P); Hydroxylation (K) (K)
	961.137	31	K.GPKGENGPVGTGPVGAAGPSGPNPPGAGSR.G + 2 Deamidated (NQ); Hydroxylation (P) (P); Hydroxylation (K) (K)
	1007.125	45	R.GPPGASGAPGPQGFQGPPEPGEQGTGPAGAR.G + 2 Deamidated (NQ); 5 Hydroxylation (P) (P)
COL3A1(P02458)	324.680	20	R.GVPGFR.G + Hydroxylation (P) (P)
	371.207	34	R.GRPGPLPGAAGAR.G + 2 Hydroxylation (P) (P)
	564.781	29	R.GLAGPPGMPGAR.G + Oxidation (M); 2 Hydroxylation (P) (P)
	610.297	18	R.GQPGVMGFPGPK.G + Oxidation (M); 2 Hydroxylation (P) (P)
	637.302	49	R.GSPGGPGAAGFPGGR.G + 2 Hydroxylation (P) (P)
	718.847	43	R.GPPGPPGTNGVPGQR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	551.621	38	K.GEMGPAGIPGAPGLIGAR.G + Oxidation (M); Hydroxylation (P) (P)
	560.609	28	R.GPAGANGLPGEKGGPPGDR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	570.629	40	R.GVAGEPGRDGLPGGPGLR.G + 3 Hydroxylation (P) (P)
	611.973	43	R.GPPGPQGLPGLAGTAGEPGR.D + 3 Hydroxylation (P) (P)
	694.701	35	R.GAPGPQPPGAPGLGIAGLTGAR.G + Hydroxylation (P) (P)
	724.000	28	R.GSDGQPPGPPGTAGFPSPGAK.G + 5 Hydroxylation (P) (P)

	732.677	39	K.GAAGPPGPPGSAGTPGLQGMPGER.G + Oxidation (M); 4 Hydroxylation (P) (P)
	578.043	22	R.GVPGFRGPAGANGLPGEKGPPGDR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	861.392	35	R.GEQGPPGPAGFPAGQNGEPGAKGER.G + 2 Deamidated (NQ); 4 Hydroxylation (P) (P)
	989.826	41	R.GPTGPIGPPGPAGQPGDKGESGAPGVPIAGPR.G + Deamidated (NQ); 4 Hydroxylation (P) (P)

Table S6: Protein identification of the historical sample "Vecchio Rifodero" in *situ* digestion with the use of polyvinyl chloride (PEG) as substrate.

Identified Protein (Uniprot Entry)	m/z	Peptide Score	Peptide
COL1A1(P02453)	450.248	18	R.GVVGLPGQR.G + Deamidated (NQ); Hydroxylation (P) (P)
	544.777	47	R.GFPGADGVAGPK.G + Hydroxylation (P) (P)
	581.786	39	R.GQAGVMGFPGPK.G + Deamidated (NQ); Oxidation (M)
	588.828	46	R.GVPGPPGAVGPAGK.D + Hydroxylation (P) (P)
	664.833	49	R.GFPGLPGPSGEPGK.Q + 2 Hydroxylation (P) (P)
	718.35	45	R.GEPGPAGLPGPPGER.G + 3 Hydroxylation (P) (P)
	780.916	81	R.GETGPAGPAGPIGPVGAR.G
	781.896	39	K.DGLNGLPGPIGPPGPR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	793.888	51	K.GANGAPGIAGAPGFPGAR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	795.912	29	R.GLTGPIGPPGPAGAPGDK.G + Hydroxylation (K) (K); Hydroxylation (P) (P)
	552.281	15	R.GERGFPLPGPSGEPGK.Q + Hydroxylation (P) (P)
	611.629	43	R.GPPGPMGPPGLAGPPGESGR.E + Oxidation (M); 2 Hydroxylation (P) (P)
	932.446	24	K.GEPGPTGIQGGPPPAGEEGK.R + 2 Hydroxylation (P) (P)
	659.34	58	K.SGDRGETGPAGPAGPIGPVGAR.G
	674.002	47	K.GEPGPTGIQGGPPPAGEEGKR.G + 2 Hydroxylation (P) (P)
	1038.003	33	K.GAPGADGPAGAPGTPGPQGIAGQR.G + 2 Deamidated (NQ); Hydroxylation (P) (P)
	712.002	17	R.GEPGPPGPAGFAGPPGADGQPGAK.G + 3 Hydroxylation (P) (P)
	713.026	32	R.GETGPAGPPGAPGAPGAPVGPAGK.S + Hydroxylation (P) (P)
	752.366	20	R.GFPGLPGPSGEPGKQGPSGASGER.G + 2 Hydroxylation (P) (P)
	764.72	38	K.GDAGPPGPAGPAGPPGPIGNVGPAGPK.G + Deamidated (NQ); Hydroxylation (K) (K); Hydroxylation (P) (P)
	822.416	36	K.GDRGETGPAGPPGAPGAPGAPVGPAGK.S + Hydroxylation (P) (P)
	829.746	48	R.GPPGSAGSPGKDGLNGLPGPIGPPGPR.G + Deamidated (NQ); 5 Hydroxylation (P) (P)
	845.075	54	R.GNDGATGAAGPPGPTGPAGPPGFPAGVAVGAK.G + Deamidated (NQ); Hydroxylation (K) (K); Hydroxylation (P) (P)
901.756	27	R.GAPGDRGEPGPPGPAGFAGPPGADGQPGAK.G + 4	

			Hydroxylation (P) (P)
	699.115	17	K.SGDRGETGPAGPAGPIGPVARGPAGPQGPR.G
	951.816	55	R.GLTGPIGPPGPAGAPGDKGEAGPSGPAGPTGAR.G + 2 Hydroxylation (P) (P)
	907.188	18	R.GRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAK. G + Deamidated (NQ); 5 Hydroxylation (P) (P)
	1028.989	27	R.GEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPSG AR.G + 6 Hydroxylation (P) (P)
COL1A2(P02465)	597.341	55	R.IGQPGAVGPAGIR.G + Deamidated (NQ)
	619.81	29	R.GFPGTPLPGFK.G + Hydroxylation (K) (K); 3 Hydroxylation (P) (P)
	423.234	24	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)
	714.371	46	R.GIPGEFLPGPAGAR.G + 2 Hydroxylation (P) (P)
	737.346	42	R.GDGGPPGATGFPGAAGR.T + 2 Hydroxylation (P) (P)
	766.899	48	R.GEPGPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)
	781.916	55	K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) (P)
	808.912	32	K.GELGPVGNPAGPAGPR.G + Deamidated (NQ); Hydroxylation (P) (P)
	824.924	59	R.GSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	686.039	61	K.EGPVGLPIDGRPGIPGAGAR.G + Hydroxylation (P) (P)
	690.001	27	R.GEVGPAGPNGFAGPAGAAGQPGAK.G + Deamidated (NQ); Hydroxylation (K) (K); Hydroxylation (P) (P)
	1066.065	63	R.GLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	754.712	50	R.GYPGNAGPVGAAGAPGPQGPVGPVVK.H + Deamidated (NQ); 2 Hydroxylation (P) (P)
	804.056	28	R.GEVGPAGPNGFAGPAGAAGQPGAKGER.G + Deamidated (NQ); Hydroxylation (K) (K); Hydroxylation (P) (P)
	811.077	53	R.GEVGLPGLSGVGPNGPNGANGLPGAK.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	825.104	29	R.GERGLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	696.093	32	R.GAPGAIGAPGPAGANGDRGEAGPAGPAGPAGPR.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	931.780	36	K.GEQGPAGPPGFQGLPGPAGTAGEAGKPGER.G + Deamidated (NQ); Hydroxylation (K) (K); 2 Hydroxylation (P) (P)
	961.122	32	K.GPKGENGPVGPPTGPVGAAGPSGPNPPGPAGSR.G + 2 Deamidated (NQ); Hydroxylation (K) (K); Hydroxylation (P) (P)
	1001.793	34	R.GPPGASGAPGPQGFQGGPPGEPGEGQTGPAGAR.G + 2 Deamidated (NQ); 4 Hydroxylation (P) (P)
	1011.836	44	K.GPSGEPGTAGPPGTPGPQGLLAPGFLGLPGSR.G + 5 Hydroxylation (P) (P)
	807.880	36	R.GPSGPPGPDGNKGEVGVGAPGTAGPSGSPGLPGER.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	COL3A1(P02458)	469.759	34
556.950		37	K.GEMGPAGIPGAPGLIGAR.G + 3 Hydroxylation (P) (P)
559.965		38	R.GVAGEPRDGLPGGPGLR.G + Hydroxylation (P) (P)
611.979		33	R.GPPGPQGLPGLAGTAGEPGR.D + 3 Hydroxylation (P) (P)
700.045		52	R.GAPGPQGGPAGPGLIAGLTGAR.G + 2 Hydroxylation (P) (P)
989.826		53	R.GPTGPIGPPGPAGQPGDKGESGAPGVPIAGPR.G + Deamidated (NQ); 4 Hydroxylation (P) (P)

	1008.980	24	R.GEPGPQGHAGAPPPGPPGNSGSPGGKGMGPAGIPGAPGLIGA R.G + Deamidated (NQ); 7 Hydroxylation (P) (P)
--	----------	----	--

Table S7: Protein identification of the historical sample "Vecchio Rifodero" in *situ* digestion with the use of polyethylene glycol (PEG) as substrate.

Identified Protein (Uniprot Entry)	<i>m/z</i>	Peptide Score	Peptide
COL1A1 (P02453)	601.793	49	R.GEPGNIGFPGPK.G + Hydroxylation (P) (P); Hydroxylation (K) (K); Deamidated (NQ)
	605.332	37	R.IGQPGAVGPAGIR.G + Hydroxylation (P) (P); Deamidated (NQ)
	611.812	19	R.GFPGTGPLPGFK.G + 2 Hydroxylation (P) (P); Hydroxylation (K) (K)
	408.545	25	R.GPAGPSGPAGKDGR.I
	476.582	39	R.GIPGEFGLPGAGAR.G + 2 Hydroxylation (P) (P)
	766.895	50	R.GEPGPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)
	781.911	31	K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) (P)
	824.923	57	R.GSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	602.319	30	K.RGSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	641.654	47	R.GERGPPGESGAAGPTGPIGSR.G + Hydroxylation (P) (P)
	686.039	51	K.EGPVGLPGIDGRPGPIGAGAR.G + Hydroxylation (P) (P)
	689.994	19	R.GEVGPAGPNGFAGPAGAAGQPGAK.G + Hydroxylation (P) (P); Hydroxylation (K) (K); Deamidated (NQ)
	1066.060	63	R.GLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	754.716	25	R.GYPGNAGPVGAAGAPGPQGPVGPVVK.H + 2 Hydroxylation (P) (P); Deamidated (NQ)
	804.384	34	R.GEVGPAGPNGFAGPAGAAGQPGAKGER.G + Hydroxylation (P) (P); Hydroxylation (K) (K); 2 Deamidated (NQ)
	811.078	54	R.GEVGLPGLSGVPGPPGNPGANGLPGAK.G + 4 Hydroxylation (P) (P); Deamidated (NQ)
	825.099	53	R.GERGLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	696.092	60	R.GAPGAIGAPGPAGANGDRGEAGPAGPAGPR.G + 2 Hydroxylation (P) (P); Deamidated (NQ)
	931.447	51	K.GEQGPAGPPGFQGLPGPAGTAGEAGKPGER.G + 2 Hydroxylation (P) (P); Hydroxylation (K) (K)
	961.135	57	K.GPKGENGPVGPPTGPVGAAGPSGPNPPGAGSR.G + Hydroxylation (P) (P); Hydroxylation (K) (K); 2 Deamidated (NQ)
1001.786	34	R.GPPGASGAPGPQGFQGGPEGPEGQTGPAGAR.G + 4 Hydroxylation (P) (P); 2 Deamidated (NQ)	
1011.84	39	K.GPSGEPGTAGPPGTPGPQGLLAPGFLGLPGSR.G + 5 Hydroxylation (P) (P)	
1076.839	22	R.GPSGPPGPDGNKGEVGVGAPGTAGPSGSPGLPGER.G +	

			4 Hydroxylation (P) (P); Deamidated (NQ)
COL1A2(P02465)	581.783	28	R.GQAGVMGFPGPK.G + Deamidated (NQ); Oxidation (M)
	664.830	41	R.GFPGLPGPSGEPGK.Q + 2 Hydroxylation (P) (P)
	780.912	53	R.GETGPAGPAGPIGPVGAR.G
	781.895	23	K.DGLNGLPGPIGPPGPR.G + 3 Hydroxylation (P) (P); Deamidated (NQ)
	793.884	47	K.GANGAPGIAGAPGFPGAR.G + 3 Hydroxylation (P) (P); Deamidated (NQ)
	557.615	17	R.GERGFPLPGPSGEPGK.Q + 2 Hydroxylation (P) (P)
	916.939	49	R.GPPGPMGPPGLAGPPGESGR.E + 2 Hydroxylation (P) (P); Oxidation (M)
	659.339	55	K.SGDRGETGPAGPAGPIGPVGAR.G
	1037.505	70	K.GAPGADGPAGAPGTPGPQGIAGQR.G + Hydroxylation (P) (P); Deamidated (NQ)
	713.026	22	R.GETGPAGPPGAPGAPGAPGVPAGK.S + Hydroxylation (P) (P)
	739.362	37	R.GETGPAGRPGEVGPVPPGPPGAGEK.G + 2 Hydroxylation (P) (P); Hydroxylation (K) (K)
	764.386	46	K.GDAGPPGAPGAPGPPGPIGNVAGAPGPK.G + Hydroxylation (P) (P); Hydroxylation (K) (K)
	822.409	28	K.GDRGETGPAGPPGAPGAPGAPGVPAGK.S + Hydroxylation (P) (P)
	829.747	52	R.GPPGSAGSPGKDGLNGLPGPIGPPGPR.G + 5 Hydroxylation (P) (P); Deamidated (NQ)
	850.405	47	R.GNDGATGAAGPPGPTGPAGPPGFPGAVGAK.G + 3 Hydroxylation (P) (P); Deamidated (NQ)
	902.083	42	R.GAPGDRGEPGPPGAPGAGPPGADGQPGAK.G + 4 Hydroxylation (P) (P); Deamidated (NQ)
	699.107	27	K.SGDRGETGPAGPAGPIGPVARGPAGPQGPGR.G
	714.111	59	R.GLTGPIGPPGAPGAPGDKGEAGPSGAPGPTGAR.G + 2 Hydroxylation (P) (P)
	960.804	22	R.GVPPGPAVGPAGKDGEAGAQQPPGAPAGER.G + 3 Hydroxylation (P) (P)
	1034.473	30	R.GLPGPPGAPGPGQGFQGGPEGEPGASGPMGPR.G + 5 Hydroxylation (P) (P); Deamidated (NQ); Oxidation (M)
1029.222	20	R.GEQGPAGSPGFQGLPGAPPPGEAGKPGEQGVPGDLGAPGSGAR.G + 6 Hydroxylation (P) (P); Deamidated (NQ)	
COL3A1(P02458)	477.756	16	R.DGLPGGPGLR.G + Hydroxylation (P) (P)
	556.785	23	R.GLAGPPGMPGAR.G + Hydroxylation (P) (P); Oxidation (M)
	637.306	30	R.GSPGGPGAAGFPGR.G + 2 Hydroxylation (P) (P)
	826.927	39	K.GEMGPAGIPGAPGLIGAR.G + Hydroxylation (P) (P); Oxidation (M)
	565.293	30	R.GVAGEPRDGLPGGPGLR.G + 2 Hydroxylation (P) (P)
	611.979	22	R.GPPGPQGLPGLAGTAGEPGR.D + 3 Hydroxylation (P) (P)

	733.004	21	K.GAAGPPGPPGSAGTPGLQGMPGER.G + 4 Hydroxylation (P) (P); Deamidated (NQ); Oxidation (M)
	989.496	33	R.GPTGPIGPPGAGQPGDKGESGAPGVPGIAGPR.G + 4 Hydroxylation (P) (P)

Table S8: Protein identification of the historical sample "Vecchio Rifodero" in *situ* digestion with the use of cellulose acetate as substrate.

Identified Protein (Uniprot Entry)	<i>m/z</i>	Peptide Score	Peptide
COL1A2(P02465)	398.2294	43	R.IGQPGAVGPAGIR.G
	634.347	39	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)
	476.585	42	R.GIPGEFGLPGAGAR.G + 2 Hydroxylation (P) (P)
	738.858	15	R.SGETGASGPPGFVGEK.G
	766.902	36	R.GEPGPAGAVGPAGAVGPR.G + Hydroxylation (P) (P)
	521.611	39	K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) (P)
	808.416	42	K.GELGPVGNPAGPAGPR.G + Hydroxylation (P) (P)
	550.279	44	R.GSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	602.313	22	K.RGSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) (P)
	605.306	24	R.TGPPGPSGISGPPGPPGAGK.E + 2 Hydroxylation (P) (P)
	711.048	53	R.GLPGVAGSVGEPGLGIAGPPGAR.G + 3 Hydroxylation (P) (P)
	1131.572	29	R.GYPGNAGPVGAAGAPGQGPVGPVVK.H + Deamidated (NQ); 2 Hydroxylation (P) (P)
	816.413	39	R.GEVGLPGLSGPVGPPGNPGANGLPGAK.G + Deamidated (NQ); Hydroxylation (K) (K); 4 Hydroxylation (P) (P)
	861.744	34	K.GENGPVGPPTGPVGAAGPSGPNPAGPAGSR.G + 2 Deamidated (NQ); Hydroxylation (P) (P)
	1001.789	23	R.GPPGASGAPGPFQGGPPGEPGEGQTGPAGAR.G + 2 Deamidated (NQ); 4 Hydroxylation (P) (P)
	1011.838	19	K.GPSGEPGTAGPPGTPGQGLLGPAGFLGPGSR.G + 5 Hydroxylation (P) (P)
	1076.844	21	R.GPSGPPGPDGNKGEVGVGAPGTAGPSGSPGLPGER.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
COL1A1(P02453)	648.838	23	R.GFPGLPGSGEPGK.Q
	781.901	33	K.DGLNGLPGIPIPPGPR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	793.389	16	K.GANGAPGIAGAPGFPGAR.G + 3 Hydroxylation (P) (P)
	795.916	64	R.GLTGPIGPPGAGAPGDK.G + 2 Hydroxylation (P) (P)
	1037.50	23	K.GAPGADGPAGAPGTPGQGIAGQR.G + Deamidated (NQ); Hydroxylation (P) (P)
	706.999	15	R.GEPGPPGAGFAGPPGADGQPGAK.G + Deamidated (NQ); 2 Hydroxylation (P) (P)
	713.030	23	R.GETGPAGPPGAPGAPGVPAGK.S + Hydroxylation (P) (P)
	764.391	48	K.GDAGPPGAGPAGPPGPIGNVGPAGPK.G + Hydroxylation (K) (K);

			Hydroxylation (P) (P)
	822.418	25	K.GDRGETGPAGPPGAPGAPGAPGPVGPAGK.S + Hydroxylation (P) (P)
	850.081	46	R.GNDGATGAAGPPGPTGPAGPPGFPVAVGAK.G + 3 Hydroxylation (P) (P)
	901.758	39	R.GAPGDRGEPGPPGAPGAGPPGADGQPGAK.G + 4 Hydroxylation (P) (P)
	902.759	51	R.GFSLQGGPPGSPGEGQPSGASGPAGPR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
	961.138	20	R.GVPGPPGAVGPAGKDGEGAGQPPGAPAGER.G + Deamidated (NQ); 3 Hydroxylation (P) (P)

Table S9: Glycation products in the historical canvas "Vecchio rifodero" digestion with the four different functionalized films. LC-MSMS raw data from *in situ* digestion with trypsin functionalized film were searched by Mascot MS/MS Ion search using COLLE database considering semitrypsin as the enzyme, deamidation on Gln and Asn, hydroxylation at Pro, carboxyethyl(K), carboxymethyl(K), Glyoxal(R) and methylglyoxal (R) as variable modifications. Individual ion score threshold provided by MASCOT software (>10) to evaluate the quality of matches in MS/MS data was used for confidence threshold in protein identification.

Nitrile butadiene rubber (NBR or nitrile gloves)			
Protein name (Uniprot Entry)	m/z	Peptide Score	Peptide
Collagen, type I, alpha 1 [Bos taurus] (AAI05185.1)	693.334	21	K.GEPGPTGIQGGPPAGEEGKR.G + 2 Hydroxylation (P) (P); Carboxymethyl (K)
	747.036	13	R.GARGEPGAPGLPGPPGERGGPSR.G + Hydroxylation (P) (P); G-H1 (R)
	971.148	28	R.GLTGPIGPPGAPAGPDKGEAGPSGPAGPTGAR.G + 2 Hydroxylation (P) (P); Carboxymethyl (K)
	1033.251	11	K.GDAGPPGAPGAPGPPGPIGNVGPAGPKGARGSGAPGATGFPGAAGR.V + 5 Hydroxylation (P) (P); Carboxymethyl (K); G-H1 (R)
	1198.306	14	R.GSEGPQGVREPGPPGAPAGAAGPAGNPAGDQPGAKGANGAPGIAGAPGFPGAR.G + 7 Hydroxylation (P) (P); G-H1 (R)
	1396.456	13	K.STGISVPGPMGSPGRGLPGPPGAPGPQGFQPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPRPGER.G + 11 Hydroxylation (P) (P); G-H1 (R)
Collagen, type I, alpha 2 [Bos taurus] (AAI49096.1)	521.278	20	K.DGRIGQPGAVGPAGIR.G + Deamidated (NQ); G-H1 (R)
	858.415	14	K.GPSGEPGTAGPPGTPGQGLLAGPFLGLPGRGER.G + Deamidated (NQ); 5 Hydroxylation (P) (P); MG-H1 (R)
	1084.004	18	R.GSDGSVGPVGPAGPIGSAGPPGFPAGPKGELGVPVGNPPGAPAGPR.G + Deamidated (NQ); 7 Hydroxylation (P) (P); Carboxymethyl (K); G-H1 (R)
Collagen, type III, alpha 1 [Bos taurus] (AAI23470.1)	1006.789	19	R.GGPGGPPQGPAGKNGETGPQGGPPGPTGSPGDK.G + 6 Hydroxylation (P) (P); Carboxymethyl (K)
	1062.226	15	K.GNDGAPGKNGERGGPGGPGQGPAGKNGETGPQGGPPGPTGSPGDK.G + Deamidated (NQ); Carboxyethyl (K); 2 Carboxymethyl (K); G-H1 (R)

	1077.511	24	R.GSPGPQGIK <u>G</u> ENGKPGPSGQNGER <u>G</u> PPGPQGLPGLAGTAGEPGR .D + Deamidated (NQ); 4 Hydroxylation (P) (P); Carboxyethyl (K); Carboxymethyl (K); G-H1 (R)
	1365.398	12	K.GERGSPPGGPGAAGFPGGRRPPGPPGSNGNPGPPSSGAPGKDG PPGPPGSNGAPGSPGISGPK.G + 3 Deamidated (NQ); Hydroxylation (P) (P); MG-H1 (R)
	Polyvinyl chloride (PVC)		
Collagen, type I, alpha 1 [Bos taurus] (AAI05185.1)	1007.119	14	K.GANGAPGIAGAPGFPGARGPSGPQGPSGPPGPK.G + 2 Deamidated (NQ); 8 Hydroxylation (P) (P); MG-H1 (R)
	841.151	12	K.GLTGSPPGSPDGTGPPGAPQDGRPPGPPGAR.G + Deamidated (NQ); 3 Hydroxylation (P) (P); Carboxymethyl (K); MG-H1 (R)
	1025.240	12	K.GARGSAGPPGATGFPGAAGRVGPPGPSNAGPPGPPGAPGKEGS K.G + Deamidated (NQ); 7 Hydroxylation (P) (P); Carboxyethyl (K); MG-H1 (R)
	1258.319	16	R.GFPPGERGVQPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ GAPGLQGMPPGER.G + 4 Deamidated (NQ); Hydroxylation (P) (P); Carboxymethyl (K); G-H1 (R); MG-H1 (R)
Collagen, type I, alpha 2 [Bos taurus] (AAI49096.1)	550.284	10	R.ADQPRSP <u>T</u> SLR <u>P</u> K.D + Hydroxylation (P) (P); Carboxyethyl (K); 2 MG-H1 (R)
	1073.003	10	K.GEPGAVGQPPGPPGSGEEGKR <u>G</u> STGEIGPAGPPGPPGLRGN <u>P</u> GSR .G + Deamidated (NQ); 4 Hydroxylation (P) (P); Carboxymethyl (K); G-H1 (R)
Collagen, type III, alpha 1 [Bos taurus] (AAI23470.1)	758.025	19	K.GEGGPPGAAGPAGGSGPAGPPGPQGVK.G + Deamidated (NQ); 2 Hydroxylation (P) (P); Carboxymethyl (K)
	1077.512	18	R.GSPGPQGIK <u>G</u> ENGKPGPSGQNGER <u>G</u> PPGPQGLPGLAGTAGEPGR .D + Deamidated (NQ); 4 Hydroxylation (P) (P); Carboxymethyl (K); Carboxyethyl (K); G-H1 (R)
	1478.938	17	R.GIPGSPGGPSDGKPGPPGSQGETGRPGPPGSPGPRGQPGVMGF PGPKGNDGAPGKNGER.G + Deamidated (NQ); 7 Hydroxylation (P) (P); Carboxymethyl (K); Carboxyethyl (K); 3 MG-H1 (R)
	Cellulose acetate		
Collagen, type I, alpha 1 [Bos taurus] (AAI05185.1)	804.722	13	K.GDAGPPGPAGPAGPPGPIGNVAGAPGPK.G + Deamidated (NQ); 5 Hydroxylation (P) (P); Carboxyethyl (K)
	1415.336	18	K.GDAGPAGPKGEPGSPGENGAPGQMGRPLGERGRPGAPGPAG AR.G + Deamidated (NQ); 2 Hydroxylation (P) (P); Carboxyethyl (K); G-H1 (R)
	1781.562	12	K.DGEAGAQQPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEA GKPGEQVPGDLGAPGPSGARGERGFPPGER.G + Deamidated (NQ); 10 Hydroxylation (P) (P); Carboxyethyl (K); 4 MG-H1 (R)
Collagen, type I, alpha 2 [Bos taurus] (AAI49096.1)	662.337	10	R.GIPGPVGAAGATGAR.G + 2 Hydroxylation (P) (P); G-H1 (R)
Collagen, type III, alpha 1 [Bos taurus] (AAI23470.1)	1120.005	13	R.GEQPPGPAGFPAGPQNGEPGAK.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
	765.105	10	R.GFPGPPGMKGPAGMPGFPGMKGHRGFDGR.N + 3 Hydroxylation (P) (P); Carboxyethyl (K); G-H1 (R)
	Polyethylene glycol (PEG)		
Collagen, type I, alpha 1 [Bos taurus] (AAI05185.1)	769.357	13	K.GANGAPGIAGAPGFPGARGPSGPQGPSGPPGPK.G + 7 Hydroxylation (P) (P); Deamidated (NQ); Carboxyethyl (K); MG- H1 (R)

Collagen, type I, alpha 2 [Bos taurus] (AAI49096.1)	983.193	14	R.DGARGAPGAIGAPGPAGANGDRGEAGPAGPAGPAGPRGSPGER. G + 4 Hydroxylation (P) (P); Deamidated (NQ); 2 G-H1 (R); MG-H1 (R)
--	---------	----	--

References

- [1] G. Ntasi, D.P. Kirby, I. Stanzione, A. Carpentieri, P. Somma, P. Cicatiello, G. Marino, P. Giardina, L. Birolo, A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects, *J. Proteomics*. 231 (2021). <https://doi.org/10.1016/j.jprot.2020.104039>.
- [2] P. Cicatiello, G. Ntasi, M. Rossi, G. Marino, P. Giardina, L. Birolo, Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings, *Anal. Chem.* 90 (2018) 10128–10133. <https://doi.org/10.1021/acs.analchem.8b01718>.
- [3] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: From bones to paintings ☆, *Microchem. J.* 126 (2016) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.
- [4] S. Longobardi, A.M. Gravagnuolo, L. De Stefano, I. Rea, P. Giardina, Self – assembling fungal proteins and their biotechnological applications, *Rend. Accad. Naz. Delle Sci. Detta Dei XL - Mem. Di Sci. Fis. e Nat.* 37 (2013) 65–76. <https://doi.org/10.4399/97888548717173>.
- [5] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A.R. Da Fonseca, T.W. Stafford, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, M.T.P. Gilbert, J. V. Olsen, Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, *J. Proteome Res.* 11 (2012) 917–926. <https://doi.org/10.1021/pr200721u>.
- [6] R. Linn, I. Bonaduce, G. Ntasi, L. Birolo, A. Yasur-Landau, E.H. Cline, A. Nevin, A. Llueras-Tenorio, Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings, *Angew. Chemie Int. Ed.* 57 (2018) 13257–13260. <https://doi.org/10.1002/anie.201806520>.
- [7] S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, *Nat. Protoc.* 11 (2016) 2301–2319. <https://doi.org/10.1038/nprot.2016.136>.
- [8] M. Mackie, P. R  ther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Pegg  , H. Howard, L. Harrison, L.J. Jensen, J. V. Olsen, E. Cappellini, Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting, *Angew. Chemie - Int. Ed.* 57 (2018) 7369–7374. <https://doi.org/10.1002/anie.201713020>.
- [9] L. De Stefano, I. Rea, E. De Tommasi, I. Rendina, L. Rotiroti, M. Giocondo, S. Longobardi, A. Armenante, P. Giardina, Bioactive modification of silicon surface using self-assembled hydrophobins from *Pleurotus ostreatus*, *Eur. Phys. J. E.* 30 (2009) 181–185. <https://doi.org/10.1140/epje/i2009-10481-y>.
- [10] S. Houmadi, R.D. Rodriguez, S. Longobardi, P. Giardina, M.C. Faure  , M. Giocondo, E. Lacaze, Self-assembly of hydrophobin protein rodlets studied with atomic force spectroscopy in dynamic mode, *Langmuir*. 28 (2012) 2551–2557. <https://doi.org/10.1021/la2028093>.
- [11] N.U. Dharmaratne, T.M.M. Jouaneh, M.K. Kiesewetter, R.T. Mathers, Quantitative Measurements of Polymer Hydrophobicity Based on Functional Group Identity and Oligomer Length, *Macromolecules*. 51 (2018) 8461–8468. <https://doi.org/10.1021/acs.macromol.8b01747>.

5. Integration of mass spectrometric, spectroscopic and microscopic techniques for the molecular characterization of archaeological remains and works of art.

Introduction

Each analysis of a work of art looks like a puzzle of thousands of missing pieces that should be put together in order to reveal its history and preserve it for next generations. Inorganic and organic residues or what's metaphorically left on the puzzle are what make chemistry challenging, since their detection is quite tough due to the complex mixture and the degradation over the years. The origin of the organic material can be directly related to its function in the case of materials such as fabrics and vessels or to the techniques leading to its creation in the case of paintings, for instance. Furthermore, the detection of organic residues in works of art can be also important to understand or to propose restoration processes, in which different natural or synthetic products have been or can be used to “renew” the artwork.

In archeological studies, different scientific disciplines help to address these challenges, in order to understand the origin of the single materials and to molecularly characterize the archaeological remains to unveil past habits and optimize how to preserve them. The scientific approaches are multiple and depend exclusively on the target of molecules to be identified. For example, spectroscopic (IR, RAMAN, XRD) and microscopic (SEM, LIBS) methods are mainly used to study the inorganic part but also as screening tools for the indication of the presence of organic molecules, without attributing them precisely. However, the fact that these approaches are non-destructive is definitively an added value in the field of cultural heritage. On the other hand, the latest advances in analytical techniques such as mass spectrometry allows microinvasive analyses that lead to a molecular characterization with extremely high accuracy of several types of molecules. Specifically, mass spectrometry effectiveness renders it one of the most analytical technique used in the field of cultural heritage [1–7].

The integration of spectrometric and spectroscopic techniques provides more chemical information to face a cultural heritage case study by several scientific points of view [8–12].

Aim of study

This chapter aimed to apply a multidisciplinary approach, by integrating spectroscopic and spectrometric analyses for the molecular identification and characterization of historical paintings, vessels, marble

sculptures and an Egyptian wooden sarcophagus. The spectroscopic analyzes were obtained in collaboration with other scientific teams. My contribution concerns the development of and application of mass spectrometric methodologies (GC-MS and / or LC-MS / MS).

Manuscripts

- Novel use of Evolved Gas Analysis/Mass Spectrometry to identify the earliest organic binding medium in Aegean wall paintings from Tel Kabri, Israel, dated to the 18th C. B.C.E R. Linn, I Bonaduce, Georgia Ntasi, Leila Birolo A Yasur-Landau, E. H. Cline, A. Nevin A. Lluveras Tenorio (Angew. Chemie Int. Ed. 57 (2018))

My specific contribution to this paper was the identification of the proteinaceous materials in the sample.

- A multidisciplinary campaign for the characterization of a XXII dynasty wooden Sarcophagus (yellow coffin). Melchiorre C., Dello Ioio L., Ntasi G., Birolo L. , Trojsi G. , Cennamo P., Barone Lumaga M.R., Fatigati G., Amoresano A., and Carpentieri A (International Journal of Conservation Science, volume 11, Issue 1, January-March 2020)

My specific contribution to this paper was the identification of the proteinaceous materials in the sample and I was also involved in the identification of lipids and carbohydrates by GC-MS.

- Organic remains in amphorae from the temple of Hera in Paestum shed some lights on solemn ceremonies (in preparation)

My specific contribution to this paper was the identification of all the organic materials in the sample.

- Wall paintings from a Roman domus in Santa Maria Capua Vetere (in preparation)

My specific contribution to this paper was the identification of all the organic materials in the sample.

- Molecular characterization of wall paintings from the Royal Palace of Caserta (in preparation)

My specific contribution to this paper was the identification of all the organic materials in the sample.

References

- [1] R.P. Evershed, C. Heron, L. John Goad, Analysis of organic residues of archaeological origin by high-temperature gas chromatography and gas chromatography - Mass spectrometry, *Analyst*. 115 (1990) 1339–1342. <https://doi.org/10.1039/an9901501339>.
- [2] M.P. Colombini, F. Modugno, *Organic Mass Spectrometry in Art and Archaeology*, John Wiley & Sons, Ltd, Chichester, UK, 2009. <https://doi.org/10.1002/9780470741917>.
- [3] M. Regert, C. Rolando, Identification of archaeological adhesives using direct inlet electron ionization mass spectrometry, *Anal. Chem.* 74 (2002) 965–975. <https://doi.org/10.1021/ac0155862>.
- [4] G. Giorgi, Overview of Mass Spectrometric Based Techniques Applied in the Cultural Heritage Field, in: *Org. Mass Spectrom. Art Archaeol.*, John Wiley & Sons, Ltd, 2009: pp. 37–74. <https://doi.org/10.1002/9780470741917.ch2>.
- [5] T.P. Cleland, E.R. Schroeter, M.H. Schweitzer, Biologically and diagenetically derived peptide modifications in moa collagens, *Proc. R. Soc. B Biol. Sci.* 282 (2015) 20150015. <https://doi.org/10.1098/rspb.2015.0015>.
- [6] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A.R. Da Fonseca, T.W. Stafford, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, M.T.P. Gilbert, J. V. Olsen, Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, *J. Proteome Res.* 11 (2012) 917–926. <https://doi.org/10.1021/pr200721u>.
- [7] I. Degano, F. Modugno, I. Bonaduce, E. Ribechini, M.P. Colombini, Recent Advances in Analytical Pyrolysis to Investigate Organic Materials in Heritage Science, *Angew. Chemie Int. Ed.* 57 (2018) 7313–7323. <https://doi.org/10.1002/anie.201713404>.
- [8] H.H. Marey Mahmoud, M. Mahmoud, H. H., A Preliminary Investigation Of Ancient Pigments From The Mortuary Temple Of Seti I, El-Qurna (Luxor, Egypt), *MAA*. 11 (2011) 99–106. <https://ui.adsabs.harvard.edu/abs/2011MAA....11...99M/abstract> (accessed December 5, 2020).
- [9] H.M. Mahmoud, A MULTI-ANALYTICAL APPROACH FOR CHARACTERIZING PIGMENTS FROM THE TOMB OF DJEHUTYEMHAB (TT 194), (2013).
- [10] A. Lluveras-Tenorio, R. Vinciguerra, E. Galano, C. Blaensdorf, E. Emmerling, M. Perla Colombini, L. Birolo, I. Bonaduce, GC/MS and proteomics to unravel the painting history of the lost Giant Buddhas of Bāmiyān (Afghanistan), *PLoS One*. 12 (2017) e0172990. <https://doi.org/10.1371/journal.pone.0172990>.
- [11] A. Casoli, S. Santoro, Organic materials in the wall paintings in Pompei: A case study of Insula del Centenario, *Chem. Cent. J.* 6 (2012) 107. <https://doi.org/10.1186/1752-153X-6-107>.
- [12] J. Blaško, R. Kubinec, B. Husová, P. Příklad, V. Pacáková, K. Štulík, J. Hradilová, Gas chromatography/mass spectrometry of oils and oil binders in paintings, *J. Sep. Sci.* 31 (2008) 1067–1073. <https://doi.org/10.1002/jssc.200700449>.

5.1 Manuscript 1: Novel use of Evolved Gas Analysis/Mass Spectrometry to identify the earliest organic binding medium in Aegean wall paintings from Tel Kabri, Israel, dated to the 18th C. B.C.E

R. Linn, I Bonaduce, Georgia Ntasi, Leila Birolo A Yasur-Landau, E. H. Cline, A. Nevin A. Lluveras Tenorio

Abstract

This study presents the identification of the organic binder used in the painted fragments from the Canaanite palace of Tel Kabri, Israel. Recently dated to the 18th C. B.C.E. by ¹⁴C, Tel Kabri is the most ancient of the Eastern Mediterranean sites in which Aegean paintings were found. The application of pigments in the painted fragments was suspected to be executed using an organic binding medium, particularly for the Egyptian Blue pigment. Samples of blue painted fragments were examined using Evolved Gas Analysis/Mass Spectrometry (EGA-MS) in order to overcome the analytical challenge imposed by highly degraded aged proteinaceous materials. The findings show the use of egg as the binder of the blue paint confirming the use of *a secco* painting technique, in the blue layers of Tel Kabri wall paintings. Results were confirmed by proteomics analysis. To the authors' knowledge, this is the earliest use of egg as a binder in Aegean wall paintings.

Introduction

The discovery of Aegean-style paintings in Eastern Mediterranean archaeological sites, is of great interest regarding the cultural influence and contacts between Bronze Age societies [1]. The study of the materials that were used in these paintings was conducted in order to better understand how the Aegean societies produced such significant works of art during the Bronze Age. It also adds to our knowledge and understanding of trade, technological aspects, materials use and artistic transfer between the Canaanite coast and the Aegean region.

Until now, the morphological and inorganic composition of Aegean paintings, taken from different archaeological sites (Crete, the Cyclades and Greek mainland) including also eastern Mediterranean sites such as Tell el Dab'a and Tel Kabri, have been studied by means of SEM-EDS, XRD, XRF, LIBS and micro Raman [1-7].

However, the study of the binders of these paintings has been rarely attempted. As was previously suggested in the literature, the paintings were executed in a fresco technique, with pigments applied directly on lime-based plaster without any organic binder, although few scholars suggested that a secco technique was used

but with no analytical confirmation [8-13]. Investigation by Brecolaki et al. (2012) on the organic binding media of Aegean wall-paintings from the Mycenaean “Palace of Nestor” in Pylos (ca.1200 B.C.E.) [4,14] and the “West House” at Mycenae (ca. 1250 B.C.E.) [15], proved for the first time the use of organic binders in Aegean style wall-paintings.

Tel Kabri was the capital of a Canaanite Kingdom during the Middle Bronze Age (ca. 1850-1650 BCE). In a palatial complex, approximately 2000 painted plaster fragments of Aegean style painting were found. Based on recent ¹⁴C analysis, the end of phase DW IV to which the wall paintings belong is dated to the end of the 18th century B.C.E. [16]. Consequently, these paintings are the earliest known among the four sites with Aegean-style wall paintings in the east Mediterranean - Tel Kabri, Qatna, Tell el Dab’a and Alalakh [17]. In a larger context, they were painted possibly before the appearance of comparable paintings in the Aegean area in the Late Minoan IA or Late Cycladic I contexts in Crete and the Cyclades [18].

Material and Methods

Samples

The fragments analysed (K1, K7, K20) belong to the defined group A by Linn et al 2017 [19]

EGA/MS

Samples (0.5 mg) for mass spectrometry EGA/MS were placed into a stainless steel cup and inserted into the micro-furnace. They underwent a thermal decomposition in an inert atmosphere (He) over the chosen heating range and evolved gaseous compounds were transferred to the mass spectrometer, directly ionized and analysed as a function of time. The instrumentation used included: Micro-furnace Multi-Shot Pyrolyzer EGA/Py-3030D (Frontier Lab) coupled with a gas chromatograph 6890 Agilent Technologies (Palo Alto, USA) equipped with a deactivated and uncoated stainless steel transfer tube (UADTM-2.5N, 0.15 mm i.d. × 2.5 m length, Frontier Lab). The GC was coupled with a 5973 Agilent single quadrupole mass selective detector (Palo Alto, USA).

Proteomic analyses

The identification of the proteinaceous material in the painting samples was carried out following the minimally invasive proteomic analytical procedure previously described with a deglycosylation pre-treatment step [27]. Briefly, 50 µL of AMBIC 50 mM containing 60 mU/µL of PNGaseF solution was added to microsamples (ca. 600–800 µg) and incubated at 37 °C for 2 h. The reaction was stopped by boiling the

sample for 2 min. Subsequently, enzymatic digestion was carried out with the urea 6 M pre-treatment previously reported [27]. The samples, incubated for 1 h in 20 μ L in 6M urea followed by sonication for 30 min at room temperature, were then 6-fold diluted with ammonium bicarbonate 10 mM pH 7.5 and enzymatic digestion carried out by the addition of 1 μ g of trypsin at 37°C for 16 hours. The supernatants were then recovered by centrifugation, filtered on 0.22 μ m PVDF membrane (Millipore), concentrated and purified using a reverse phase C18 Zip Tip pipette tip (Millipore). Peptides were eluted with 20 μ L of a solution made up of 50% Acetonitrile, 0.1% Formic acid in Milli-Q water and analysed by LC-MS/MS. The eluate was analyzed by LC-MSMS on a LTQOrbitrap-XL (ThermoScientific, Bremen, Germany). Samples were injected onto a capillary chromatographic system consisting of a 2 cm length trapping column (C18, ID 100 μ m, 5 μ m) and a 20 cm C18 reverse phase silica capillary column (ID 75 μ m, 5 μ m) (Nanoseparations). Peptides were fractionated with acetonitrile-based eluents (Solvent A: 0.2% formic acid, 2% acetonitrile in water; Solvent B: 0.2% formic acid, 5% water in acetonitrile) with a gradient from 10 to 60% of B in 70 min, and then from 60 to 100% in 5 min, at 0.25 μ L/min flow rate. MS analysis was performed with a resolution set to 30,000, and mass range from m/z 400 to 1,800 Da. The five most intense doubly, triply and fourthly charged ions were selected and fragmented in the ion trap by using CID fragmentation. Number of precursors selected for tandem-MS in each scan cycle: top 5 peaks; Mass window for precursor ion selection: 2 (m/z); Normalised collision energy: 35.0; Dynamic exclusion settings: 120 sec.

Data handling

The acquired MS/MS spectra were transformed in Mascot Generic files (.mgf) format and used to query the SwissProt database 2015_02 (547,599 sequences; 195,014,757 residues) with Chordata as taxonomy restriction for protein identification with licenced version of MASCOT software (www.matrixscience.com) version 2.4.0. No fixed chemical modification was inserted, but possible oxidation of methionine residues, and deamidation at asparagines and glutamines were considered as variable modifications, and possible hydroxylation at lysines and proline. Only proteins presenting two or more peptides were considered as positively identified. Individual ion score threshold provided by MASCOT software to evaluate the quality of matches in MS/MS data was 34. Spectra with MASCOT score of <10 having low quality were rejected.

Results and discussion

Following the discovery of new fragments in the 2008-2011 excavation seasons, the painted fragments from Tel Kabri were investigated in terms of their polychromy [6]. The blue color, identified as Egyptian Blue, was the most common color in the palette (Figure 1), in agreement with other studies on Aegean wall paintings

[3, 4, 20]. Due to the powdery appearance of the blue paint layers (Figure 2a), the bad adherences of the color to the support, and the distinctive flat interface between the paint layer and its support, as well as the fact that in a typical cross-section of paint layers a clear preparation layer was found (Figure 2b), it was suggested that an organic binding medium might have been used [6].

The study of such ancient paintings is a big challenge for the analytical chemist. It has been demonstrated that during aging, proteinaceous materials become highly thermally stable [21, 22] and show very little solubility, challenging GC-MS and MS proteomics based approaches [23, 24]. Evolved gas analysis mass spectrometry (EGA/MS) has the advantage of giving rapid indications regarding the presence of organic media without the need for extraction. The technique is based on analytical pyrolysis and yields temperature-resolved mass spectrometric information on the gaseous products evolved from samples upon progressive heating [25].



Figure 1: One of the fragments with blue design that was uncovered during 2008-2011 excavation seasons, showing probably a part of a wing (Photo: R. Linn).

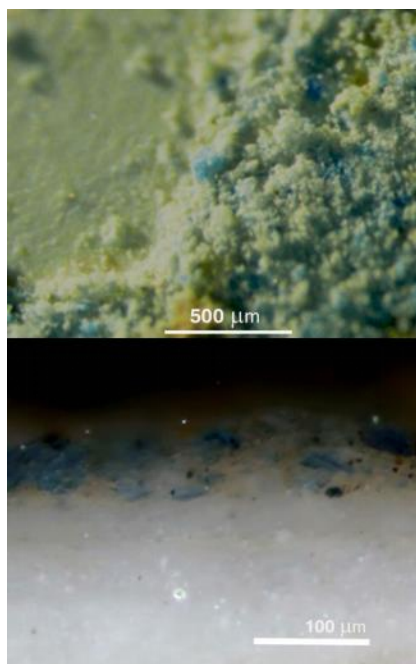


Figure 2 :a) (top) A Micrograph of the blue paint layer showing the powdery appearance of the paint layer (right) and the flat smooth surface that is below (Photo: R. Linn); b) (bottom) A Micrograph of a cross section showing the blue paint layer with Egyptian Blue particles on top of two distinctive white preparation layers suggesting the use of a secco painting technique. Below these layers, is the white lime plaster layer (Photo: R. Linn).

EGA profiles of the samples are shown in Figure 3a, b, c. Data clearly show that both the paint layer of Egyptian blue as well as the preparation layer contain egg (which could be from either egg yolk or whole egg). The Total Ion Thermogram (TIT) profile of both sub-samples (Figure 3a) shows a main degradation step (B) centered at 420°C. The average mass spectra in the range 410-430°C of both samples (Figure 3c) are dominated by fragment ions (m/z 67, 91, 107, 117, 131) corresponding to nitrogen and oxygen containing aromatic compounds (i.e. pyrrole and alkyl-pyrrole, indole and alkyl-indole, phenol and alkyl-phenol, toluene, styrene and ethyl-cyanobenzene) associated with the thermal decomposition of the more thermally stable portion of a protein [21]. The temperature range in which the evolution of aromatic compounds occurs, as well as the absence of fragmented ions ascribable to diketopiperazines, indicate that the material is highly thermally stable and degraded [22,23]. The TIT curve of the blue paint layer also shows a well-resolved peak (A) at around 300°C (Figure 3a). The average mass spectrum in this temperature range (290-310°C) of both samples (Figure 3b), shows the presence of fragmented ions ascribable to hexadecanenitrile and octadecanenitrile (m/z 194, 208, 222, 236) [26].

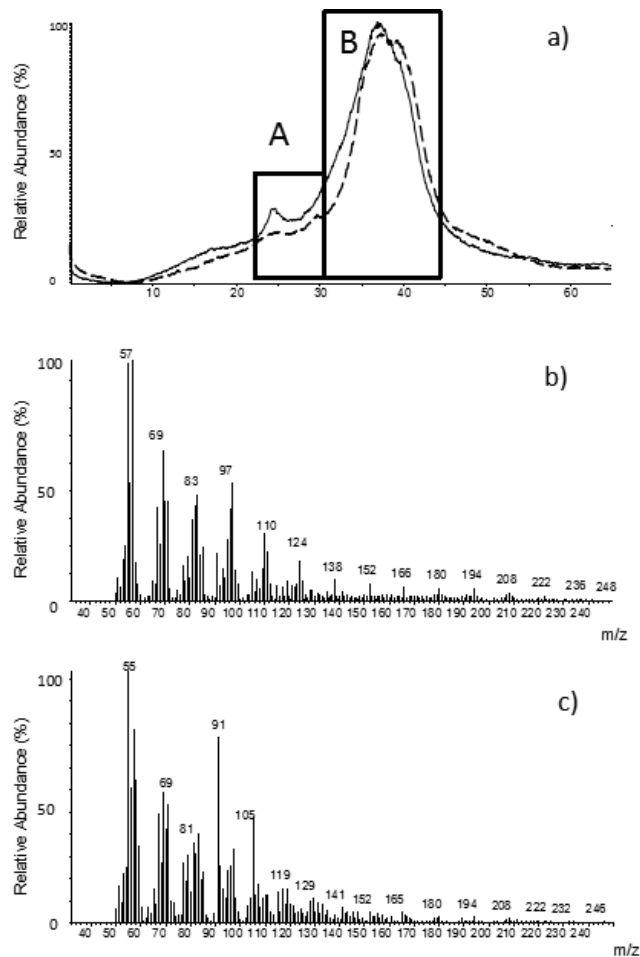


Figure 3: a) Total Ion Thermogram (TIT) of sample K1 showing Egyptian blue paint layer (straight line) and a preparation layer (dotted line); b) Mass spectra of area A; c) Mass spectra of area B. One of the fragments with blue design that was uncovered during 2008-2011 excavation seasons, showing probably a part of a wing (Photo: R. Linn)

Hexadecanitrile and octadecanitrile are known markers of egg yolk [26] and have been identified in extremely degraded samples collected from artificially aged pigmented paint reconstructions containing egg yolk as paint binder [22] and in samples of polychromy on a clay sculpture from the 6th century C.E. [21]. The formation of hexadecanitrile and octadecanitrile has been related to the pyrolysis of the products of a chemical interaction – probably with the formation of covalent bonds of carbonyl moieties of egg lipids and the primary ammine group of Lys and Arg in egg proteins [21]. The absence of fragmented ions related to the

thermal decomposition of the glycerolipid component and cholesterol confirms the observation that the organic material is highly degraded [22].

From an analytical point of view, EGA/MS has not been used so far to identify paint binders in samples from artwork and archaeological materials. To confirm the findings, two samples belonging to the same paint fragments were analysed by MS based proteomics, according to the analytical procedure as implemented to be more sensible in detecting egg proteins in paint samples [28]. Lysozyme C from *Gallus gallus* was unequivocally detected in both samples (9 and 11 peptides, respectively, Table 1), clearly confirming the identification of egg performed by EGA/MS. As far as we are aware, this is the second oldest identification of an organic material used as binder in wall paintings in the literature. The earliest use of egg as binding material has been suggested in the wall paintings from the Neolithic to the Early Bronze age periods (3400-2700 B.C.E.) in the Domus de Janas chamber tombs at Sardinia [27].

The data obtained proves that the Aegean wall paintings from Tel Kabri were executed in a *secco* technique using egg as an organic binding medium. This result is in agreement with the study of Brecolaki et al. (2012) on Mycenaean wall paintings from Pylos, Greece (ca.1200 B.C.E)[14]. However, the current study allows us to extend the use of an organic binder in Aegean-style wall paintings back an additional 500 years and to a wider geographical area including the Eastern Mediterranean. The use of egg as a binder in the painted fragments of Tel Kabri is earlier than the earliest known use of egg in wall paintings of the Bronze Age in the Aegean proper. Since the Kabri paintings are Aegean in their style and technique, this is an indirect evidence that organic binding material and a *secco* painting technique were probably part of the Aegean Bronze Age art from its early stages in the Middle Bronze Age II and III periods. It was used in the Aegean region as well as other locations outside the region, which could therefore bring us to the conclusion that the *a secco* technique was used in Aegean style paintings more extensively than was thought before.

Table 1: Identification of the proteins in the samples from Tel Kabri by overnight trypsin digestion in heterogeneous phase and LCMSMS analysis. Proteins were identified searching UniprotSprot database, with Chordata as taxonomy restriction, with MS/MS Ion search Mascot software (Matrix Science). Only identification of proteins with at least two peptides with individual ion score above the significance threshold (>34), were considered as significant.

Sample	Identified protein (Accession number)	Protein score	Sequence coverage (%)	Matched sequence (Oxidation of methionine, hydroxylation of proline and lysine, deamidation Gln and Asn were inserted as variable modifications in the MSMS Ion Search Program).
K7 (31090)	Lysozyme C (P00698)	482	56	CELAAAMKR CELAAAMKR + Oxidation (M) HGLDNYR HGLDNYRGYSLGNWVCAAKFESNFNTQATNR GYSLGNWVCAAKFESNFNTQATNR GYSLGNWVCAAKFESNFNTQATNR + Deamidated (N) FESNFNTQATNRNTDGSTDYGILQINSR FESNFNTQATNRNTDGSTDYGILQINSR + Deamidated (N) NTDGSTDYGILQINSR KIVSDGNGMNAWVAWR IVSDGNGMNAWVAWR IVSDGNGMNAWVAWR + Deamidated (N) CKGTDVQAWIR
K20 (31091)	Lysozyme C (P00698)	645	56	CELAAAMKR CELAAAMKR + Oxidation (M) HGLDNYR HGLDNYRGYSLGNWVCAAKFESNFNTQATNR HGLDNYRGYSLGNWVCAAKFESNFNTQATNR + Deamidated (N) HGLDNYRGYSLGNWVCAAKFESNFNTQATNR + 2 Deamidated (NQ) GYSLGNWVCAAKFESNFNTQATNR GYSLGNWVCAAKFESNFNTQATNR + Deamidated (N) FESNFNTQATNR FESNFNTQATNRNTDGSTDYGILQINSR FESNFNTQATNRNTDGSTDYGILQINSR + Deamidated (N) NTDGSTDYGILQINSR KIVSDGNGMNAWVAWR

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area)

References

- [1] A. Brysbaert, in: *The Case of the Painted Plaster*. Equinox, London-Oakville. 2008.
- [2] A., Brysbaert, *Technology and Social Agency in Bronze Age Aegean and Eastern Mediterranean Painted Plaster*, PhD in the Department of Archaeology, University of Glasgow, 2004
- [3] A. Brysbaert, K. Melessanaki, D. Anglos, *Journal of Archaeological Science* 2006, 33(8), 1095–1104.
- [4] H. Brecoulaki, C. Zaitoun, S.R. Stocker, J. L. Davis, A.G. Karydas, M.P. Colombini, U. Bartolucci, *Hesperia: The Journal of the American School of Classical Studies at Athens* 2008, 77(3), 363-397.
- [5] R. E Jones, E. Photos-Jones, *British School at Athens Studies* 2005,13, 199-228
- [6] R. Linn, E.H. Cline, A. Yasur-Landau, *Journal of Archaeological Science Reports* 2017, 13, 466–475.
- [7] P. Westlake, P. Siozos, A. Philippidis, C. Apostolaki, B. Derham, A. Terlix, V. Perdikatsis, R. Jones, D. Anglos, *Anal Bioanal Chem* 2012, 402,1413–1432
- [8] A.P. Chapin in *The Oxford Handbook of the Bronze Age Aegean* (Ed.: E.H. Cline) Oxford University Press, Oxford, 2010, pp. 223-236.
- [9] M.A.S. Cameron, *The painted signs on fresco fragments from the 'House of the Frescoes'*. *Kadmos*, 1968, 7, 45-64.
- [10] M.A.S. Cameron, R.E. Jones, S.E. Philippakis. *Annual of the British School at Athens*, 1977, 72, 122-184.
- [11] *Fresco: a Passport into the Past. Minoan Crete through the Eyes of Mark Cameron.*(Ed: D. Evely) British School at Athens e Museum of Cycladic Art, Athens. 1999.
- [12] M.L. Lang, in *The Palace of Nestor at Pylos in Western Messenia*, vol. II. *The Frescoes*. Princeton, 1969.
- [13] B. Niemeier, W.D. Niemeier in *The Wall Paintings of Thera: Proceedings of the First International Symposium* (Ed.: S. Sherratt) Petros M. Nomikos and the Thera Foundation, Athens, 2000, pp. 763-793.
- [14] H. Brecoulaki, A. Andreotti, I. Bonaduce,, M.P. Colombini, A. Lluveras, *Journal of Archaeological Science*, 2012, 39, 2866-2876.
- [15] H. Brecoulaki, S. Sotiropoulou, V. Perdikatsis, A. Lluveras-Tenorio, I. Bonaduce, M.P. Colombini, in *The Wall Paintings of the West House at Mycenae* (Ed.: H. Brecoulaki), INSTAP Academic Press. Philadelphia, 2017, pp. 147-158
- [16] F. Höflmayer, A. Yasur-Landau, E.H. Cline, M.W. Dee, B. Lorentzen, S. Riehl, *Radiocarbon* 2016, 27,
- [17] A. Yasur-Landau, E.H. Cline, A.J. Koh, D. Ben-Shlomo, N. Marom, A. Ratzlaff, I. Samet, *Journal of Field Archaeology* 2015, 40, 607– 625.
- [18] E.H. Cline, A. Yasur-Landau, A. Koh, *Journal of Ancient Egyptian Interconnection* 2017, 13, 43–47.

- [19] R. Linn, E.H. Cline, A. Yasur-Landau, Technological study of Middle Bronze Age painted plaster fragments from the Canaanite Palace of Tel Kabri, Israel - materials and painting techniques, *J. Archaeol. Sci. Reports*. 13 (2017) 466–475. <https://doi.org/10.1016/j.jasrep.2017.03.053>.
- [20] Mycenaean Wall Painting in Context: New Discoveries, Old Finds Reconsidered. Μελετήματα (Eds.: H. Brecolouaki, J.L. Davis, S.R. Stocke) Institute of Historical Research, Section of Greek and Roman Antiquity (KERA), National Hellenic Research Foundation. Athens, 2015
- [21] S. Orsini, F. Parlanti, I. Bonaduce, *Journal*
- [22] S. Sotiropoulou, G. Sciutto, A. Lluveras Tenorio, J. Mazurek, I. Bonaduce, S. Prati, R. Mazzeo, M. Schilling, M. P. Colombini, *Microchemical Journal*, 2017, accepted
- [23] Rif ila
- [24] I. Bonaduce, M., Cito, M.P., Colombini, *Journal of Chromatography A* 2009, 1216, 5931-5939.
- [25] S. Materazzi, S. Vecchio, *Applied Spectroscopy Reviews* 2011, 46, 261-340.
- [26] Bonaduce, I., M.P. Colombini. *Rapid Communications in Mass Spectrometry* 2003, 17, 2523-2527.
- [27] L. Rampazzi, L. Campo, F. Cariati, G. Tanda, M.P. Colombini, *Archaeometry* 2007, 49(3), 559-569.
- [28] Vinciguerra, R.; Galano, E.; Vallone, F.; Greco, G.; Vergara, A.; Bonaduce, I.; Marino, G.; Pucci, P.; Amoresano, A.; Birolo, L.: Deglycosylation step to improve the identification of egg proteins in art samples. *Analytical chemistry* 2015, 87, 10178-10182

5.2 Manuscript 2: A multidisciplinary campaign for the characterization of a XXII dynasty wooden Sarcophagus (yellow coffin).

Melchiorre C., Dello Ioio L., Ntasi G. Birolo L. , Trojsi G. , Cennamo P. , Barone Lumaga M.R. Fatigati G., Amoresano A. and Carpentieri A.

Abstract

The characterization of historical artefacts at a molecular level is becoming an increasingly important aspect for cultural heritage [1]. All the developments in extraction, separation and analytical methodologies can be helpful for challenging tasks, such as the identification of chemical compounds (proteinaceous binders, oils, varnishes) used by artists [2]. Any information thus obtained, is fundamental to investigate the executive techniques and the constitutive materials, at the same time they can also be used to develop protocols for conservation treatment. In this paper, we present the molecular characterization of the organic components extracted from the preparation layers and from the pictorial surface of an Egyptian wooden sarcophagus. The artefact is part of the Egyptian Collection held at the Archeological National Museum of Naples (IT) and belongs to a specific type known with the name of “yellow coffin”, dated at the beginning of the XXII Dynasty. Following the identification of the wood used to build the coffin and the classification of the inorganic components of the preparations, we performed on each selected layer an extraction of the organic components, which, according to differences in chemical physical properties, were subsequently divided into three categories (monosaccharides, lipids and proteins) and separately analyzed. Pigments are not the subject of the current study. The ingenuity of our methodology relies in the use of powerful analytical methodologies (i.e. high resolution MS) which led to the unambiguous identification of heterogeneous molecules.

Introduction

The characterization of the materials used for historical artworks at a molecular level is becoming an increasingly important aspect for cultural heritage [1]. Nowadays all the developments in extraction, separation and analytical methodologies can be helpful for challenging tasks, such as the reconstruction of the “chemical palette” used by artists [2-4].

Information on the manufacturing technique and on the materials (which is the aim of this investigation) are a great source of knowledge useful also for conservation treatments. However, there are several difficulties to be overcome. Primarily, in order to preserve the integrity of the artefact, the amount (size, weight, etc.) of samples to take for the analysis must be as small as possible, thus requiring the use of a very sensitive

analytical methodology. Additionally, ageing, storing conditions and previous conservation treatment attempts can affect the intrinsic nature of the original compounds thus introducing further heterogeneity in the molecules to be identified [5, 6]; therefore the analytical methodology must be able to deal with very complex matrices.

Because of these difficulties, a widescreen procedure able to collect as much information as possible, avoiding numerous sampling procedures, is rather desirable. To this aim, a multi-analytical approach can represent the best choice.

According to literature data, numerous approaches for the characterization of the organic components of artefacts are focused on compounds with the same (or similar) chemical and physical properties [7], while some others [8] suggest the possibility of using a single generic procedure capable of extracting analytes with diverse properties, followed by different analytical strategies. Our approach is based on the latter assumption; the novelty of our methodology relies on the combination of analytical methodologies, which led to the simultaneous and unambiguous identification of a large number of different analytes on the same sample, thus greatly improving the quality and reliability of the results.

Artefacts from ancient Egypt represent an interesting case of study; as it is in fact well-known, natural compounds were used for the realization of utensils, sacred objects, artworks, etc. For example, animal proteins (glue, albumen), polysaccharides elements (resin, gum) and fatty acids esters/oils (beeswax) were used (among the others) as adhesives [9]; similarly, gelatin, glue, gum, albumen and beeswax were mixed as binders with pigments [9]. All listed natural compounds are characterized by significative differences in both chemical and physical properties and can be grouped into three main categories: proteins, polysaccharides and fatty acids.

Our protocol was applied to the characterization of the organic materials extracted from the preparation and paint layers of the lid of an Egyptian wooden coffin (Fig. 1).

A cross-section (Fig. 2) image of a fragment taken from the lid of the coffin clearly shows the presence of four layers above the wooden support, therefore in order to collect informations on all the different layers, three micro samples on the surface of the lid were selected from exposed parts, as shown in Fig. 1.

After the recognition of the wood used to build the coffin, the first step of our workflow was the extraction of the organic molecules. Several extraction procedures have been tested by different groups [10-13], we selected the one based on ammonia solutions that seemed to return the best results [8, 14].

The second step of our protocol was the fractionation (based on the solubility in different organic solvents) of analytes into the three categories (polysaccharides, lipids and proteins) previously indicated.

The third step of our workflow (aimed at the identification of extracted analytes) was based on mass spectrometric analysis by GC-MS (for monosaccharides and lipids) and LC-MSMS (for proteins), preceded by appropriate derivatization and/or digestion procedures depending on the molecules to be identified. As nowadays widely accepted, MS represents one of the most established methodology in the diagnosis in conservation and fine arts, because of its sensibility as well as its ability to analyze complex matrices.

Sarcophagus description

The sarcophagus which was the subject of this study belongs to the Drosso-Picchianti Collection, sold to the Real Museum of Naples on February 27th 1828, as attested by a document about a "mummy of a priest" purchased for the price of 240 ducats (Archivio Storico, Soprintendenza ai Beni Archeologici delle province di Napoli e Caserta, IV B 10, 27/02/1828). The artefact (which can be dated to the beginning of the XXII Dynasty) is an inner anthropoid coffin and belongs to a specific type known in Egyptology as "yellow coffin"; this type has a rich decoration painted on a yellow base layer, from which the name derives. According to its iconography and inscriptions, the coffin is datable towards the end of the XXI and the beginning of the XXII Dynasty (959-889 BC ca.) [15-17]. The coffin was made using different wooden planks, joined together with circular dowels or tenon and mortise joints, finally adding separately carved parts (the face and the hands).

The different layers of plaster on the wood are used to hide the irregularities of the material, such as filling the pores of the wood and the gaps between the boards, creating a smooth surface ready to receive the decoration.

Both lid and base show three different stages of preparation (Fig 1). Immediately on the last layer, there is a red drawing, used to outline the figures initially, which was then filled with vivid colours for the final rendering of the shapes. Only the lid shows an inhomogeneous colored varnish layer on the surface; instead, the base does not have this layer except for some areas.

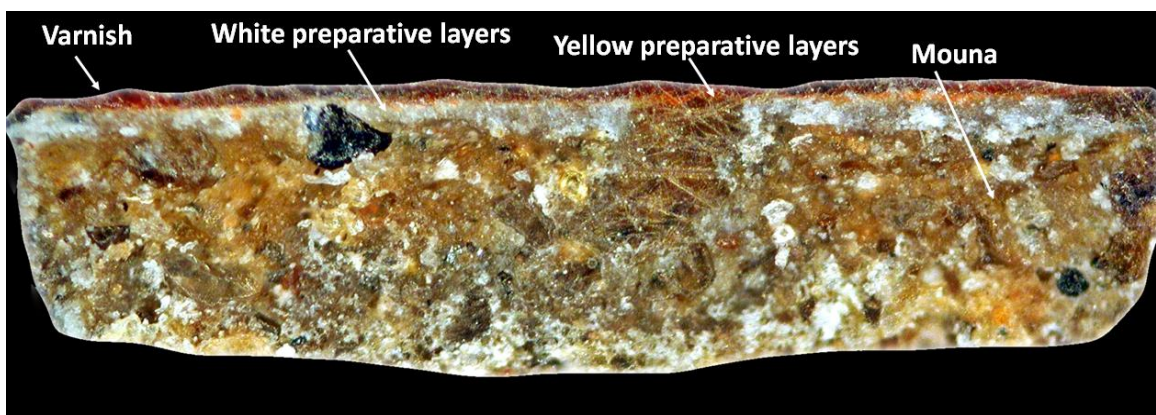


Figure 1: Cross section of a fragment taken from the lid of the sarcophagus observed with an optical microscope (100X).

Materials and Methods

Chemicals

Urea, ammonium bicarbonate (AMBIC), Trypsin TPCK-treated and of N,O-bis- (trimethylsilyl)-acetamide (TMS) from bovine pancreas were from Sigma-Aldrich. Acetonitrile, formic acid, chloroform and methanol were purchased from Baker. Acetone, n-hexane, and chloroform were supplied by Sigma-Aldrich; acetyl chloride, acetonitrile (ACN), were provided by Fluka; formic acid and methyl alcohol anhydrous were supplied by Carlo Erba; bidistilled water (MilliQ) was provided by Millipore. All other reagents and solvents were of the highest purity available from Carlo Erba, Rodano, Italy.

Sample description

For the characterization of the constituent materials, four micro-samples were collected from parts that were exposed because of some lacks.

Four samples (A, B, C and D) were collected as follow:

Sample A, white preparation layer (Fig 2);

Sample C, yellow preparation layer (Fig 2)

Samples B and D, painted layers (Fig 2)

For the identification of the wood, two samples were collected as follow:

Sample 1, wood from the lid of the coffin;

Sample 2, wood from the base of the coffin.

The dimension of each sample was considerably small: about 1 mm³ volume and about 1 mg weight.



Figure 2: The Egyptian wooden sarcophagus under investigation. Sampling areas are highlighted

Wood identification

Light microscopy and scanning electron microscopy (SEM) observations of wood section features were performed for the identification of the woods. We referred to analytical keys texts [18, 19]. Terminology according to Schweingruber (1990) was adopted for the descriptions of characters useful for wood identification.

For light microscopy, transverse and longitudinal (tangential and radial) handmade sections of sampled woods were observed with a Zeiss Axiolab microscope (Zeiss, Jena, Germany) and photographed with a Nikon Digital Sight DS-L1 (Nikon, Tokyo, Japan). For SEM observations, wood samples were coated with gold to about 30 nm. The samples were observed under a VEGA3 TESCAN environmental scanning electron microscope at an accelerating voltage of 20 kV.

Cross sections optical microscopy

Observations of cross sections were carried out on a Nikon Eclipse L150 microscope [20].

Extraction of the organic component

In order to extract the organic fraction from each sample, according to literature data [8, 14] we used the ammonia solution based procedure summarized in Fig 3. More analytically, 400 μL of 2.5 M NH_3 were added to each sample; the extraction was performed in an ultrasonic bath at room temperature for 30 min and then for 120 min at room temperature. The procedure was repeated twice. The extracts were evaporated to dryness under vacuum and then suspended in 100 μL of tri-fluoro-acetic acid (TFA) 1%; to separate the polar compounds (proteins and polysaccharides) from the non-polar ones (lipids) we performed a liquid-liquid extraction by the addition of 200 μL of diethyl ether (three times). Subsequently, proteins were separated from the polysaccharides by the addition of 200 μL Methanol in the in -20°C for 120 min. All the extracts were analyzed separately.

Polysaccharides fraction

Aliquots (100 μl) of the polysaccharides containing fraction were dried under vacuum and submitted to methanolysis. The reaction was performed by adding 500 μl of methanolic HCl (125 μl of acetyl chloride were added to 2,5 ml of methyl alcohol anhydrous) to the samples; the re-N-acetylation of the monosaccharides mixture was performed by adding 500 μl of methanol, 10 μl of pyridine and 50 μl of acetic anhydride at room temperature for 15 min. Sugars were finally trimethylsilylated in 200 μl of N,O-bis-(trimethylsilyl)-acetamide (TMS) at 70°C for 15 min. Each sample was dried down under nitrogen, dissolved in 50 μl of hexane and centrifuged to remove the excess of solid reagents. The hexane supernatant (1/60) was used for the GC-MS analysis.

Lipids fraction

After liquid-liquid extraction (Fig 4) lipids containing fractions were dried under vacuum. Transesterification was performed by adding 150 μl of sulphuric acid to 850 μl methanol, 95°C for 16 hours. Subsequently, the pH was corrected to neutrality using 2 ml of a solution of ammonium bicarbonate 100 mg/ml. The excess of salts was then removed by water. The resulting fatty acids methyl esters were dried under nitrogen flow and then suspended in 1 ml of n-hexane. 1 μl of the sample was analyzed by GC-MS.

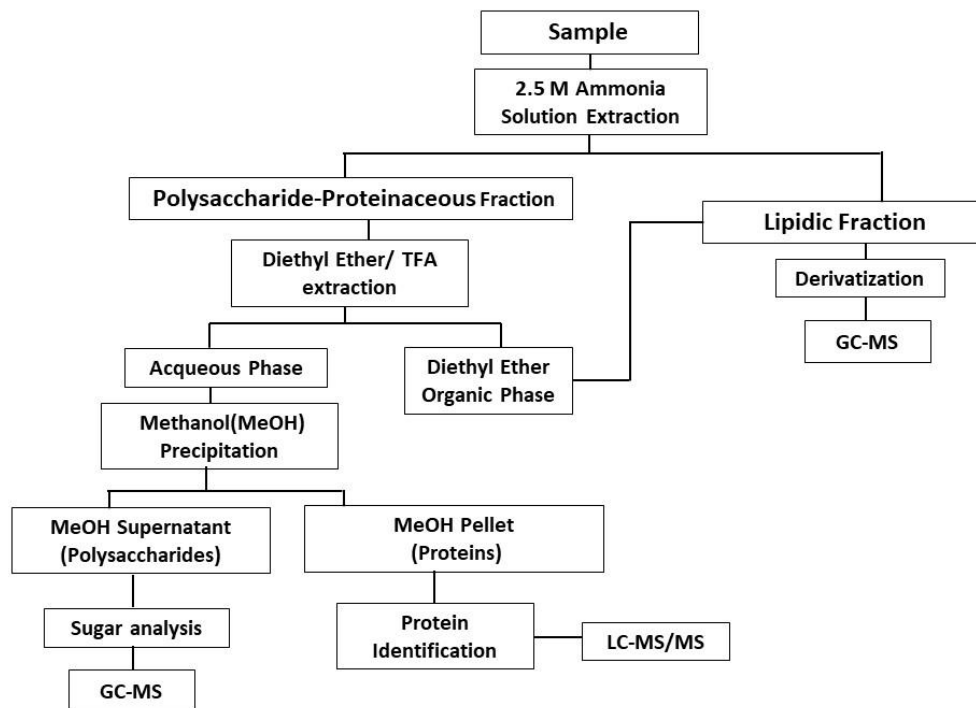


Figure 3 : Flow chart describing the principal steps of the protocol used to extract analytes.

GC-MS analysis

GC-MS analyses were performed on an ISQ-QD quadrupole mass spectrometer (Thermo Fisher scientific) equipped with a TRACE™ 1300 Gas Chromatograph using a Zebron ZB-5HT Inferno (5%-Phenyl-95%-Dimethylpolysiloxane) fused silica capillary column (Column 30 m x 0.32 mm x 0.10 µm) from Phenomenex.

The injection temperature was 250°C, the oven temperature was held at 70°C for 2 min and then increased to 230°C at 20°C/min, increasing to 240°C at 20°C/min and finally to 270°C at 20°C/min and held for 3 min. Electron Ionization mass spectra were recorded by continuous quadrupole scanning at 70eV ionization energy, in the mass range of m/z 30-800. Mass spectra assignment was generally based on the direct match with the spectra of NIST library, if the correlation match index was higher than 95%, the identification was considered reliable. Each analyses was repeated in triplicate.

Proteins fraction.

The proteins containing fractions were evaporated under vacuum and the pellet suspended in 10 µL Urea 6 M. Disulfide bonds were reduced and alkylated. Reduction was performed by using a 10:1 (DTT: cysteins) molar ratio. After 2 h incubation at 37°C, iodoacetamide was added to perform carboxyamidomethylation using an excess of alkylating agent of 5:1 to the moles of thiolic groups. The mixture was then incubated in the dark at room temperature for 30 min. The alkylation reaction was stopped by addition of formic acid to an acidic pH. Samples were 6-fold diluted with ammonium bicarbonate 10 mM pH 7.5 and enzymatic digestion carried out by the addition of 1 µg of trypsin at 37°C for 16 hours. The supernatants were then recovered by centrifugation, filtered on 0.22 µm PVDF membrane (Millipore), and peptides were desalted and concentrated by in-house C18 extraction stage tips as described by [21]. Peptides were eluted by 20 µL of 50% Acetonitrile, 0.1% Formic acid solution and analyzed by LC-MS/MS on an LTQ Orbitrap XL™ Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany). C-18 reverse phase capillary column 75 µm*10 cm (Thermo Fisher Scientific), was performed using a flow rate of 300 nl/min, with a gradient from eluent A (0.2% formic acid in 2% acetonitrile) to eluent B (0.2% formic acid in 95% acetonitrile). The following gradient conditions were used: t=0 min, 5% solvent B; t=10 min, 5% solvent B; t=90 MIN, 50% solvent B; t=100 min, 80% solvent B; t=105 min, 100% solvent B; t=115 min, 100% solvent B; t=120 min; 5% solvent B. Peptides analysis was performed using data-dependent acquisition of one MS scan followed by CID fragmentation of the five most abundant ions. For the MS/MS experiment, we selected the three most abundant precursors and subjected them to sequential CID-MS/MS acquisitions. For the MS scans, the scan range was set to 400–1800 m/z at a resolution of 60,000, and the automatic gain control

(AGC) target was set to 1×10^6 . For the MS/MS scans, the resolution was set to 15,000, the AGC target was set to 1×10^5 , the precursor isolation width was 2 Da, and the maximum injection time was set to 500 ms. The CID normalized collision energy was 35%; AGC target was set to 1×10^5 . Data were acquired by Xcalibur™ software (Thermo Fisher Scientific). Each analyses was repeated in triplicate.

MASCOT identification

The acquired MS/MS spectra were transformed in Mascot Generic files (.mgf) format and used to query the SwissProt database 2015_04 (548,208 sequences; 195,282,524 residues), with Chordata as taxonomy restriction for protein identification. A licensed version of Mascot software (www.matrixscience.com) version 2.4.0 was used with trypsin as enzyme; 3, as allowed number of missed cleavage; 10 ppm MS tolerance and 0.6 Da MS/MS tolerance; peptide charge from + 2 to + 3. No fixed chemical modification was inserted, but possible oxidation of methionine, deamidation of asparagine and glutamine were considered as variable modifications. When collagen proteins were identified, a further identification run was carried out, with the insertion of hydroxylation on lysine and proline as variable modifications, since more confident identifications are commonly obtained for these proteins by taking into consideration their extensive post-translational modifications. Only proteins presenting two or more peptides were considered as positively identified. Individual ion score threshold provided by Mascot software to evaluate the quality of matches in MS/MS data was applied and the protein with scores lower than this were rejected.

Results and Discussion

Wood identification

In order to identify the wood species of the lid and the base of the coffin, wood samples were taken as shown in Fig 1 (samples 1 and 2).

The following histological characteristics allowed identification of the wood samples:

Sample 1 (lid sample): *Acacia* sp. Rays generally 3- to 4 seriate, average height 10 to 20 cells. Rays homogeneous to slightly heterogeneous. Perforation plates simple, no spiral thickenings of the vessel wall. Prismatic crystals in long chains in the axial parenchyma (Fig 4 C, D).

The principal trees growing in Egypt during dynastic times and whose timber was employed by carpenters and joiners were *Acacia*, *Sycamore*, *Fig* and *Tamarisk*. *Acacia* wood has been identified in many artefacts (coffins, boxes etc.) produced from the Predynastic Period up to the Late 5th century A.D [9]. Nile *Acacia* is

native to Egypt, growing on the Nile banks, and is seldom straight growing, of great length or girth, although it is very hard, heavy and durable [22].

The wooden sample 1 is from a species belonging to the genus *Acacia* (Fig 4 C, D). The correspondence of the characteristic anatomical features with those detected in the wood of *Acacia nilotica*, identified by [22] in a Saqqara coffin dated Late Period (712-332 BC), would suggest that *Acacia nilotica* wood was used also in this case for the support of the sarcophagus under investigation.

Sample 2 (base sample) : *Cupressus sempervirens*. Resin canals absent. Rays homocellular. Average height of rays 10 to 15 cells. Tracheid pits uniseriate (Fig 4 A, B). This sample has been taken from a wooden element probably added during XIX Century conservation treatment.

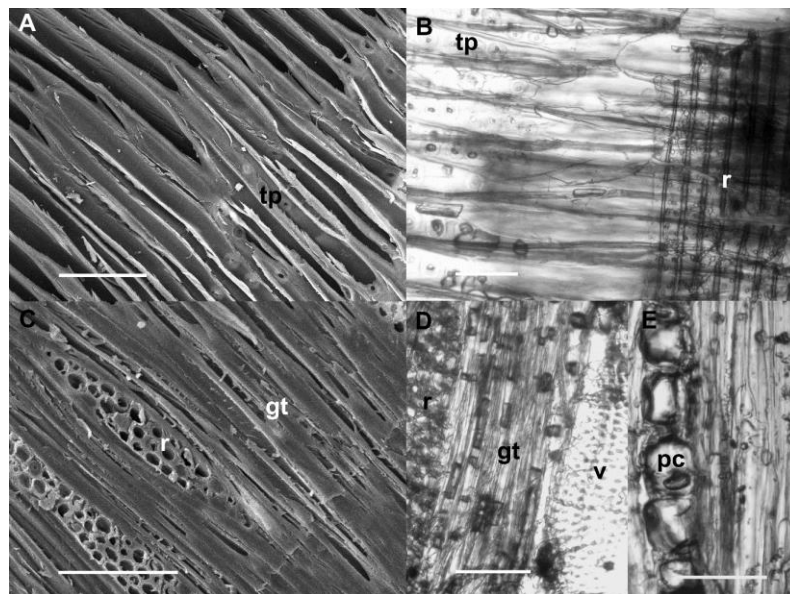


Figure 4: A–B *Cupressus sempervirens* wood of sample 2 (base sample). A Wood surface showing tracheids with uniseriate pits (tp). B Radial section showing tracheids with uniseriate pits (tp), and a ray (r). C–D: *Acacia* sp. wood of sample 1 (lid sample). C Wood surface showing ground tissue (gt) and rays (r). D Tangential section showing a ray (r), ground tissue (gt), and a vessel (v). E Radial section showing prismatic crystals (pc). A, C, scanning electron microscopy; B, D, light microscopy. Scale bars: A, C= 100µm; B, D =50 µm; E=25 µm.

Layers characterization

The examination of a cross section of a sample taken from the lid (Fig 1), confirms that the artefact has three different layers above the wooden support.

The first visible layer in cross section shows an inhomogeneous varnish on the top; immediately below there is a yellow preparation layer, characteristic of this type of sarcophagi belonging to the "yellow coffin" series [25]; finally, there is a white preparation layer and the last layer of preparation called "*mound*" (not the subject of our study) directly in contact with the wooden support.

In order to characterize the inorganic component of the materials of the white and yellow preparation layers (samples A and C), we performed an X-ray diffraction analysis (XRD). Results (summarized in Tab 1) clearly show that both layers contain anhydrite (anhydrous calcium sulfate).

Table 1: Results of XRD analyses of the preparative layers of the sarcophagus

Sample	Identification
A-White preparation layer	Anhydrite
C-Yellow preparation layer	Anhydrite

We then focused on the characterization of the organic components of the white preparation layer, of the yellow preparation layer, of the surface below the decoration and of the paint layer of the coffin (respectively samples A, C, B and D, Fig 2).

To deeply investigate the nature of organic material eventually present in the samples, we applied an extraction multistep procedure able to separate three main class of molecules, namely polysaccharides, lipids and proteins, by using different extraction solvents and taking advantage of the different physical and chemical properties of the target molecules. Our procedure is described in Materials and Methods section and summarized in Fig 3.

In order to identify putative binders constituted by natural polymers, Methanol supernatant fraction (Fig 3) was subjected to methanolysis, TMS derivatization and GC-MS analysis (Fig 5 for sample A) to investigate the presence of the monosaccharides thus produced thus leading to the identification of polysaccharides present.

Taking advantage of the fragmentation spectra and of the retention time (compared with the ones of standard mixtures) we could identify the analytes. As reported in Tab 2 we could detect Galactose, Glucuronic Acid, Ramnose and Arabinose in all the samples taken in consideration. In particular for sample A and C, the co-presence (as long as the relative ratio) of all identified monosaccharides is a very important

result since these sugars represent the main constituents of Arabic gum [26-27] one of the most widespread binder used by ancient Egyptians [9,28] and our data confirm the presence of this polymer in the samples.

Table 2: TMS Monosaccharides profile obtained by GC-MS analysis.

SAMPLE	RT(MIN)	MONOSACCHARIDES	ABBREVIATION	%
A	14,30	Rhamnose	Rha	5,2
	16,25	Arabinose	Ara	25,3
	18,34	Galactose	Gal	43,8
	22,03	Glucuronic acid	GlcA	25,8
B	13,49	Rhamnose	Rha	2,3
	15,71	Arabinose	Ara	29,5
	18,46	Galactose	Gal	68,2
C	14,3	Rhamnose	Rha	18,2
	16,24	Arabinose	Ara	28,1
	18,35	Galactose	Gal	32,9
	22,07	Glucuronic acid	GlcA	20,8
D	18,57	Galactose	Gal	2,2
	22,08	Glucuronic acid	GlcA	97,8

The analysis of the lipids containing fractions, led to results summarized in Tab 3 and showed in Fig 6 for sample A. Thanks to our approach, we could identify palmitic and stearic acid (saturated fatty acids) and oleic acid (monounsaturated fatty acid) as the most abundant components. Identified fatty acids reported in Tab 3 might suggest very little to the identification of the materials since they are among the most common species in many different biological matrices, moreover their relative ratio is affected by aging . Their co-presence though is compatible with the putative presence of egg [29] likely used for the painting techniques. This hypothesis is strengthened by proteins identification.

Table 3 : Fatty Acids Methyl Ester profile obtained for each sample by GC-MS analysis. *(C:N) indicates the number of carbon atoms (C) and double bonds (N) in the fatty acid side chains.

SAMPLE	RT(MIN)	Fatty Acids Methyl Esters	C:N	%
A	19,21	Myristic acid	14:0	1,7
	22,12	Palmitic acid	16:0	48,3
	24,12	Oleic acid	18:1	8,0
	24,4	Stearic acid	18:0	41,9
B	19,65	Palmitic acid	C16:0	24,0
	20,83	1-4-Methyl Palmitic acid	C17:0	2,0
	21,53	Linoleic acid	C18:2	1,8
	21,62	Oleic acid	C18:1	26,3
	21,9	Stearic acid	C18:0	45,9
C	19,65	Palmitic acid	16:0	18,9
	21,61	Oleic acid	18:1	14,6
	21,89	Stearic acid	18:0	66,5
D	19,95	Palmitic acid	16:0	52,0
	21,62	Oleic acid	18:1	7,8
	21,9	Stearic acid	18:0	40,1

Methanol pellet was submitted to a MS based proteomics protocol in order to identify proteins eventually present, according to the analytical procedure reported in Materials and Methods. As detailed in Tab 4, collagen was unambiguously identified in the preparation layers (samples A and C), suggesting the use of animal glue in the manufacture of the sarcophagus itself. Moreover, in the external layers (samples B and D) we could identify exclusively lysozyme, one of the most abundant proteins in albumen. The identification of this protein, suggest that egg was probably used in the realization of the external layers of the sarcophagus, most likely connected to the painting technique. As reported in literature, egg yolk was used for tempera painting technique, though in some cases the entire egg was used for the same technique, the identification of lysozyme (one of the most abundant protein in albumen) clearly confirm this option [30].

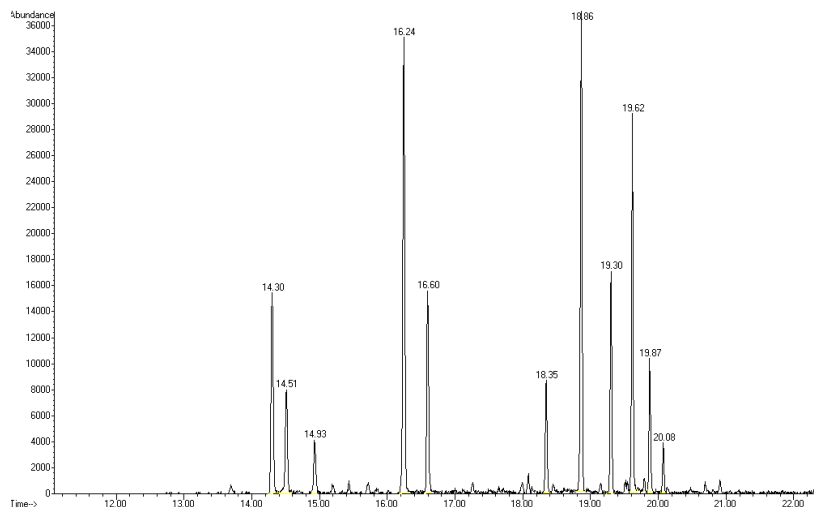


Figure 5: TIC showing the profile of extracted and TMS derivatized monosaccharides extracted from sample A.

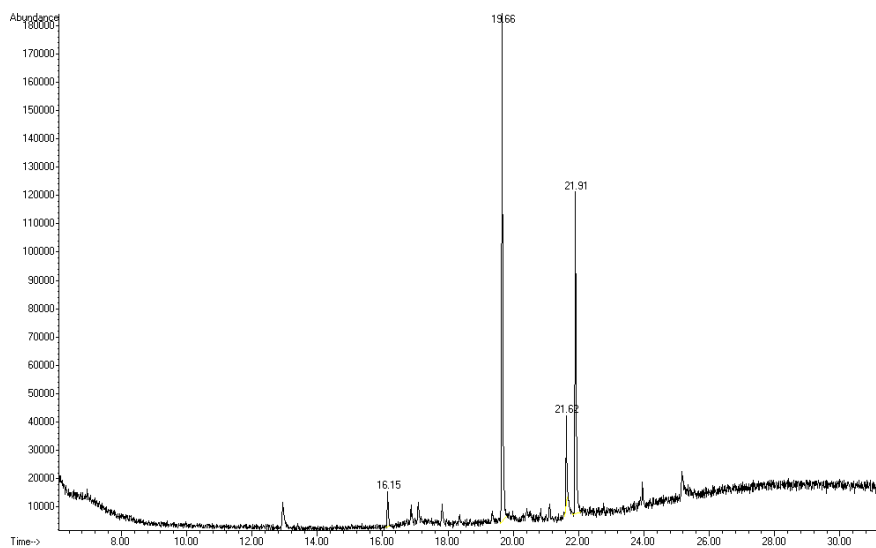


Figure 6: TIC showing the profile of extracted lipids from sample A.

Table 4: Identification of the proteins in the samples from Sarcophagus by overnight trypsin digestion. Proteins were identified searching Uniprot database, with all entries as taxonomy restriction, with MS/MS Ion search Mascot software (Matrix Science). Only identification of proteins with at least two peptides with individual ion score above the significance threshold (>34), were considered as significant.

Sample	Identified protein (Accession number)	Score	Sequence coverage (%)	Matched sequence (Oxidation of methionine, hydroxylation of proline and lysine, deamidation Gln and Asn were inserted as variable modifications in the MSMS Ion Search Program).
D	Lysozyme (P00700)	48	12	R.CKGTDVQAWIR.G -.KVFGR.C
C	Collagen alpha-1(I) (P02453)	764	41	R.GVVGLPGQR.G + Hydroxylation (P) R.GRPGAPGAGAR.G + 2 Hydroxylation (P) R.GVQGPPGAGPR.G + Deamidated (NQ); Hydroxylation (P) R.GQAGVMGFPGPK.G + Deamidated (NQ); Oxidation (M); Hydroxylation (P) R.GPSGPQGSPGPPGPK.G + Hydroxylation (K) R.GSAGPPGATGFPGAAGR.V + 2 Hydroxylation (P) R.GETGPAGPAGPIGPVGAR.G R.GLTGPIGPPGAGAPGDK.G + 2 Hydroxylation (P) K.DGEAGAQQGPPGAGPAGER.G + Hydroxylation (P) R.VGPPGPSGNAGPPGPPGAGK.E + Deamidated (NQ); 3 Hydroxylation (P) R.VGPPGPSGNAGPPGPPGAGK.E + Deamidated (NQ); 3 Hydroxylation (P) R.GPPGPMGPPGLAGPPGESGR.E + Oxidation (M); 2 Hydroxylation (P) K.GEPGPTGIQPPGAGEEGK.R + 2 Hydroxylation (P)
	Collagen alpha-2(I) (P02465)	419	24	K.TGPPGAGQDGRPGPPGPPGAR.G + Deamidated (NQ); 4 Hydroxylation (P) R.GEPGNIGFPGPK.G + Deamidated (NQ); Hydroxylation (K) (K); Hydroxylation (P) R.IGQPGAVGPAGIR.G + Hydroxylation (P) R.GEAGPAGPAGPR.G R.GIPGPVGAAGATGAR.G + Hydroxylation (P) R.GAPGAIGAPGAGANGDR.G + Deamidated

				(NQ); 2 Hydroxylation (P) K.GAAGLPGVAGAPGLPGPR.G + 3 Hydroxylation (P) R.GPPGESGAAGPTGPIGSR.G + Hydroxylation (P) R.GSTGEIGPAGPPGPPGLR.G + 2 Hydroxylation (P) K.GEPGAPGENGTPGQTGAR.G + 2 Deamidated (NQ); 2 Hydroxylation (P) R.GPNGDSGRPGEPGLMGPR.G + Deamidated (NQ); 3 Hydroxylation (P) R.GPPGNVGNPGVNGAPGEAGR.D + 3 Deamidated (NQ); 3 Hydroxylation (P) R.TGPPGPSGISGPPGPPGAGK.E + Hydroxylation (K) (K); 3 Hydroxylation (P)
	Collagen alpha-1(III) (P02458)	278	24	R.GPAGANGLPGEK.G + Deamidated (NQ); Hydroxylation (P) K.GESGAPGVPIAGPR.G + 2 Hydroxylation (P) R.GPPGPPGTNGVPGQR.G + 2 Deamidated (NQ); 3 Hydroxylation (P) K.DGASGHPGPIGPPGPR.G + 2 Hydroxylation (P) R.GETGPAGPSGAPGPAGSR.G + Hydroxylation (P) R.GPTGPIGPPGAGQPGDK.G + Deamidated (NQ); 2 Hydroxylation (P) R.GVAGEPGRNGLPGGGLR.G + Deamidated (NQ); 3 Hydroxylation (P) R.GGPGERGEQPPGAGFPAGQNGEPGAKGERGAPGEK.G + Deamidated (NQ); Hydroxylation (K); 4 Hydroxylation (P) R.GENGSPGAPGAPGHPGPPGVPAGK.S + Deamidated (NQ); 4 Hydroxylation (P)
	Lysozyme (P00700)	120	24	-KVFGRC R.HGLDNYR.G R.CELAAAMKR.H R.CKGTDVQAWIR.G
B	Lysozyme (P00700)	68	17	-KVFGRC R.HGLDNYR.G
A	Collagen alpha-1(I) (P02453)	89	11	R.GVVGLPGQR.G + Hydroxylation (P) R.GVQGPPGAGPR.G + Deamidated (NQ); Hydroxylation (P)

				R.GEPGPAGLPGPPGER.G + 3 Hydroxylation (P)
	Collagen alpha-2(I) (P85154)	72	4	R.GEAGPAGPAGPAGPR.G R.GAPGAIGAPGPAGANGDR.G + Deamidated (NQ); 2 Hydroxylation (P)

Conclusions

In ancient Egypt, the varying design and structure of coffins reflected the chronology of historical periods. Sarcophagi of royals and nobles were made of precious imported wood, whereas coffins of middle class or poor people were generally made of local wood, which was of lower quality and properties [22-31].

A wide screen characterization of materials, by analysing artefacts at a molecular level, is gaining an even more increasing importance. Data collected can be used not only to determine the most suitable conservation methodology and materials to be used, but also to increase historical information. Many different protocols have been developed in the last few years and many efforts have been made in order to gain a complete picture of the nature of binders, of the composition of the preparation layers, painting techniques etc.; our work try to answer all these analytical questions. By using our multistep protocol, we could obtain the unambiguous identification of heterogeneous molecules in a single campaign.

Our data show the presence (in the preparation layers) of anhydrite mixed with an organic matrix composed essentially by Arabic gum. Anhydrite was widely used by Egyptians for the preparation layers of sarcophagi, as used in Sekhemkhet pyramid in Saqqara [9,32-33]. This identification requires a consideration: as in fact already known, high humidity can cause the partial dissolution of the calcium sulphate and its continuous recrystallization thus causing numerous problems to the pictorial film (swellings, falls, etc.). Furthermore, the presence of salts (such as nitrates, chlorides, together with cations such as magnesium, potassium and sodium) could increase the solubility of calcium sulphate due to the increase in ionic strength. This aspect explains the significant cracks on the lid and the base, as well as the widespread abrasion of the paint film on the lid.

Arabic gum was identified in the preparation layers (samples A and C) however the presence of some monosaccharides ascribable to the same polysaccharide were detected in the decorative layers (D and B) thus suggesting its use in the preparation of the coffin. Arabic Gum was widely used in Egyptian tempera painting. This gum is made from Acacia exudate and forms a transparent, slightly viscous film soluble in water. The tempera technique used by the Egyptians was to dilute the pigments in water and then add the gum in order to give body and robustness to the colour backgrounds.

Identified lipids are widespread in many biological matrices and their identifications, if not related to other results (in particular with proteins) could not suggest any precise information in the classification of the biological matrix present. On the contrary, if all the results obtained by our approach are considered as a whole, many informations can be deduced.

In the decorative layers (samples B and D), we could in fact identify lysozyme as the only protein present. This information combined with fatty acids identified, whose co-presence is compatible with the presence of egg, suggests its presence in the decoration of the sarcophagus.

Usually egg yolk was mainly used for painting techniques but here lysozyme (one of the most abundant protein in albumen) was the only protein identified. Decorative layers are the external ones, they interact with the environment and this could explain the fact that proteins present in these samples might probably be degraded or modified.

We could identify proteins in all the layers taken under consideration. The presence of collagen in the preparation layers, suggests that animal glue was used in the construction of the sarcophagus.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area) and from POR, Parco Archeologico Urbano di Napoli (PAUN).

References

- [1] R. Vinciguerra, *Proteomics and Cultural Heritage*, (2016).
- [2] V. Pauk, P. Barták, K. Lemr, *Characterization of natural organic colorants in historical and art objects by high-performance liquid chromatography*, *Journal of separation science*, 37 (2014) 3393-3410.
- [3] J. Romero-Noguera, I. Martín-Sánchez, M.d.M. López-Miras, José Miguel; Bolívar-Galiano, Fernando *Biodeterioration patterns found in dammar resin used as art material* *Electronic Journal of Biotechnology*, vol. 13, núm. 3, mayo, pp. 1-8 *Pontificia Universidad Católica de Valparaíso*, *Electronic Journal of Biotechnology*, 13 (2010) 1-8.
- [4] S. Rinaldi, *La fabbrica dei colori: pigmenti e coloranti nella pittura e nella tintoria*, Il bagatto, 1986.
- [5] G. Leo, I. Bonaduce, A. Andreotti, G. Marino, P. Pucci, M.P. Colombini, L. Birolo, *Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings*, *Analytical chemistry*, 83 (2011) 2056-2064.
- [6] A. Lluveras-Tenorio, R. Vinciguerra, E. Galano, C. Blaensdorf, E. Emmerling, M.P. Colombini, L. Birolo, I. Bonaduce, *GC/MS and proteomics to unravel the painting history of the lost Giant Buddhas of Bāmiyān (Afghanistan)*, *PLoS one*, 12 (2017) e0172990.

- [7] W. Chen, W. Li-Qin, Y. Lu, Z.-Z. Ma, *Application of gas chromatography-mass spectrometry for the identification of organic compounds in cultural relics*, Chinese Journal of Analytical Chemistry, 41 (2013) 1773-1779.
- [8] A. Lluveras, I. Bonaduce, A. Andreotti, M.P. Colombini, *GC/MS analytical procedure for the characterization of glycerolipids, natural waxes, terpenoid resins, proteinaceous and polysaccharide materials in the same paint microsample avoiding interferences from inorganic media*, Analytical chemistry, 82 (2009) 376-386.
- [9] A. Lucas, J. Harris, *Ancient Egyptian materials and industries*, Courier Corporation, 2012.
- [10] M.P. Colombini, R. Fuoco, A. Giacomelli, B. Muscatello, *Characterization of proteinaceous binders in wall painting samples by microwave-assisted acid hydrolysis and GC-MS determination of amino acids*, Studies in Conservation, 43 (1998) 33-41.
- [11] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, *Proteomic strategies for cultural heritage: from bones to paintings*, Microchemical Journal, 126 (2016) 341-348.
- [12] C. Tokarski, E. Martin, C. Rolando, C. Cren-Olivé, *Identification of proteins in renaissance paintings by proteomics*, Analytical chemistry, 78 (2006) 1494-1502.
- [13] W. Fremout, S. Kuckova, M. Crhova, J. Sanyova, S. Saverwyns, R. Hynek, M. Kodicek, P. Vandenabeele, L. Moens, *Classification of protein binders in artist's paints by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry: an evaluation of principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA)*, Rapid communications in mass spectrometry, 25 (2011) 1631-1640.
- [14] G. Gautier, M.P. Colombini, *GC-MS identification of proteins in wall painting samples: a fast clean-up procedure to remove copper-based pigment interferences*, Talanta, 73 (2007) 95-102.
- [15] A. Niwiński, G.L. Nicola, M. Egizio, *Sarcofagi della XXI dinastia:(CGT 10101-10122)*, Ministero per i beni e le attività culturali-Soprintendenza al Museo delle ..., 2004.
- [16] R. Cantilena, P. Rubino, *La Collezione egiziana del museo archeologico di Napoli*, 1989.
- [17] AA.VV. *Guida alla collezione egizia del MANN*, 2016.
- [18] F. Schweingruber, *Anatomie europäischer Hölzer ein Atlas zur Bestimmung europäischer Baum-, Strauch- und Zwergstrauchhölzer Anatomy of European woods an atlas for the identification of European trees shrubs and dwarf shrubs*, Verlag Paul Haupt, Stuttgart, (1990).
- [19] S.L. S. Berti, N. Macchioni, L. Sozzi - I.Va.L.S.A. - CNR (Firenze), Progetto Xyloteca, Banca dati Legnami e software di riconoscimento specie per chiavi dicotomiche, versione per Sistemi operativi MS-Windows. Operating systems (in Italian). I.Va.L.S.A. G. L. Vottero, Ecodata, Italy in, 2002.
- [20] *Documenti Normal* 1985.
- [21] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A. da Fonseca, T.W. Stafford Jr, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, *Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins*, Journal of proteome research, 11 (2011) 917-926.
- [22] A. Abdrabou, M. Abdallah, I.A. Shaheen, H.M. Kamal, *The application of multispectral imaging and reflectance transformation imaging to an ancient Egyptian polychrome wooden stele, wood, laque and furniture group*, in: The 18th Triennial Conference of ICOM-CC, Copenhagen, Denmark, on, 2017, pp. 4-8.
- [23] H. Greaves, *Micromorphology of the bacterial attack of wood*, Wood Science and technology, 3 (1969) 150-166.
- [24] C. Björdal, T. Nilsson, *Waterlogged archaeological wood—a substrate for white rot fungi during drainage of wetlands*, International Biodeterioration & Biodegradation, 50 (2002) 17-23.
- [25] R. Salem, A. Owais, R. Salama, *THE CONSERVATION OF AN EGYPTIAN POLYCHROME WOOD COFFIN FROM LATE PERIOD*, Egyptian Journal of Archaeological and Restoration Studies, 6 (2016) 23.
- [26] M.E. Osman, P.A. Williams, A.R. Menzies, G.O. Phillips, *Characterization of commercial samples of gum arabic*, Journal of Agricultural and food chemistry, 41 (1993) 71-77.

- [27] A. Gallina, E. Fiorese, P. Pastore, F. Magno, *Identification and semi-quantitative determination of gum arabic in wines by GC-MS and size exclusion chromatography*, *Annali di Chimica: Journal of Analytical, Environmental and Cultural Heritage Chemistry*, 94 (2004) 177-184.
- [28] P.T. Nicholson, I. Shaw, *Ancient Egyptian materials and technology*, (2000).
- [29] S. Kuckova, R. Hynek, M. Kodicek, *MALDI-MS applied to the analysis of protein paint binders*, *Organic mass spectrometry in art and archaeology*, 1 (2009) 165-170.
- [30] M.P. Colombini, G. Gautier, *GC/MS in the characterisation of protein paint binders*, John Wiley & Sons: Chichester, United Kingdom, 2009.
- [31] R. Gale, P. Gasson, N. Hepper, G. Killen, *Wood, Ancient Egyptian Materials and Technology*. Cambridge, (2000) 334-371.
- [32] P. Hatchfield, R. Newman, *Ancient Egyptian gilding methods*, Sound View Press, 1991.
- [33] D. Bigelow, *Gilded wood: conservation and history*, Falk Art Reference, 1991.

5.3 Manuscript 3:Organic remains in amphorae from the temple of Hera in Paestum shed lights on solemn ceremonies

Abstract

FT-IR spectroscopy in the latest years has become an essential tool in the field of paleoproteomics as a non-invasive approach for the molecular characterization of archaeological objects and works of art [4–7]. Despite its high efficiency in the identification of minerals, in case of complex mixtures of organic molecules, it cannot be used as an exclusive tool for chemical analysis, but their integration with other analytical techniques allows to define a detailed molecular picture. In this work we exploited the features of FT-IR spectroscopy for the molecular characterization of the inorganic component of amphorae from the temple of Heraion and as a pre-screening tool for the investigation of the presence of organic molecules. Selected samples were subjected to GC-MS and LC-MS-MS analyses, shedding lights to their use in solemn ceremonies in the city of Paestum.

Author's contribution

Leila Birolo and Georgia Ntasi, designed GC-MS and LC-MS/MS experiments; Georgia Ntasi conducted GC-MS and LC-MS/MS experiments; Nunzia De Riso and Alessandro Vergara designed FT-IR experiments ; Nunzia De Riso conducted FT-IR experiments. All authors are involved in data analysis and wrote the paper.

Introduction

Near the temple of Hera, Paestum (S/E of the sanctuary of Hera) in 1950 a new area of interest (zone C) was found by Zancani, of which we have only short and scarce information in a note of 1959 [8]. Essentially, the whole area is still unexplored. The excavation campaigns of the Federico II University was concentrated in this area where a significant action of soil reclamation has been carried out. For the first time, structures already identified in the 1950s have been brought to light, defining their plan, function and chronology. The archaeological investigations, associated with an intense work of survey of the territory, have clarified some problems, discovered interesting news and recovered the function of this area with respect to both the life of the sanctuary and the relationship with the city of Paestum. For Zancani [9], the building from the archaic age, where the potteries were found, was never the site of an official cult, but was nevertheless of some importance in the religious life of the sanctuary. It was probably used for propitiatory practices, reserved for private rites, or used for collective and solemn ceremonies; the presence of amphorae, jugs, goblets and cups of various forms seems to support this hypothesis.

Materials and Methods

Samples

Fragments from twelve amphorae were collected from the temple of Hera in Paestum (DC_49, DC_67, DC_68, DC_70, DC_72, DC_50, DC_53 DC_59, DC_69, DC_54, DC_56, DC_71). The samples were grated and 2-5 mg was used for FT-IR analyses, 1 mg for GC-MS and 10 ug for LC-MS/MS analyses.

Chemicals

Urea, ammonium bicarbonate (AMBIC), Trypsin TPCK-treated and of N,O-bis-(trimethylsilyl)-acetamide (TMS) from bovine pancreases were from Sigma-Aldrich. Acetonitrile, formic acid, chloroform and methanol were purchased from Baker. Acetone, n-hexane, and chloroform were supplied by Sigma-Aldrich; acetyl chloride, acetonitrile (ACN), were provided by Fluka; formic acid and methyl alcohol anhydrous were supplied by Carlo Erba; bidistilled water (MilliQ) was provided by Millipore cartridge equipment. All other reagents and solvents were of the highest purity available from Carlo Erba, Rodano, Italy.

FT-IR spectroscopy

Infrared absorption spectra were collected with an FT-IR Nicolet 5700, in the range of 4000-400 cm^{-1} , using the Omnic software. The spectrometer is equipped with a KBr beam splitter and a MCT-B detector. Spectra were recorded by accumulating 128 scans, with a resolution of 4 cm^{-1} . Samples were typically diluted to 1-5 % in KBr.

Gas-Chromatography Mass Spectrometry

Samples were treated and analyzed as reported in [10].

Liquid-Chromatography Mass Spectrometry.

Samples were treated and analyzed as reported in [11].

Results

Pre-screening by FT-IR spectroscopy

The twelve samples received from Heraion were investigated by FT-IR in ATR mode. The samples showed different features, and were clustered in 3 groups accordingly. The first group includes DC_49, DC_67, DC_68, DC_70, DC_72 (Figure 1), the second DC_50, DC_53 DC_59, DC_69 (Figure 2), and the third group of samples includes DC_54, DC_56, DC_71 (Figure 3).

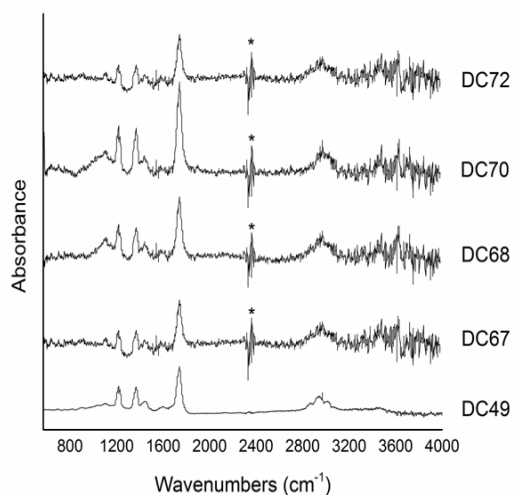


Figure 1: FT-IR spectra for the first group of samples.

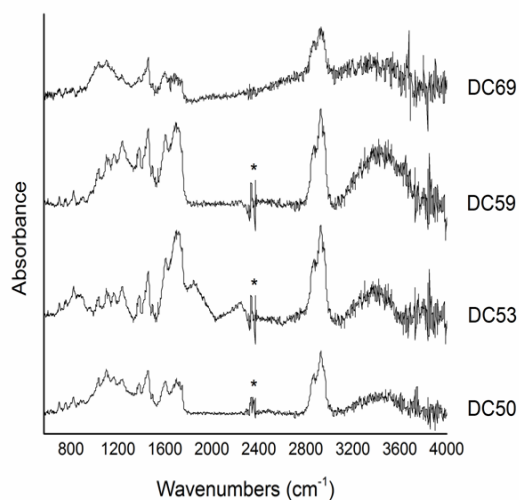


Figure 2: FT-IR spectra for second group of samples

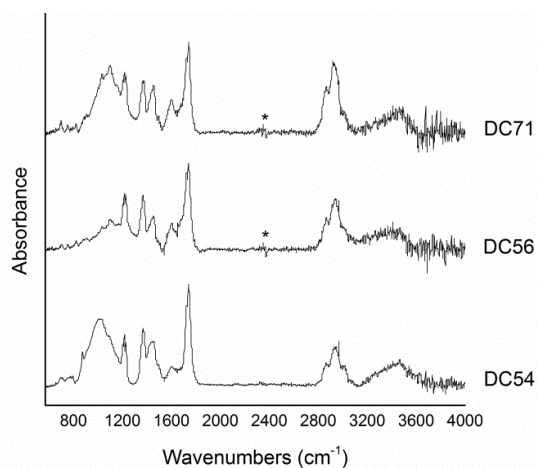


Figure 3: FT-IR spectra for the third group of samples

First of all, a brief survey on the possible contaminants from soil, typically composed by minerals, organic molecules, air and water. Inorganic substances in soil can be grossly divided in primary (silicates) and secondary (clays, hematite, carbonates) minerals, whereas organic substances are mainly humic acids. In our spectra, bands from mineral components from soil were almost never observed. Specifically, Si-O stretching (around 1050-1070 cm^{-1}) silicates/silica is not significantly detected (with the exception of sample DC54), whereas quartz is typically detected by the 777 and 695 cm^{-1} narrow bands. Clays can also be excluded by the absence of any Si-OH band around 3600-3700 cm^{-1} . Al-OH band around 909 cm^{-1} can be also used to check interferences from aluminosilicates. Similarly, neither the strong carbonates band at 1384-1407 cm^{-1} [12] is prominent in our samples. Finally, hematite contribution to the FT-IR spectrum in the investigated range (above 500 cm^{-1}) are expected weak at 535 cm^{-1} related to Fe-O stretching mode [13], and consistently not detectable in our work. The above absences, and the low amount of sample, suggested us not to use any oxidative pyrolysis (treatment at 600° C) to subtract the inorganic contamination from soil [14]. Moreover, the band from humic acid around 1705-1720 cm^{-1} (COOH stretching) could interfere with lipid assignment, but only if accompanied by other humus markers such as 1510-1540 cm^{-1} band [13]. Our samples did not reveal significant water content related to the two bands at 1636 cm^{-1} and 3430 cm^{-1} (possibly detectable only in groups 2), and therefore the microwave session at 100 °C, elsewhere used for soil analysis [14], was herein avoided.

The first group of samples in Figure 1 is characterized by a weak and broad band around 2920 cm^{-1} due to the asymmetric $\nu(\text{CH}_2$ and $\text{CH}_3)$, a strong band at 1739 cm^{-1} assigned to $\nu(\text{C}=\text{O})$ stretching in ester lipids, and some medium bands at 1448 cm^{-1} (various symmetrical and asymmetrical bending $\delta(\text{CH}_3)$ vibration partially overlapped to CH_2 scissoring, $\delta_s\text{CH}_2$), 1365 cm^{-1} (CH_2 twisting, 1217 cm^{-1}), vibration in C-O stretching [15], suggesting the presence of lipids, partly unsaturated. The weak bands around 908, 820 698 and 528 cm^{-1} suggest the presence of sugars [15].

The second group is characterized by the a strong 1698 cm^{-1} band, which can be assigned as C=O vibration of different protein structures [15], as well as the 1455 cm^{-1} band can be assigned to CH_2 scissoring, $\delta_s\text{CH}_2$. The CH stretching bands are very strong with well detectable OH/ H_2O band around 3450 cm^{-1} .

The third group, in addition to the signals reported by the first group, shows weak signals also around 874 and 696 cm^{-1} within the range that contains the various absorption of C-O bending frequencies of the polysaccharide component [16].

Conclusively, all samples are rich in organic components and, due to the very small amount and given the

profile similarity between the samples of each group, the richest spectrum of each group was selected as representative of the group. In particular, from the first group we chose sample DC 72, in the hypothesis of the presence of lipids, partly unsaturated. From the second group, sample DC 54 was selected, and we hypothesize the presence of proteins and other substances. Moreover, in sample DC 59 from the third group we suggest the presence of sugars and proteins. GC-MS and LC-MS/MS analyses were performed in order to further validate the FT-IR analysis results and to provide chemical details.

Table 1: FTIR wave numbers assignments in the four analyzed samples.

Group	Sample	FT-IR (cm ⁻¹) wavenumbers	Putative assignment
1st	DC_49	2968(w, br)	Asymmetric ν (CH ₃ /CH ₂)
		1736(s)	ν (C=O) stretching in ester lipids
		1445(m)	lipids, bending δ (CH ₃), δ_s CH ₂
		1365(m), 1217(m)	lipids, CH ₂ twisting, C-O stretching
		908(w), 820(w), 696(w), 528(w)	Sugars
	DC_53	2968(w, br)	Asymmetric ν (CH ₃ /CH ₂)
		1740(s)	ν (C=O) stretching in ester lipids
		1450(m)	lipids, bending δ (CH ₃), δ_s CH ₂
		1370(m), 1220(m)	lipids, CH ₂ twisting, C-O stretching
		908(w), 821(w), 698(w), 528(w)	Sugars
	DC_67	2968(w, br)	Asymmetric ν (CH ₃ /CH ₂)
		1739(s)	ν (C=O) stretching in ester lipids
		1446(m)	lipids, bending δ (CH ₃), δ_s CH ₂
		1365(m), 1217(m)	lipids, CH ₂ twisting, C-O stretching
		906(w), 820(w), 701(w), 528(w)	Sugars
	DC_68	2969(w, br)	Asymmetric ν (CH ₃ /CH ₂)
		1740(s)	ν (C=O) stretching in ester lipids
		1447(m)	lipids, bending δ (CH ₃), δ_s CH ₂
		1370(m), 1213(m)	lipids, CH ₂ twisting, C-O stretching
		908(w), 820(w), 696(w), 528(w)	Sugars
	DC_70	2973(w, br)	Asymmetric ν (CH ₃ /CH ₂)
		1739(s)	ν (C=O) stretching in ester lipids
		1447(m)	lipids, bending δ (CH ₃), δ_s CH ₂
		1366(m), 1214(m)	lipids, CH ₂ twisting, C-O stretching
905(w), 820(w), 698(w), 528(w)		Sugars	
DC_72	2970(w, br)	Asymmetric ν (CH ₃ /CH ₂)	
	1740(s)	ν (C=O) stretching in ester lipids	

		1446(m)	lipids, bending $\delta(\text{CH}_3)$, $\delta_s\text{CH}_2$
		1370(m), 1213(m)	lipids, CH_2 twisting, C-O stretching
		908(w), 820(w), 699(w), 528(w)	Sugars
2nd	DC_50	3400(br)	OH/H ₂ O band
		1454(s)	protein, CH_2 scissoring ($\delta_s\text{CH}_2$)
		1697(m)	protein, C=O vibration
		908(w), 818(w), 696(w)	Sugars
	DC_59	3400(br)	OH/H ₂ O band
		1454(m)	protein, CH_2 scissoring ($\delta_s\text{CH}_2$)
		1698(s)	protein, C=O vibration
		904(w), 817(w), 696(w)	Sugars
	DC_69	3400(br)	OH/H ₂ O band
		1458(s)	protein, CH_2 scissoring ($\delta_s\text{CH}_2$)
		1698(m)	protein, C=O vibration
	3rd	DC_54	2939(m)
1741(vs)			$\nu(\text{C}=\text{O})$ stretching in ester lipids
1447(m)			lipids, bending $\delta(\text{CH}_3)$, $\delta_s\text{CH}_2$
1370(m), 1213(m)			lipids, CH_2 twisting, C-O stretching
1050(vs)			Si-O stretching
872(m), 694(w)			polysaccharide C-O bending
DC_56		2939(m)	Asymmetric $\nu(\text{CH}_3/\text{CH}_2)$
		1736(vs)	$\nu(\text{C}=\text{O})$ stretching in ester lipids
		1452(m)	lipids, bending $\delta(\text{CH}_3)$, $\delta_s\text{CH}_2$
		1365(m), 1220(m)	lipids, CH_2 twisting, C-O stretching
		877(w), 696(w)	polysaccharide C-O bending
DC_71		2925(vs)	Asymmetric $\nu(\text{CH}_3/\text{CH}_2)$
		174(vs)	$\nu(\text{C}=\text{O})$ stretching in ester lipids
		1452(m)	lipids, bending $\delta(\text{CH}_3)$, $\delta_s\text{CH}_2$
		1365(m), 1220(m)	lipids, CH_2 twisting, C-O stretching
		874(w), 694(w)	polysaccharide C-O bending

GC-MS analysis

A multistep fractionation procedure of the samples based on the ammonia extraction of proteins and polysaccharide materials, in order to separate them from lipid and resinous materials was followed and polysaccharidic and lipidic fractions were subjected to GC-MS analysis [10] after appropriate derivatization. Tables 2-4 present the methyl ester derivatives of the fatty acids (FAMES) extracted from the lipidic fractions, whereas Fig 5 lists the TMS derivatives after sugar analysis of the polysaccharide fractions. GC-MS analysis identifies the sealing material that had been used in the amphoraes for impermeabilization. Specifically, in all polysaccharidic fractions, abietanic structures were identified, indicating the presence of resins and attributing them to *Pinaceae* family [17]. In particular, dehydroabietic acid, 7-oxodehydroabietic acid as also methyl 7-oxoabieta-9(11),8(14),12-trien-18-oate are oxidation products of abietic acid due. Moreover, compounds deriving from resin thermal decomposition were also detected. Prolonged heating cause decarboxylation of dehydroabietic acid, generating dehydroabietin and norabietamine. Norabietamine is further dehydrogenated to tetrahydroretene and retene. Finally, the dealkylation of retene leads to phenanthrene [18] [19] [20]. The simultaneous presence of retene (Fig 2) and phenanthrene compounds (tables 1-3) highlights that the resin have been subjected to extensive thermal processes in the presence of wood obtained from plants of the *Pinaceae* [21,22].

GC-MS analyses revealed important information about amphoraes' content since several secondary metabolites of animal origin were identified.

In all samples two lactose isomers were detected, demonstrating that some milk had been inside the amphoraes [23]. The identification of lactose after the reactions of methanolysis and R,N acetylation, thus after sugar analysis, means that the organic remains of milk are trisaccharides of lactose such as lactosucrose. During methanolysis, the glucosidic bond between the sugar and lactose breaks and the released lactose is identified. Because of lactose and glucopyranose extremely high chemical structure similarity, comparison of MS spectra between analytical standard of lactose and the identified isomers was performed as a further validation. The standard of lactose was not subjected to sugar analysis but was only derivatized in order to maintain the disaccharide structure. Fig 6 reports the comparison of MS spectra of one of the lactose isomers with the MS spectra of lactose. Moreover, the evidence of the presence of milk comes also from the identification of several glucopyranosiduronic isomers (Fig 2). Glucopyranosiduronic acid derivatives belong to the family of glucuronides, comprising glucuronic acid linked to another chemical

molecule via a glycosidic bond, provoked from enzyme or bacteria, in a chemical process known as glucuronidation [24]. Considering the identification of two lactose isomers we propose that glucopyranosiduronic derivatives are mammalian secondary metabolites in glucuronidation pathway. On the basis of the chemical structure of glucopyranosiduronic isomers it can be proposed that some bacterial uridine 5'-diphospho-glucuronosyltransferases (UDPGT) catalyzed the reaction of conjugation of 6-(1-hydroxyethyl)-3,5-dimethyl-4a,5,6,7-tetrahydroquinazoline-2,4(1H,3H)-dione to glucuronic acid producing the identified glucopyranosiduronic acid derivatives. The whole reaction is summarized in the fig 4. The detection of *Androst-8-c-3-one-17,19 diacetoxo 4,4 dimethyl* in the sample DC72 as well as the identification of retinol in the sample DC 59 are also indicative of products of animal origin such as milk [25]. However, the identification of fatty acids such as stearic, palmitic (samples DC72 & DC54) as also the presence of oleanolic acid in all samples (Fig2), suggest that beside milk, some oil was also in the amphoraes. Although olive skin contains the highest level of oleanolic acid, making it thus a likely source, the specification of the oil origin will require a more rigorous fatty acids quantification [26][27]. The presence of some other natural compounds in the sample DC72 is also confirmed from the identification of triterpenes such as squalene [28,29] and barrigenol [30] (table 2). LC-MS/MS analysis was carried out for further confirmation.

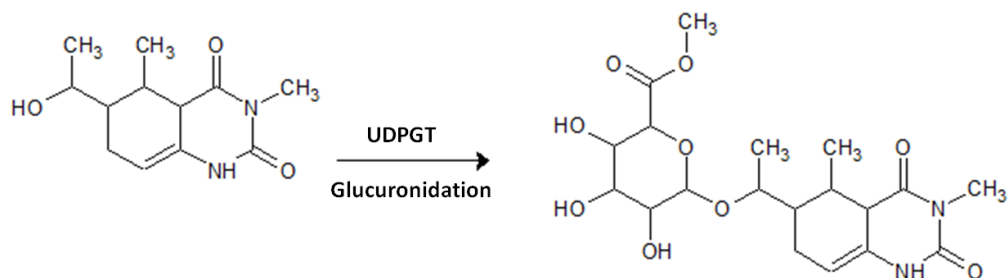


Figure 4: Glucuronidation of 6-(1-hydroxyethyl)-3,5-dimethyl-4a,5,6,7-tetrahydroquinazoline-2,4(1H,3H)-dione by formation of a β -glucosidic bond leading to the identified compound: D-Glucopyranosiduronic acid, 6-(1-hydroxyethyl)-3,5-dimethyl-4a,5,6,7-tetrahydroquinazoline-2,4(1H,3H)-dione.

Table 2: GC-MS analysis of sample DC_72. Identification was done by comparison of the retention time and mass spectra to those of standards in the instrument manufacture database NIST MS Search 2.0.

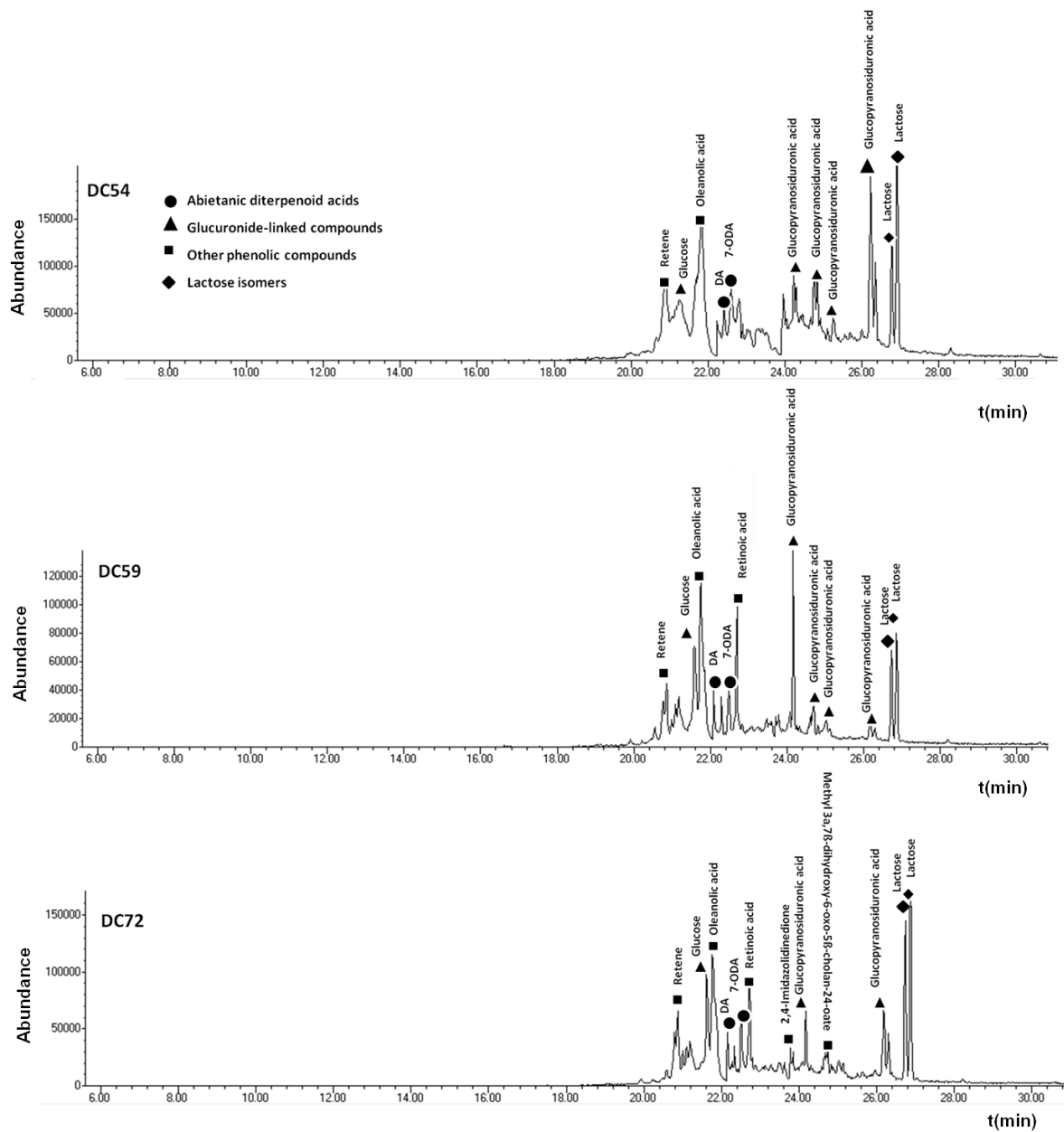
RT (min)	Name
14,14	2-tert-butyl-4-(1,1,3,3-tetramethylbutyl) phenol
16,36	Palmitic acid, methyl ester
17,64	4b,8-dimethyl-2-isopropylpheanthrene,
17,94	7-isopropyl-1-methyl-1,2,3,4,4a,9,10,10a-octahydrophenanthrene
18,78	Stearic acid, methyl ester
19,35	5,7,9(11)-androstatriene,3-hydroxy-17-oxo-
20,23	7-isopropyl-1,1,4°-trimethyl-1,2,3,4,4a,9,10,10a-octahydrophenanthrene
20,92	Podocarpa-6,8,11,13-tetraen-15-oic acid
21,50	R1-Barrigenol
22,25	Diglycidyl bisphenol A
23,00	Methyl 7-oxoabieta-9(11),8(14),12-trien-18-oate
24,19	Androst-8-c-3-one-17,19 diacetoxy4,4 dimethyl
24,72	Squalene

Table 3: GC-MS analysis of sample DC_54 Identification was done by comparison of the retention time and mass spectra to those of standards in the instrument manufacture database NIST MS Search 2.0.

RT(min)	Name
16,39	Palmitic acid, methyl ester
17,66	4b,8-dimethyl-2-isopropylphenanthrene
17,96	7-isopropyl-1-methyl-1,2,3,4,4a,9,10,10a-octahydrophenanthrene
18,80	Stearic acid, methyl ester
20,94	Podocarpa-6-8-11-13-trien-15-oic acid, 13-isopropyl-, methyl ester
21,53	19-nor-5,7,9(10)-pregnatrien-3-ol,20-carbomethoxy
23,00	Methyl 7-oxoabieta-9(11),8(14),12-trien-18-oate

Table 4: GC-MS analysis of sample DC_59 Identification was done by comparison of the retention time and mass spectra to those of standards in the instrument manufacture database NIST MS Search 2.0.

RT(min)	Name
17,13	Pimaric acid
17,88	10,18-bisnorabieta-8,11,13-triene
18,45	16-octadecenoic acid, methyl ester
19,47	1-phenanthrenecarboxy- tetradecahydro-7[2-methoxy-2-oxoethylidene)-1,4a,8-trimethyl-9-oxo-,methyl ester]
19,78	Phenanthrene,1-methyl-7-(1-methylethyl)-phenanthrene,7-isopropyl-1-methyl
20,18	Retinol
20,89	1-phenanthrenecarboxylic acid
21,46	9,10-anthracenedione,2-(1,1-dimethylethyl)-anthraquinone,2-tert-butyl
21,71	1-phenanthrenecarboxylic acid
21,87	Gibb-2-ene-1,10-dicarboxylic acid



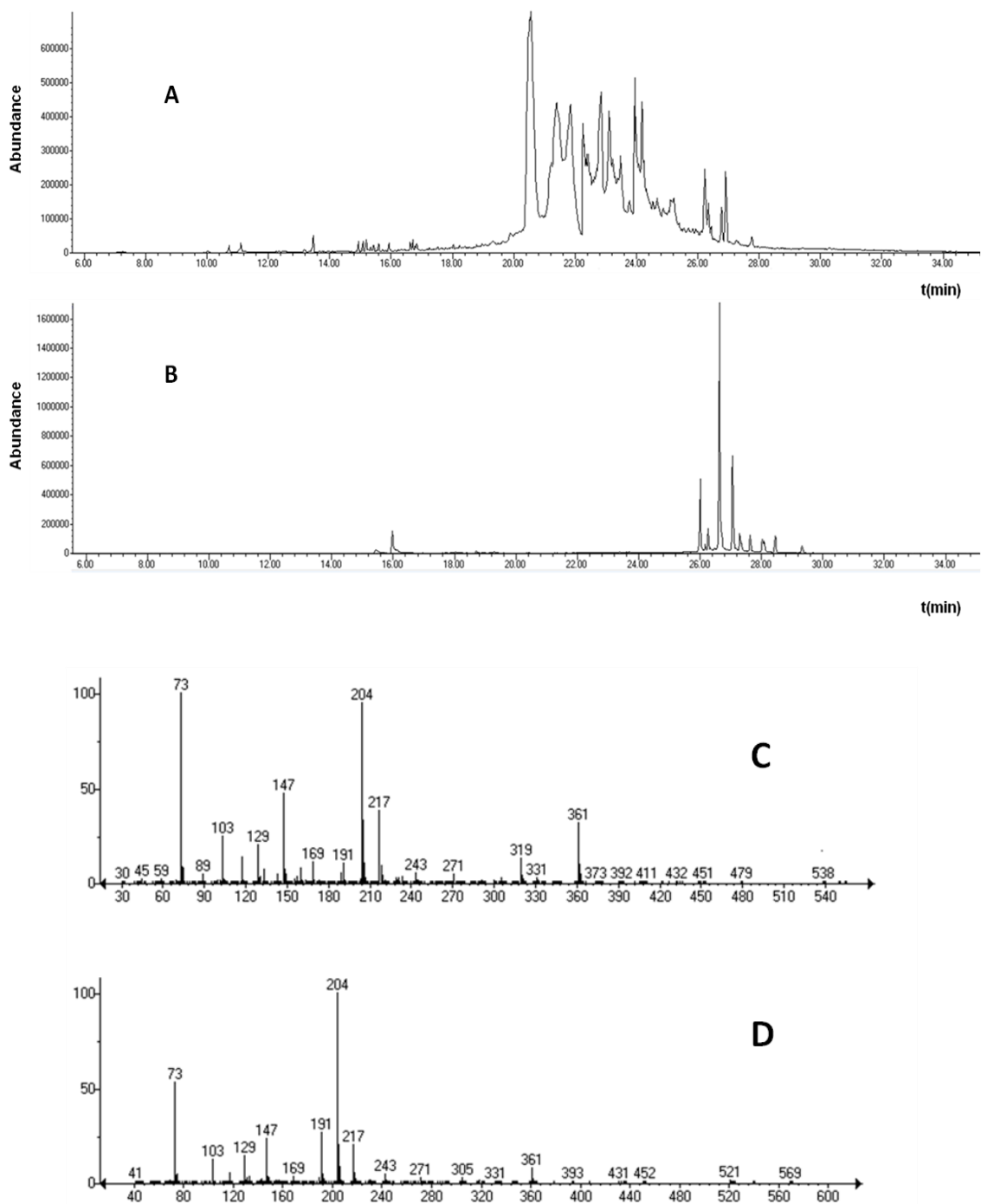


Figure 6: Lactose identification by GC/MS analysis in sample DC54. In the upper panels comparison of retention time between the sample DC54 (A) and the analytical standard of lactose (B) is reported. In the lower panels the MS spectrum one of lactose isomers (26.8 min) in the sample DC54 (C) is compared to the MS spectrum of the same isomer in the analytical standard of lactose (D). All compounds are TMS derivatives.

LC-MS/MS analysis

The results of FTIR and GC-MS analysis were integrated with a proteomic analysis, that confirmed the aforementioned results, since in the samples DC 54 and DC 59 beta galactosidase was confidently identified, thus definitively assessing the presence of milk. Beta galactosidases are glycoside hydrolase enzymes that catalyze the hydrolysis of β -galactosides into monosaccharides through the breaking of a glycosidic bond [31]. In the sample DC 72 no proteins were identified confidently, in agreement with the FT-IR analysis.

Table 6: Proteins identified in the samples DC 54 and DC 59. Raw data were searched by Mascot MS/MS Ion search using SwissProt database. No fixed chemical modification was inserted, and oxidation of methionine and the transformation of N-terminal glutamate or glutamine to pyroglutamate were considered as variable modifications. Individual ion score threshold provided by MASCOT software (>15) to evaluate the quality of matches in MS/MS data was used for confidence threshold in protein identification.

Sample	Identified protein (Accession number)	Score	Sequence coverage (%)	Matched sequence
DC54	Beta-galactosidase <i>Escherichia coli</i> (P00722)	109	14	L.ITTAHAWQHQGK.T R.LPSEFDLSAFLR.A M.TMITDSLAVVLQR.R V.LITTAHAWQHQGK.T R.LSGQTIEVTSEYLFH.H R.LQGGFVWDWVDQSLIK.Y R.VTVSLWQGETQVASGTAPFGGEIIDER .G
DC59	Beta-galactosidase <i>Escherichia coli</i> (P00722)	57	11	M.TMITDSLAVVLQR.R + Oxidation (M) R.LSGQTIEVTSEYLFH.H R.LQGGFVWDWVDQSLIK.Y R.YGLYVVDEANIETHGMVPMNR.L + 2 Oxidation (M) R.VVQPNATAWSEAGHISAWQQWR.L R.VTVSLWQGETQVASGTAPFGGEIIDER .

Conclusions

A multidisciplinary and multi-methodological approach that integrates the results obtained from FT-IR, GC-MS and LC-MS/MS analyses allowed to identify the remains of organic materials on the amphora shard of the Hera temple in Paestum. Using as pre-screening step the FT-IR analyses, samples were separated in three groups according to the presence of lipids, sugars and proteins. More specifically, in the first group we hypothesize the presence of organic compounds such as lipids and sugars while the other two groups showed some bands beyond 3000 cm^{-1} ascribable to the presence of some proteins. One sample was selected as representative from each group and subjected to GC-MS and proteomic analyses.

From GC-MS analysis, the resins used in the impermeabilization of the vessels were identified. Several abietic compounds and their products of degradation allowed to attribute the resin to *Pinaceae* family. Moreover, androsteroid compounds and glucuronide metabolites, and the unambiguous identification of lactose revealed that these amphorae contained milk. This result was further validated by the identification of beta galactosidase in the proteomic analysis [31]. Finally, the presence of oleanolic acid in all samples as also the co-presence of stearic and palmitic acid in two of them, suggest the use of oil, that might be either added to enhance the stiff properties of resins or has been mixed with milk [32][33]. Both hypotheses are plausible. Therefore, the results from the combination of spectroscopic and spectrometric analytical techniques propose the use of amphorae for some conservation of milk, aromatized with other natural compounds. These results are a small piece of the puzzle and contribute to disclose rituals during solemn ceremonies in the undiscovered temple of Hera in Paestum.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 722606, TEMPERA (Teaching Emerging Methods in Palaeoproteomics for the European Research Area) and from POR, Parco Archeologico Urbano di Napoli (PAUN).

References

- [1] G. Giorgi, Overview of Mass Spectrometric Based Techniques Applied in the Cultural Heritage Field, in: *Org. Mass Spectrom. Art Archaeol.*, John Wiley & Sons, Ltd, 2009: pp. 37–74. <https://doi.org/10.1002/9780470741917.ch2>.
- [2] I. Degano, F. Modugno, I. Bonaduce, E. Ribechini, M.P. Colombini, Recent Advances in Analytical Pyrolysis to Investigate Organic Materials in Heritage Science, *Angew. Chemie Int. Ed.* 57 (2018) 7313–7323. <https://doi.org/10.1002/anie.201713404>.
- [3] J. Blaško, R. Kubinec, B. Husová, P. Přikryl, V. Pacáková, K. Štulík, J. Hradilová, Gas chromatography/mass spectrometry of oils and oil binders in paintings, *J. Sep. Sci.* 31 (2008) 1067–1073. <https://doi.org/10.1002/jssc.200700449>.
- [4] S. Vahur, A. Teearu, P. Peets, L. Joosu, I. Leito, ATR-FT-IR spectral collection of conservation materials in the extended region of 4000–80 cm^{-1} , *Anal. Bioanal. Chem.* 408 (2016) 3373–9. <https://doi.org/10.1007/s00216-016-9411-5>.
- [5] N. Salvadó, S. Butí, M.J. Tobin, E. Pantos, A.J.N.W. Prag, T. Pradell, Advantages of the use of SR-FT-IR microspectroscopy: applications to cultural heritage., *Anal. Chem.* 77 (2005) 3444–51. <https://doi.org/10.1021/ac050126k>.
- [6] D. Lazidou, D. Lampakis, I. Karapanagiotis, C. Panayiotou, Investigation of the Cross-Section Stratifications of Icons Using Micro-Raman and Micro-Fourier Transform Infrared (FT-IR) Spectroscopy, *Appl. Spectrosc.* 72 (2018) 1258–1271. <https://doi.org/10.1177/0003702818777772>.
- [7] C. Daher, L. Bellot-Gurlet, A.-S. Le Hô, C. Paris, M. Regert, Advanced discriminating criteria for natural organic substances of cultural heritage interest: spectral decomposition and multivariate analyses of FT-Raman and FT-IR signatures., *Talanta.* 115 (2013) 540–7. <https://doi.org/10.1016/j.talanta.2013.06.014>.
- [8] M. I. M., *Atti e Memorie della Società Magna Grecia, 1926–30 (Vols. I–IV)*. Rome: Palazzo Taverna. Annual Subscription, 100 lire. - *L'opera della Società Magna Grecia nei primi 10 anni*. By U. Zanotti-Bianco. From *Annales Institutorum*, 1931. Rome: Palazzo Ricci, 19, *J. Hell. Stud.* (1932). <https://doi.org/10.2307/627206>.
- [9] P. Zancani Montuoro, *Heraion alla force del Sele: I: Altre metope del primo thesauros*, *Atti e Mem. Della Soc. Magna Grecia.* (1964).
- [10] P. Cennamo, M. Rosaria, B. Lumaga, G. Fatigati, A. Amoresano, A. Carpentieri, CONSERVATION SCIENCE A MULTIDISCIPLINARY ASSESSMENT TO INVESTIGATE A XXII, 11 (2020) 25–38.
- [11] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: From bones to paintings ☆, *Microchem. J.* 126 (2016) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.
- [12] S. Rossi, M. Petrelli, D. Morgavi, F.P. Vetere, R.R. Almeev, R.L. Astbury, D. Perugini, Role of magma mixing in the pre-eruptive dynamics of the Aeolian Islands volcanoes (Southern Tyrrhenian Sea, Italy), *Lithos.* 324–325 (2019) 165–179. <https://doi.org/10.1016/j.lithos.2018.11.004>.
- [13] B. Jananee, V. Thangam, A. Rajalakshmi, Investigation of soils by thermal and spectroscopic analysis, *Chem. Eng. Commun.* (2019). <https://doi.org/10.1080/00986445.2019.1680370>.
- [14] R.J. Cox, H.L. Peterson, J. Young, C. Cusik, E.O. Espinoza, The forensic analysis of soil organic by FTIR, *Forensic Sci. Int.* 108 (2000) 107–116. [https://doi.org/10.1016/S0379-0738\(99\)00203-0](https://doi.org/10.1016/S0379-0738(99)00203-0).
- [15] M. Cavagna, R. Dell'Anna, F. Monti, F. Rossi, S. Torriani, Use of ATR-FTIR microspectroscopy to monitor autolysis of *Saccharomyces cerevisiae* cells in a base wine, *J. Agric. Food Chem.* 58 (2010) 39–45. <https://doi.org/10.1021/jf902369s>.
- [16] S.H. Gordon, S.C. Whatel, B.C. Wheeler, C. James, Multivariate FTIR analysis of substrates for protein, polysaccharide, lipid and microbe content: Potential for solid-state fermentations, *Biotechnol. Adv.* 11 (1993) 665–675. [https://doi.org/10.1016/0734-9750\(93\)90035-L](https://doi.org/10.1016/0734-9750(93)90035-L).
- [17] B. Wiyono, S. Tachibana, D. Tinambunan, Chemical Compositions of Pine Resin, Rosin and Turpentine

- Oil from West Java, Indones. J. For. Res. (2006). <https://doi.org/10.20886/ijfr.2006.3.1.7-17>.
- [18] R.P. Evershed, C. Heron, L. John Goad, Analysis of organic residues of archaeological origin by high-temperature gas chromatography and gas chromatography - Mass spectrometry, *Analyst*. (1990). <https://doi.org/10.1039/an9901501339>.
- [19] M.D. Petit-Dominguez, J. Martinez-Maganto, MCF fast derivatization procedure for the identification of resinous deposit components from the inner walls of roman age amphorae by GC-MS, *Talanta*. (2000). [https://doi.org/10.1016/S0039-9140\(99\)00330-6](https://doi.org/10.1016/S0039-9140(99)00330-6).
- [20] C. Azemard, M. Menager, C. Vieillescazes, Analysis of diterpenic compounds by GC-MS/MS: contribution to the identification of main conifer resins, *Anal. Bioanal. Chem.* 408 (2016) 6599–6612. <https://doi.org/10.1007/s00216-016-9772-9>.
- [21] J. Font, N. Salvadó, S. Butí, J. Enrich, Fourier transform infrared spectroscopy as a suitable technique in the study of the materials used in waterproofing of archaeological amphorae, *Anal. Chim. Acta.* 598 (2007) 119–127. <https://doi.org/10.1016/j.aca.2007.07.021>.
- [22] I. Jerković, Z. Marijanović, M. Gugić, M. Roje, Chemical profile of the organic residue from ancient amphora found in the adriatic sea determined by direct GC and GC-MS analysis, *Molecules*. 16 (2011) 7936–7948. <https://doi.org/10.3390/molecules16097936>.
- [23] F.J. Moreno, A. Montilla, M. Villamiel, N. Corzo, A. Olano, Analysis, structural characterization, and bioactivity of oligosaccharides derived from lactose, *Electrophoresis*. 35 (2014) 1519–1534. <https://doi.org/10.1002/elps.201300567>.
- [24] G.J. Dutton, Glucuronide-Forming Enzymes, in: *Concepts Biochem. Pharmacol.*, Springer Berlin Heidelberg, 1971: pp. 378–400. https://doi.org/10.1007/978-3-642-65177-9_24.
- [25] A. Goyon, J.Z. Cai, K. Kraehenbuehl, C. Hartmann, B. Shao, P. Mottier, Determination of steroid hormones in bovine milk by LC-MS/MS and their levels in Swiss Holstein cow milk, *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* 33 (2016) 804–816. <https://doi.org/10.1080/19440049.2016.1175186>.
- [26] J. Orsavova, L. Misurcova, J. Vavra Ambrozova, R. Vicha, J. Mlcek, Fatty acids composition of vegetable oils and its contribution to dietary energy intake and dependence of cardiovascular mortality on dietary intake of fatty acids, *Int. J. Mol. Sci.* 16 (2015) 12871–12890. <https://doi.org/10.3390/ijms160612871>.
- [27] J. Pollier, A. Goossens, Oleanolic acid, *Phytochemistry*. 77 (2012) 10–15. <https://doi.org/10.1016/j.phytochem.2011.12.022>.
- [28] M.A. Lozano-Grande, S. Gorinstein, E. Espitia-Rangel, G. Dávila-Ortiz, A.L. Martínez-Ayala, Plant Sources, Extraction Methods, and Uses of Squalene, *Int. J. Agron.* 2018 (2018). <https://doi.org/10.1155/2018/1829160>.
- [29] S.K. Kim, F. Karadeniz, Biological Importance and Applications of Squalene and Squalane, in: *Adv. Food Nutr. Res.*, Academic Press Inc., 2012: pp. 223–233. <https://doi.org/10.1016/B978-0-12-416003-3.00014-7>.
- [30] X. Zhang, S. Zhang, Y. Yang, D. Wang, H. Gao, Natural barrigenol-like triterpenoids: A comprehensive review of their contributions to medicinal chemistry, *Phytochemistry*. 161 (2019) 41–74. <https://doi.org/10.1016/j.phytochem.2019.01.017>.
- [31] P.D. Uchil, A. Nagarajan, P. Kumar, β -galactosidase, *Cold Spring Harb. Protoc.* 2017 (2017) 774–779. <https://doi.org/10.1101/pdb.top096198>.
- [32] M.P. Colombini, G. Giachi, F. Modugno, E. Ribechini, Characterisation of organic residues in pottery vessels of the Roman age from Antinoe (Egypt), in: *Microchem. J.*, 2005: pp. 83–90. <https://doi.org/10.1016/j.microc.2004.05.004>.
- [33] J. Cuní, P. Cuní, B. Eisen, R. Savizky, J. Bové, Characterization of the binding medium used in Roman encaustic paintings on wall and wood, *Anal. Methods*. 4 (2012) 659–669. <https://doi.org/10.1039/c2ay05635f>.
- [34] A. Pichot, Études de lettres Theatra et spectacula Théâtres et amphithéâtres : outils de romanisation

en Maurétanie ?, in: [Http://Journals.Openedition.Org/Edl](http://Journals.Openedition.Org/Edl), Faculté des lettres de l'Université de Lausanne, 2011: pp. 171–192. <https://doi.org/10.4000/edl.113>.

[35] G. Renda, S. Gigli, A. Amato, S. Venticinque, B. Di Martino, F.R. Cappa, Mobile devices for the visit of anfiteatro campano in Santa Maria Capua Vetere, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Springer, Berlin, Heidelberg, 2012: pp. 281–290. https://doi.org/10.1007/978-3-642-34234-9_28.

[36] M. Gelzo, M. Grimaldi, A. Vergara, V. Severino, A. Chambery, A. Dello Russo, C. Piccioli, G. Corso, P. Arcari, Comparison of binder compositions in Pompeian wall painting styles from Insula Occidentalis, *Coke Chem.* 8 (2014) 65. <https://doi.org/10.1186/s13065-014-0065-0>.

[37] M. Gelzo, M. Grimaldi, A. Vergara, V. Severino, A. Chambery, A. Dello Russo, C. Piccioli, G. Corso, P. Arcari, Comparison of binder compositions in Pompeian wall painting styles from Insula Occidentalis, *Chem. Cent. J.* 8 (2014) 65. <https://doi.org/10.1186/s13065-014-0065-0>.

[38] G. Corso, M. Gelzo, A. Chambery, V. Severino, A. Di Maro, F.S. Lomoriello, O. D'Apolito, A. Dello Russo, P. Gargiulo, C. Piccioli, P. Arcari, Characterization of pigments and ligands in a wall painting fragment from Liternum archaeological park (Italy), *J. Sep. Sci.* 35 (2012) 2986–2993. <https://doi.org/10.1002/jssc.201200490>.

[39] A. Casoli, S. Santoro, Organic materials in the wall paintings in Pompei: A case study of Insula del Centenario, *Chem. Cent. J.* 6 (2012) 107. <https://doi.org/10.1186/1752-153X-6-107>.

[40] J. Cuní, What do we know of Roman wall painting technique? Potential confounding factors in ancient paint media analysis, *Herit. Sci.* 4 (2016) 44. <https://doi.org/10.1186/s40494-016-0111-4>.

[41] L. Birolo, A. Tomeo, M. Trifuoggi, F. Auriemma, L. Paduano, A. Amoresano, R. Vinciguerra, C. De Rosa, L. Ferrara, A. Giarra, A. Luchini, C. De Maio, G. Greco, A. Vergara, A hypothesis on different technological solutions for outdoor and indoor Roman wall paintings, *Archaeol. Anthropol. Sci.* 9 (2017) 591–602. <https://doi.org/10.1007/s12520-016-0408-y>.

[42] S. Gunasekaran, G. Anbalagan, S. Pandi, Raman and infrared spectra of carbonates of calcite structure, *J. Raman Spectrosc.* 37 (2006) 892–899. <https://doi.org/10.1002/jrs.1518>.

[43] H.M. Mahmoud, A MULTI-ANALYTICAL APPROACH FOR CHARACTERIZING PIGMENTS FROM THE TOMB OF DJEHUTYEMHAB (TT 194), (2013).

[44] H.H. Marey Mahmoud, M. Mahmoud, H. H., A Preliminary Investigation Of Ancient Pigments From The Mortuary Temple Of Seti I, El-Qurna (Luxor, Egypt), *MAA.* 11 (2011) 99–106. <https://ui.adsabs.harvard.edu/abs/2011MAA....11...99M/abstract> (accessed December 5, 2020).

[45] F. Ospitali, D. Bersani, G. Di Lonardo, P.P. Lottici, “Green-earths”: vibrational and elemental characterisation of glauconites, celadonites and historical pigments, (2007) 40–41.

[46] R.J.H. Clark, P.J. Gibbs, K.R. Seddon, N.M. Brovenko, Y.A. Petrosyan, Non-destructive in situ identification of cinnabar on ancient Chinese manuscripts, *J. Raman Spectrosc.* 28 (1997) 91–94. [https://doi.org/10.1002/\(sici\)1097-4555\(199702\)28:2/3<91::aid-jrs67>3.0.co;2-x](https://doi.org/10.1002/(sici)1097-4555(199702)28:2/3<91::aid-jrs67>3.0.co;2-x).

[47] P. Baraldi, C. Baraldi, R. Curina, L. Tassi, P. Zannini, A micro-Raman archaeometric approach to Roman wall paintings, *Vib. Spectrosc.* 43 (2007) 420–426. <https://doi.org/10.1016/j.vibspec.2006.04.029>.

[48] E.A. Lalla, G. Lopez-Reyes, A. Sansano, A. Sanz-Arranz, J. Martínez-Frías, J. Medina, F. Rull-Pérez, Raman-IR vibrational and XRD characterization of ancient and modern mineralogy from volcanic eruption in Tenerife Island: Implication for Mars, *Geosci. Front.* 7 (2016) 673–681. <https://doi.org/10.1016/j.gsf.2015.07.009>.

[49] M. Alfè, V. Gargiulo, R. Di Capua, F. Chiarella, J.N. Rouzaud, A. Vergara, A. Ciajolo, Wet chemical method for making graphene-like films from carbon black, *ACS Appl. Mater. Interfaces.* 4 (2012) 4491–4498. <https://doi.org/10.1021/am301197q>.

[50] D.C. Smith, M. Bouchard, M. Lorblanchet, An initial Raman microscopic investigation of prehistoric rock art in caves of the Quercy District, S. W. France, *J. Raman Spectrosc.* 30 (1999) 347–354. [https://doi.org/10.1002/\(SICI\)1097-4555\(199904\)30:4<347::AID-JRS379>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-4555(199904)30:4<347::AID-JRS379>3.0.CO;2-A).

- [51] D.L.A. De Faria, S. Venâncio Silva, M.T. De Oliveira, Raman microspectroscopy of some iron oxides and oxyhydroxides, *J. Raman Spectrosc.* 28 (1997) 873–878. [https://doi.org/10.1002/\(sici\)1097-4555\(199711\)28:11<873::aid-jrs177>3.0.co;2-b](https://doi.org/10.1002/(sici)1097-4555(199711)28:11<873::aid-jrs177>3.0.co;2-b).
- [52] S. Das, M.J. Hendry, Application of Raman spectroscopy to identify iron minerals commonly found in mine wastes, *Chem. Geol.* 290 (2011) 101–108. <https://doi.org/10.1016/j.chemgeo.2011.09.001>.
- [53] I.M. Bell, R.J.H. Clark, P.J. Gibbs, Raman spectroscopic library of natural and synthetic pigments (pre- \approx 1850 AD), *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 53 (1997) 2159–2179. [https://doi.org/10.1016/S1386-1425\(97\)00140-6](https://doi.org/10.1016/S1386-1425(97)00140-6).
- [54] J. Jehlička, H.G.M. Edwards, K. Osterrothová, J. Novotná, L. Nedbalová, J. Kopecký, I. Němec, A. Oren, Potential and limits of Raman spectroscopy for carotenoid detection in microorganisms: Implications for astrobiology, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 372 (2014). <https://doi.org/10.1098/rsta.2014.0199>.
- [55] Ugo Zezza, *Petrografia microscopica*, La Goliardica Pavese, 1976.
- [56] A.P.D. Peccerillo, *Introduzione alla petrografia ottica*, Morlacchi Editore, 2003. <https://www.amazon.it/Introduzione-alla-petrografia-ottica-CD-ROM/dp/8888778276> (accessed December 5, 2020).
- [57] N. Szczepańska, P. Kubica, B. Kudłak, J. Namieśnik, A. Wasik, Stabilities of bisphenol A diglycidyl ether, bisphenol F diglycidyl ether, and their derivatives under controlled conditions analyzed using liquid chromatography coupled with tandem mass spectrometry, *Anal. Bioanal. Chem.* 411 (2019) 6387–6398. <https://doi.org/10.1007/s00216-019-02016-5>.
- [58] V. Beltran, N. Salvadó, S. Butí, T. Pradell, Ageing of resin from Pinus species assessed by infrared spectroscopy, *Anal. Bioanal. Chem.* 408 (2016) 4073–4082. <https://doi.org/10.1007/s00216-016-9496-x>.
- [59] G. Kaklamanos, G. Theodoridis, T. Dabalís, Determination of anabolic steroids in bovine urine by liquid chromatography-tandem mass spectrometry, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 877 (2009) 2330–2336. <https://doi.org/10.1016/j.jchromb.2009.03.033>.
- [60] H. Zhang, J.M. Yue, Hasubanan type alkaloids from *Stephania longa*, *J. Nat. Prod.* 68 (2005) 1201–1207. <https://doi.org/10.1021/np0500833>.
- [61] P. Ranchana, M. Ganga, M. Jawaharlal, M. Kannan, Investigation of Volatile Compounds from the Concrete of *Jasminum auriculatum* Flowers, *Int.J.Curr.Microbiol.App.Sci.* 6 (2017) 1525–1531. <https://doi.org/10.20546/ijcmas.2017.611.180>.
- [62] H.M. Usman, N. Abdulkadir, M. Gani, H.M. Maiturare, Bacterial pigments and its significance, (2017). <https://doi.org/10.15406/mojbb.2017.04.00073>.
- [63] P. Sabbagh, A. Ebrahimzadeh Namvar, The eminence status of bacterial pigments under different aspects, *Microbiol. Medica.* 32 (2017). <https://doi.org/10.4081/mm.2017.7247>.
- [64] Il Palazzo - Sito web non ufficiale della Reggia di Caserta, (n.d.). <https://www.reggiadicasertaunofficial.it/it/reggia/il-palazzo/> (accessed November 24, 2020).
- [65] 18th-Century Royal Palace at Caserta with the Park, the Aqueduct of Vanvitelli, and the San Leucio Complex - Documents - UNESCO World Heritage Centre, (n.d.). <https://whc.unesco.org/en/list/549/documents/%23ABevaluation> (accessed November 24, 2020).
- [66] Dictionnaire amoureux de Versailles - Franck FERRAND - Google Libri, (n.d.). https://books.google.it/books?id=qTdIAQAAQBAJ&redir_esc=y (accessed November 24, 2020).
- [67] J. Mazurek, M. Svoboda, M. Schilling, GC/MS Characterization of Beeswax, Protein, Gum, Resin, and Oil in Romano-Egyptian Paintings, *Heritage.* 2 (2019) 1960–1985. <https://doi.org/10.3390/heritage2030119>.
- [68] Rapid differentiation of beeswaxes using EGA-MS and Py-GC / MS, (n.d.) 8862.
- [69] M.P. Colombini, G. Giachi, F. Modugno, E. Ribechini, Characterisation of organic residues in pottery vessels of the Roman age from Antinoe (Egypt), *Microchem. J.* 79 (2005) 83–90. <https://doi.org/10.1016/j.microc.2004.05.004>.
- [70] I. Jerković, Z. Marijanović, M. Gugić, M. Roje, Chemical profile of the organic residue from ancient

amphora found in the adriatic sea determined by direct GC and GC-MS analysis, *Molecules*. 16 (2011) 7936–7948. <https://doi.org/10.3390/molecules16097936>.

[71] Y. Zhang, G. Wu, C. Chang, Y. Lv, W. Lai, H. Zhang, X. Wang, Q. Jin, Determination of Origin of Commercial Flavored Rapeseed Oil by the Pattern of Volatile Compounds Obtained via GC–MS and Flash GC Electronic Nose, *Eur. J. Lipid Sci. Technol.* 122 (2020) 1–7. <https://doi.org/10.1002/ejlt.201900332>.

[72] A. Lluveras-Tenorio, J. Mazurek, A. Restivo, M.P. Colombini, I. Bonaduce, The Development of a New Analytical Model for the Identification of Saccharide Binders in Paint Samples, *PLoS One*. 7 (2012) e49383. <https://doi.org/10.1371/journal.pone.0049383>.

[73] S. Bogdanov, B.P. Science, Beeswax : Uses and Trade, *The Beeswax*. (2009) 1–16.

[74] C. Brøns, K.L. Rasmussen, M.M. Di Crescenzo, R. Stacey, A. Lluveras-Tenorio, Painting the Palace of Apries I: Ancient binding media and coatings of the reliefs from the Palace of Apries, Lower Egypt, *Herit. Sci.* 6 (2018) 1–20. <https://doi.org/10.1186/s40494-018-0170-9>.

5.4 Manuscript 4: Wall paintings from a Roman domus in Santa Maria Capua Vetere

Abstract

Pigments, ligands and mortars of wall paintings from a Roman Domus in Santa Maria Capua Vetere were studied via a multi-methodological approach. Optical, Raman and Scanning Electron microscopy were combined with GC-MS and LC-MS/MS analyses in order to achieve a comprehensive picture of the chemical components of some fine wall paintings recently excavated in Santa Maria Capua Vetere. Some archaeological indications suggest that the Domus' owner might be a relative of the Emperor August. The four-layers preparation, the rich pigment palette (both natural and synthetic ones) and the variety of the organic ligands used support the hypothesis of a very wealthy owner.

Author's contribution

Leila Birolo and Georgia Ntasi, designed GC-MS and LC-MS/MS experiments; Georgia Ntasi conducted GC-MS and LC-MS/MS experiments;

Introduction

The city of Santa Maria Capua Vetere is located in the ruins of ancient Capua, a city of Roman origin in southern Italy. The history of the city tells of an intermittent cycle of splendor and ruins. During Roman times the city lived a very prosperous era, so rich to be named Campania felix by the Romans. After the domination of the Oscii and Etruscans, in 841 the city was almost destroyed by the incursions by the Saracens. Thanks to its rich history, the city can boast the beauty of various archeological monuments such as: the Campano amphitheater, the arch of Hadrian and a Domus [34,35]. In 2018 some excavations brought to light a multi-level site consisting of a basement and a Roman villa. The molecular characterization of murals in terms of colors and/or organic binders, cross-referenced to literature data [33,36–40] can reveal important aspects about its welfare state.

Aim of study

Raman micro-spectroscopy, GC-MS and Optical microscopy were used to characterize and identify layers of mortars and a large variety of pigments and binders in order to draw a picture of the preparation technique. The data obtained can contribute to the diagnosis of the current conservation state of the wall paintings and give information on the welfare state of the Domus owners.

Materials and methods

Samples

In 2018, a house dating back to the Roman era was unearthed. Four fragments of wall paintings selected on site were analysed (187, 151 A, 151 B, 151 C), with special attention to pigments, binders and mortars.

Raman Microscopy

A confocal Raman microscope (Jasco, NRS-3100) was used to obtain Raman spectra. The 514 nm line of an air-cooled Ar⁺ laser (Melles Griot, 35 LAP431 220) was injected into an integrated Olympus microscope and focused to a spot diameter of approximately 3 μm by a 20x objective with a final 4 mW power at the sample. A holographic notch filter was used to reject the excitation laser line. Raman backscattering was collected using a diffraction lattice of 1200 grooves/mm and 0.01-0.20 mm slits, corresponding to an average spectral resolution up to 1 cm^{-1} . Because of the heterogeneity of the samples, acquisition time was not the same for all. In some cases a time of about 90 seconds was required, for others even of several minutes due to problems such as fluorescence. To collect a complete dataset from a Peltier-cooled 1024x128 pixel CCD 31 photon detector (Andor DU401BVI). Raman measurements were finally triplicated for the purpose of reproducibility for each spot sampled. Wavelength calibration was performed by using cyclohexane as a standard.

Optical Microscopy

Stereoscopic microscope and the polarized light microscope were used. The stereo-microscope used is an Axio Zoom V16 microscope, which mounts Plan- NEOFLUAR Z 1.0x/0.25 - FWD 56mm and Apo Z 1.5x FWD 30mm lenses, both with a parfocal length of 133 mm. The microscope has a built-in table for analysis in transmitted light and dark field. With Axio Zoom V16 it is possible to zoom large areas with the maximum depth of field with high magnifications at very high resolutions. The high depth of field and the high resolution, ideal for three-dimensional morphological studies, are given by the combination of 10x/23 zoom with a high numerical aperture of NA 0.25. Thanks to these particular optics it is possible to make measurements of micrometric structures with a reproducibility of 99%. The Axio Zoom V16 is equipped with a dedicated camera, Axio Cam ICc5. The image management software is Zeiss Axiovision 4.7 with active modules: Z-Stack, Extended Focus and Panorama. The polarized light microscope used is the Model Zeiss Axio Imager Am1 in transmitted light and reflected light, which mounts LC EC Epiplan-NEOFLUAR 20x/0.22 HD DIC, LC EC Epiplan-NEOFLUAR 50x/0 lenses.55 HD DIC (reflect light observations), EC Plan-NEOFLUAR

2.5x/0.075, EC Plan-NEOFLUAR 10x/0.3 Pol, EC Plan-NEOFLUAR 20x/0.5 Pol, EC Plan-NEOFLUAR 40x/0.9 Pol. The camera used is AxioCam ICc5 while the image management software is Zeiss Axiovision 4.7 with active module Panorama.

Organic molecule analyses (GC-MS)

Samples were fractionated, derivatized and analyzed as reported in [10].

Protein extraction

Because of a significant presence of metals and carbonates was observed by spectroscopic techniques, a demineralization step was performed as a pre-treatment step. In particular, about 300 μL of a solution of 0.5M EDTA pH~8 were added to wall fragments for 10 days at room temperature, refreshing the solution every 2 days. After centrifugation for 2 min at 10,000 rpm, the collected supernatant was rejected. After 10 days, the pellet was subjected to a denaturation step with Urea. Therefore, 10 μL of a solution of 6M Urea were added to micro-samples and incubated for 10 min at room temperature, followed by sonication for 20 min. Urea was then 6-fold diluted with water. Subsequently, trypsin was added to a final concentration of 0.1 $\mu\text{g}/\mu\text{L}$ in Ambic 10mM. After incubation at 37°C for 16 hours, the supernatants were recovered by centrifugation at 10,000 rpm, and the peptide mixture was filtered on 0.22 μm PVDF membrane (Millipore), concentrated and purified using a reverse-phase C18 Zip Tip pipette tip (Millipore). Peptides were eluted with 20 μL of a solution made of 50% Acetonitrile, 50% Formic acid 0.1% in Milli-Q water and analysed by LC-MS/MS

LC-MS/MS analyses

Samples were analyzed as reported in [9].

Results and Discussion

In 2018, a house dating back to the Roman era was unearthed. Two fragments of the wall paintings 151 and 187 selected on the site were analyzed with special attention to pigments, binders and mortars. Because of the color complexity the sample 151 was subdivided in three layers (151 A, 151 B, 151 C).

The wall paintings represent an example of preparative techniques adopted by Romans in a house probably belonging to an August Emperor-related owner. Several methodologies were used to characterize a) four layers of mortars (optical microscopy), b) a large variety of pigments, such as Egyptian blue, green earth,

cinnabar, red ochre, amorphous carbon (Raman and optical microscopy), c) several organic binders (GC-MS). The preparation of these wall paintings looks very articulated and rich of chemical components (both natural and synthetic), indicating prestige and wealthy of the domus owner.

Raman Analysis

Only the pictorial layers (first level) of the samples were investigated by Raman spectroscopy for pigment analysis (samples 187, 151A, 151B, 151C). The spectra are shown in Figure 1 and summarized in Table 1.

The samples consist of well-preserved colored fragments of red, yellow, blue, green and white.

The Raman spectra demonstrated blue and green. The blue spots of samples 151A exhibit Raman peaks at 1089 cm^{-1} and $430, 467, 572, 757, 989, 1015, 1087\text{ cm}^{-1}$, respectively assigned to *Carbonate* (trace 7) [41][42] and *Cuprorivaite* ($\text{CuSi}_4\text{O}_{10}$) famous as Egyptian blue (trace 1) [43][44].

The green spot of the sample 151A shows signals belonging to the “Green earths” (trace 3). In particular, signals are attributable to Glauconite. The chemical composition of *Glauconite* is approximately $(\text{K,Na})(\text{Fe}^{3+},\text{Al},\text{Mg})_2(\text{Si,Al})_4\text{O}_{10}(\text{OH})_2$, similar to *Celadonite*, but with some Al due to a partial substitution of Al for Si in the tetrahedrally coordinated [45].

The samples 151A, 151B and 151C show different shades of red. The reddish colour, often tending to yellow or brown, appear to be composed of iron oxides and hydroxides, while the layers of bright red colour are mainly made up of *Cinnabar* (HgS) immersed in a carbonate matrix.

As shown in trace 2, the red spot belonging to 151A shows very strong signals at $254, 287$ and 344 cm^{-1} , easily assigned to HgS [46] [47], while, both the red spot of 151B and 151C show a tonality given by several components. In these samples there are signals attributable to both *Hematite and Goethite* [48,49] (trace 6 and trace 8), the main difference being in the presence of a third component in the sample 151 B (signal at 1296 cm^{-1} , a signal at 1611 cm^{-1} of lower intensity that can be assigned to amorphous carbon (trace 6) [49,50]). Here, the inversion of intensity between the D and G bands of coal is due to the overlap of the latter with Hematite.

It must be pointed out that the Goethite was not detected by SEM analysis. This might be probably caused by a partial “*Goethite-Hematite*” conversion, possibly due to the laser exposure [48,51,52]. It is worth noting that most likely goethite itself is a product of oxidation of an initial hematite, so accidentally the laser-induced Raman assignment can correspond to the actual starting pigment.

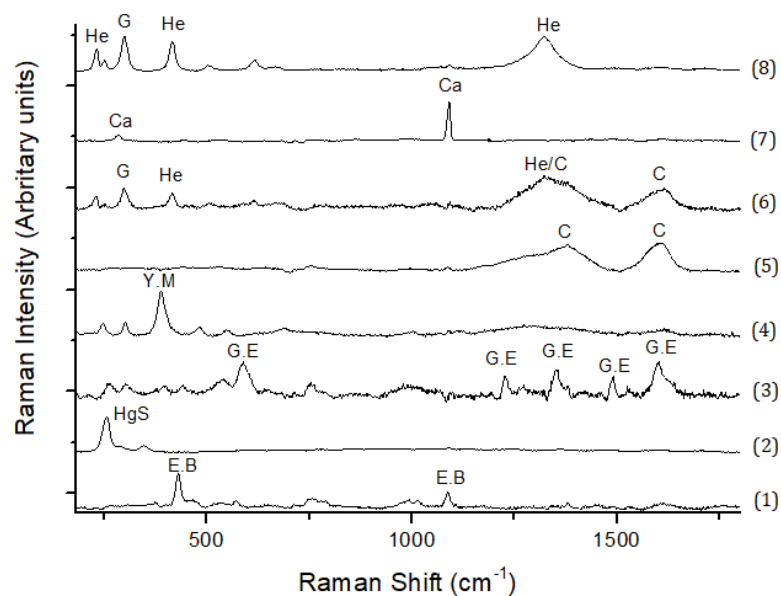


Figure 1: Representative Raman spectra of samples from S.M. Capua Vetere reported in Table 3.2.1. He (hematite); C (carbon); Ca (carbonate); E.B (Egyptian blue); G.E (Green earths); HgS (Cinnabar); G (Goethite).

The white spot of the 151 C sample can be easily assigned as carbonate because of the peaks at 285 and 1091 cm^{-1} (trace 7), however the latter could also be a component of the second layer and not of the first one. Small crystals of magnetite and pyrite were also found in the samples (data not shown). The yellow spot of the sample 151 B exhibits Raman bands typical of $\text{Fe}(\text{OH})_3$, whose synthetic counterpart (typical of the 19th century) is known as *Mars yellow* (trace 4) [53]. In this case, Raman analysis only reveals one minor component of the pigment, in fact the elemental analysis of the yellow pigments are highly heterogeneous most likely corresponding to the mineral limonite (mixture of hydroxide/oxides of iron, zinc and titan).

Although the spectra of samples 187 do not show Raman bands of any pigments, they are characterized by trace of β -carotene, indicating a poor biodeterioration of the pictorial layer (see Table 1) [54].

Table 1: Summary of the results by Raman micro-spectroscopy of wall painting samples from S.M. Capua Vetere. s (strong), m (medium), w (weak), sh (shoulder), br (broad), vs (very strong), vw (very weak). Raman traces in Figure 1

Sample	Spot	Raman bands (cm-1)	Putative assignment	Reference
187	Green	967(w); 1010(m); 1161(vs); 1193(m); 1275(w); 1358(w); 1395(vw); 1453(w); 1527(vs)	carotenoids	[54]
151 A	Blue	430(s); 467(w); 572(w); 757(w); 989(w); 1015(w); 1087(m)	CaCuSi ₄ O ₁₀	[43,44]
	Light Red	257(s); 284(w); 350(m); 1089(w)	HgS, Carbonate	[42,46,47]
	Green	542(m); 589(s); 753(w); 1226(m); 1355(m); 1491(m); 1602(s)	Green Earths	[45]
151 B	Yellow	246(m); 301(m); 388(vs); 482(m); 547(m); 687(m)	Fe(OH) ₃ , Amorphous carbon	[49,53]
	Dark	1381(s); 1606(s)	Amorphous carbon	[49]
		230(m); 297(m); 415(m); 615(w); 1322(s); 1613(m)	Amorphous carbon, Hematite, Goethite	[48,49,51,52]
	White	275(w); 1089(s)	Calcite	[42]
	Light Red	247(w); 301(m); 390(s); 480(w); 551(m); 683(w); 1002(w); 1296(m)	Hematite; Goethite	[48,51,52]
151 C	White	285(m); 1091(vs)	Calcite	[42]
	Dark Red	232(m); 251(w); 416(vs); 670(w); 506(m); 617(m); 1086(w); 1322(vs)	Hematite, Calcite	[51,52]

Analysis with stereo microscope and petrographic microscope

The samples were analysed in two different perspectives, the first is a macroscopic observation on 3D materials with a stereo microscope, the second is a petrographic observation on thin section with the petrographic microscope. The experiments herein reported were carried out by Dr. Emanuela Rossi from Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse. The SEM-EDS data, collected previously by Dr. Rossi, were included to complete the thesis discussion. The samples are made up of wall fragments inclusive of the pictorial layers. The layers have different thicknesses, textures and mineralogical compositions (Figure 2 and Figure 3).



Figure 2: Macroscopic image of sample 151

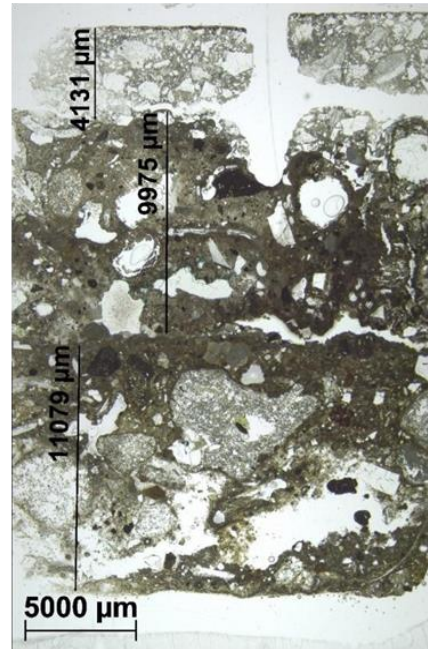


Figure3: Thin section image in transmitted light of sample 151 C

The fragments are generally composed of three layers. The first layer consists in a pictorial level with variable colors (level 1), the second layer has a very light color and a homogeneous texture at first sight (level 2), the third layer has a darker color and is characterised by the presence of a matrix with different colors and geotechnical characteristics (level 3). In the 151 C sample there are four layers (Figure 3) the first three layers are equivalent to those mentioned above, the fourth layer is very similar to the third but has a yellowish color and is more compact (level 4). The 187 sample consists only of level 3.

Level 1

Generally, the level 1 has a variable thickness between 300 and 11 μm with different, sometimes overlapping, colors (Figure 4). The most common color is red, in various shades, but there are also always green, in various shades, yellow, black, and a very light pink as well as white slightly yellowed. Often the same sample has different colors simultaneously (Figure 5).

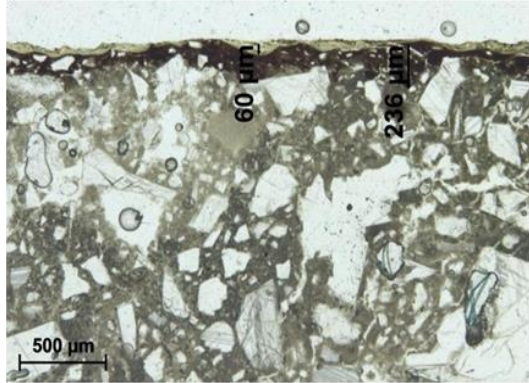


Figure 4: Thin section image in transmitted light of sample 151 C. Pictorial level with different colors sometimes overlapping.

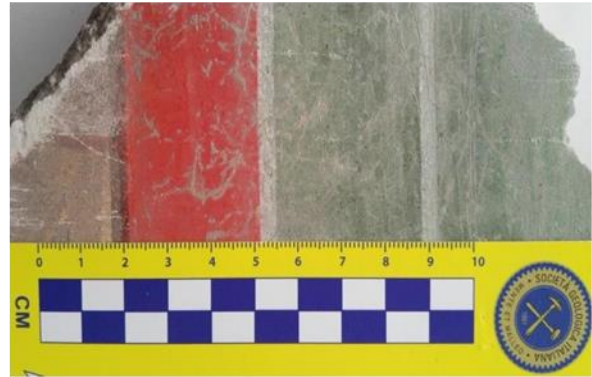


Figure 5: Macroscopic image shows different colors of frescoed in 151 samples.

The portions of the frescoed walls analysed present locally incrustations of soil and/or profound incisions sometimes parallel to each other (Figure 6), in other cases there are holes of different depths and extensions (Figure 7), in which the layer of paint has been completely removed. In the reddish pigments there are small opaque fragments of black color, randomly diffused. In the areas where the green color is present, we can observe blue fragments and some small black fragments (Figure 8).



Figure 6: Stereomicroscope image of 151 C sample. Reddish frescoed covered with impurity and with parallel incision.



Figure 7: Stereomicroscope image of level 1 in 151 B sample. The hole shows the presences of yellow layer under the red one.

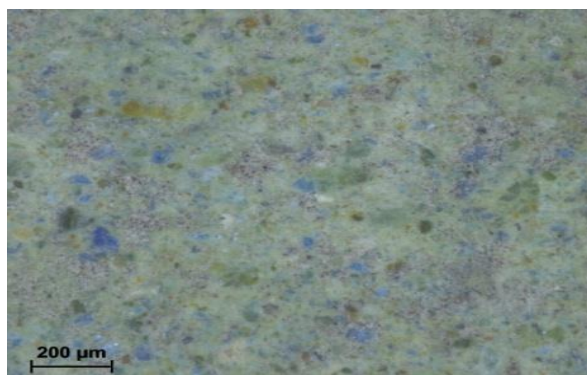


Figure 8: Green frescoed at stereomicroscope, we can observe blue fragments (sample 151).

After a careful analysis with petrographic microscope, the pictorial layers characterized by a reddish color, often tending to yellow or brown, appear respectively composed of oxides and hydroxides (Figure 9), while the layers of bright red color are mainly made up of cinnabar (HgS ; Figure 10). Associated with iron hydroxides, the presence of cinnabar in variable proportions has almost always been found. The yellowish color is instead given by iron hydroxides always immersed in a very fine carbonate matrix. The analysis of the bright red color confirmed the presence of cinnabar crystals always immersed in a carbonate matrix.

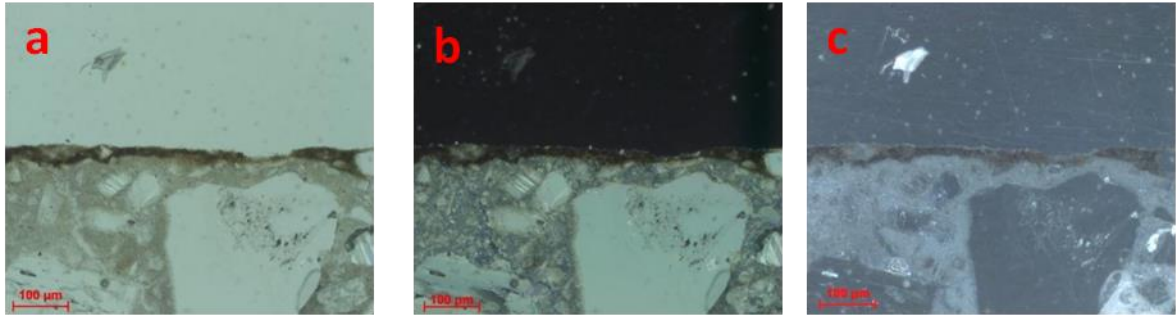


Figure 9: Iron oxides and hydroxides layer in pictorial level of 151 C. Images at: a) polarized transmitted, plane light (//); b) polarized transmitted, crossed light (+); c) dark field in reflected light.

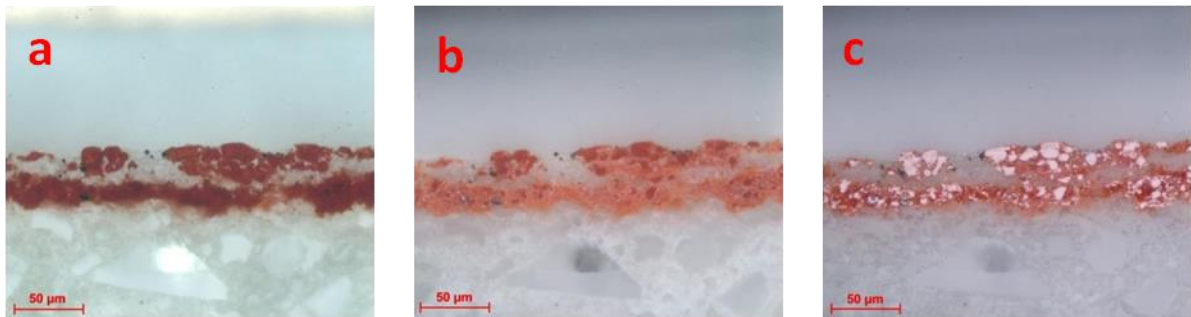


Figure 10: Cinnabar layer in pictorial level of 151 C. Images at: a) polarized transmitted, plane light (//); b) reflected light; c) dark field in reflected light.

For sample 151 C, the analyses have shown some particularities given by the presence of a more external level consisting mainly of very compact but strongly altered calcium phosphate (Figure 11) covering a more internal level consisting of carbonate matrix and iron oxides. This macroscopical layer has a very dark red color.

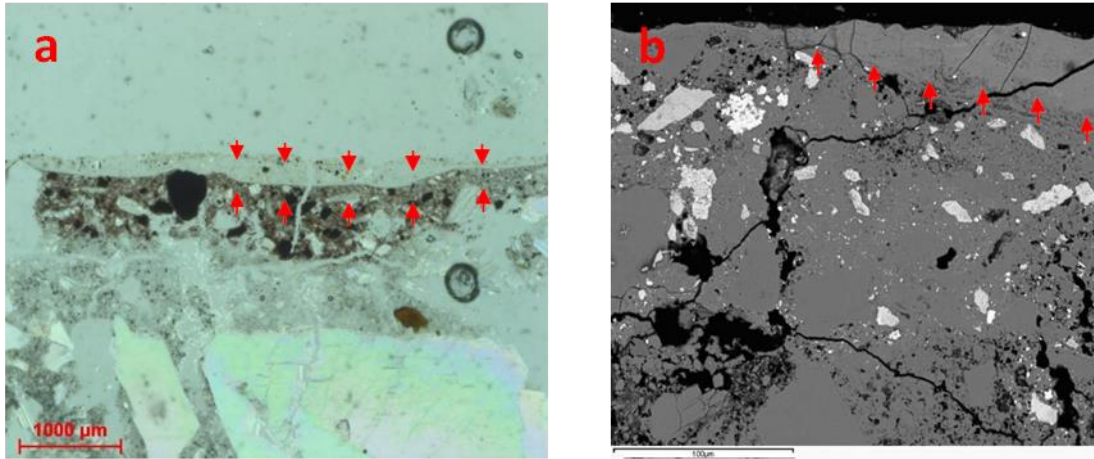


Figure 11: Altered calcium phosphate layer cover a level with carbonate matrix and iron oxides; a) polarized transmitted, plane light (//) image; b) photomicrograph in BSE (151 C sample).

The green pigments are mainly made up of a strongly pleochroic mineral (Figure 11a), with colors varying from light green to a darker green, typical of glauconite $((K,Na)(Mg,Fe^{2+},Fe^{3+})(Fe^{3+},Al)(Si,Al)_4O_{10}(OH)_2)$ and/or chlorite (clinochlore $(Mg_5Al(AlSi_3O_{10})(OH)_8)$ [55,56]. The blue fragments, on the other hand, are weakly pleochroic with turquoise colors of variable intensity. These turquoise crystals also have strong birefringence and high interference colors (Figure 12b and c), all typical characteristics of copper-rich silicate minerals. Quantitative chemical analyses confirmed the presence of glauconite while for turquoise crystals they showed the presence of cuprorivaite.

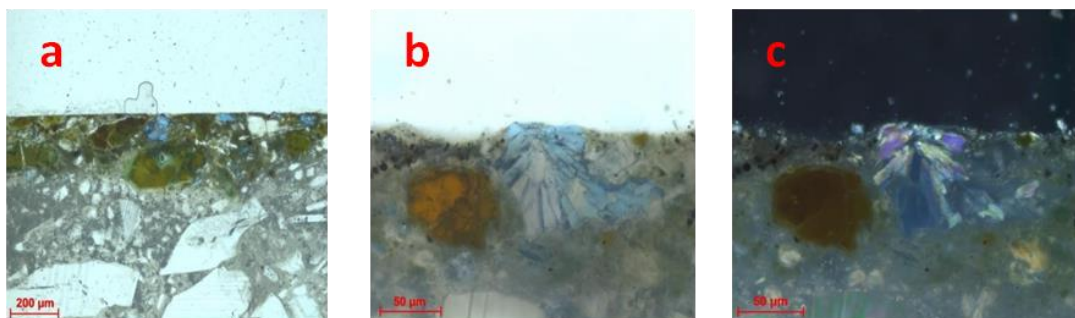


Figure 12: Sample 151. Polarized transmitted light (//) images of green pleochroic mineral (a) and of turquoise crystals (c). Turquoise crystals at polarized transmitted light (+).

The white samples consist mainly of micro crystals of calcium carbonate while the areas with black coloring (Figure 13) are made up of pale-yellow micro crystals (visible even in reflected light) while crossed nicols (+) are always extinct (it could be organic matter). Chemical analyses have confirmed the presence of calcite in a very fine mass for its whitish color.

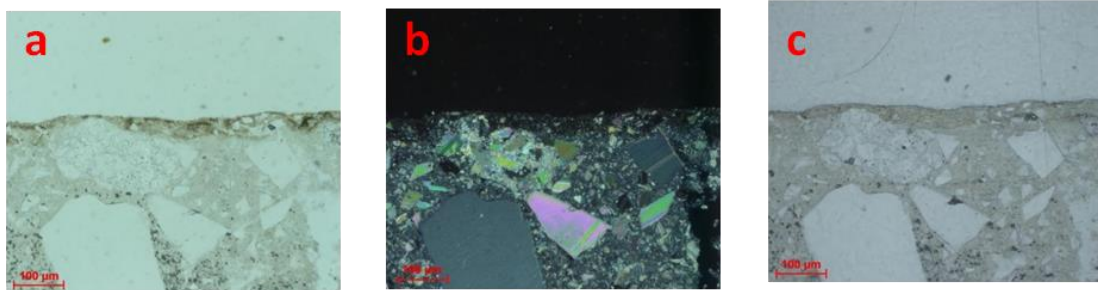


Figure 13: black coloring frescoed of 151 B sample. Images at: a) polarized transmitted light ;(b) polarized transmitted light (+); c) reflected light.

Level 2

The second layer is white-grey in color and its thickness varies greatly from 4 to 8 mm. In a thin section it appears to be made up of a very fine matrix in which fragments of calcite (CaCO_3) crystals of highly variable dimensions, even millimetres, are immersed. Only in the sample 187 the crystals appear to consist also of dolomite ($\text{CaMg}(\text{CO}_3)_2$). In the sample 187, instead, the fragments were mostly dolomite.

Level 3

The third level is generally separated from the second one in a sharp way and this is particularly visible when observing the sample with the petrographic microscope in crossed Nicols mode (Figure 14). The third layer looks like a pyroclastic rock, it seems like tuff, however, after careful observation you can identify it as a *pozzolana* kept together by a matrix. The microscopic observation shows the presence of fragments of pumice more or less zeolithized or feldspatized (Figure 15), minerals of feldspar group crystals, plagioclase, clinopyroxene (Figure 16), mica, strongly altered leucite, apatite minerals group and oxides in smaller quantities. Chemical analyses have been of fundamental importance to confirm the hypothesis of the presence of *pozzolana*, in fact they have highlighted the presence of calcium carbonate for the composition of the matrix of this level, even if it had variable silicon contents up to 8%. They also showed the presence of crystallized glass in K-feldspar, anorthic plagioclase, diopside, phlogopite, fluorapatite and iron oxides.

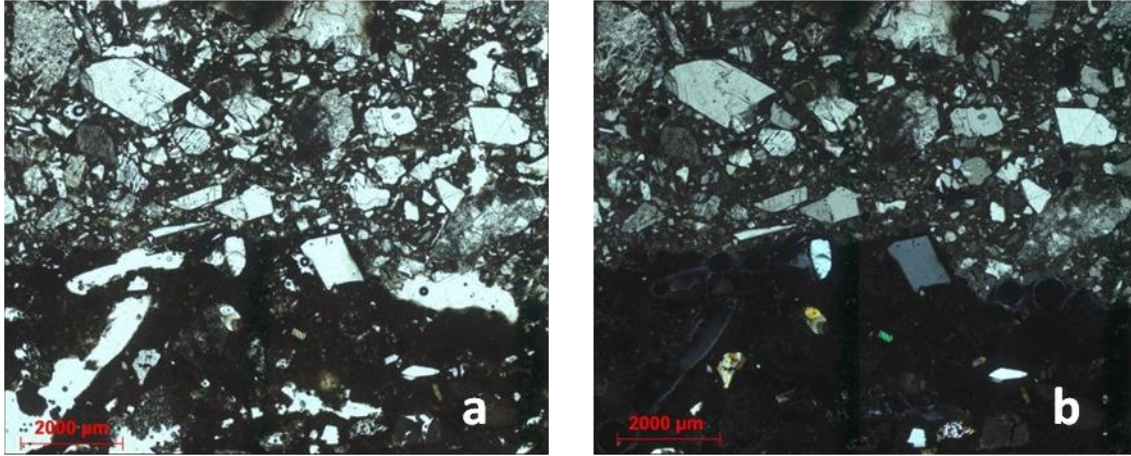


Figure 14: Contact between level 2 (with fragment of calcite crystals) and level 3 (with clinopyroxene, feldspar and pumice). Images at: a) polarized transmitted light (//); b) polarized transmitted light (+). Sample 151 B.

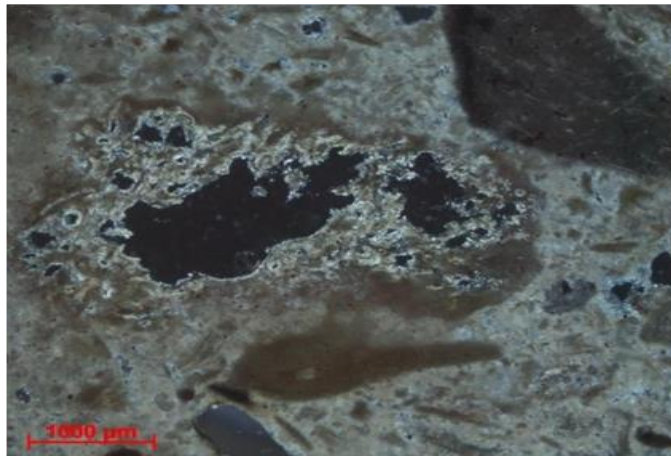


Figure 15: Big fragments of pumice feldspatized in transmitted light (+). Sample 151 C

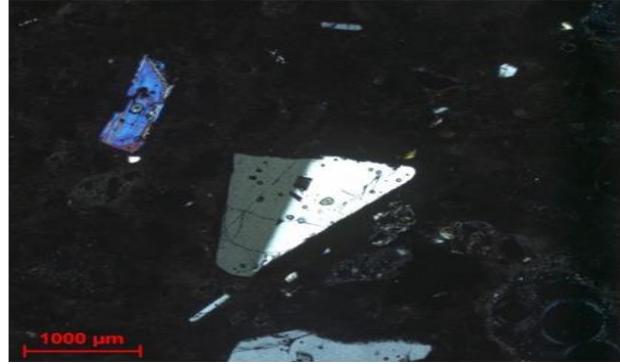


Figure 16 : Fragment of diopside, k-feldspar, apatite crystals and feldspatized pumice. Image at polarized transmitted light (+). Sample 151.

Level 4

The fourth layer was only detected in the 151 C sample. It appears under a polarizing microscope, similar to level 3. On careful observation, however, the fourth layer appears with a higher concentration of pumice and glass, a darker matrix and also the crystalline fragments are in the minority compared to level 3 (Figure 17). Chemical analyses have shown that the matrix has a silicate component of up to 18%, while the rest of the matrix continues to be made up of calcium carbonate. Furtherly, they confirm the presence of higher contents in feldspatized pumice and glass.

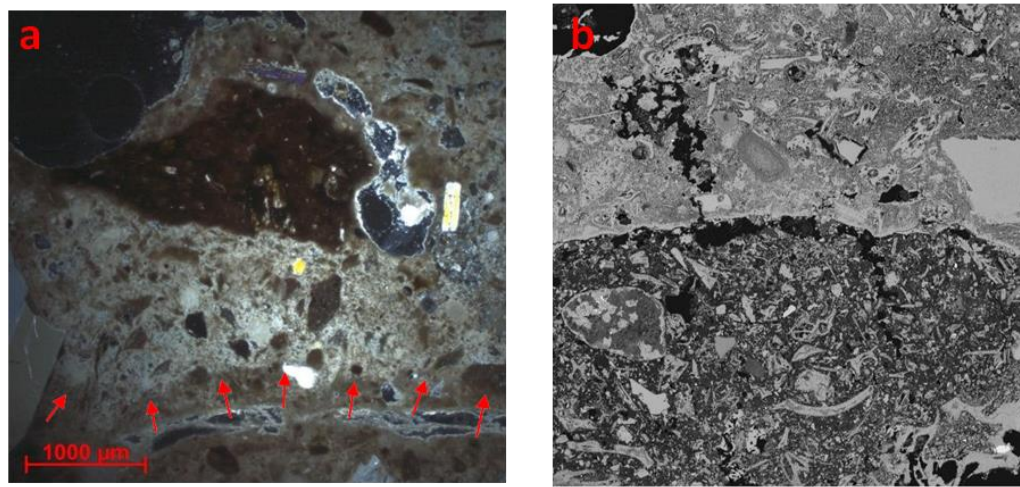


Figure 17: Contact between level 3 and level 4 with major content of pumice and silicon in matrix. Images at: a) polarized transmitted light (//); b) Photomicrograph in BSE. Sample 151 B.

GC-MS Analysis of organic binders

A multistep protocol was applied to the four wall fragments and the polar fractions where sugars and compounds of medium polarity were expected were separated from the unpolar fractions where mostly lipids were expected. The polar fraction was subjected to sugar analysis and chemical derivatization with TMS, whereas the unpolar fraction was subjected to transesterification and GC-MS analysis. Table 2 reports the compounds identified in the polar fraction of each sample. No sugars were detected in any sample but instead, flavonoids, steroids, compounds with abietanic skeleton were identified. In all samples, diglycidyl bisphenol A that is a characteristic organic compound of epoxy resins [57] and 10,18-bisnorabieta-8,11,13-triene that is a degradation product of abietic acid [20,58], were detected. The co-presence of these compounds illustrates the use of epoxy resins as organic binder. More specifically, the samples 151 A and 151 B show a similar chemical profile, with the presence of steroid compounds (Cholest-5-en-19-al,3 β -hydroxy-,cyclic ethylene mercaptal, acetate and hexastrol) that are mostly detected in species of animal origin [59], thus suggesting the presence of an organic binder. On the other hand, stephaboline in the sample 151 C and the semi-long fatty acids that were detected in sample 187 are chemical compounds that are found in plants [60,61]. Also several fatty acids were detected as methyl ester derivatives (FAMES) in the unpolar fractions (table 3). In particular, palmitic and stearic acid, two of the most common molecules that can be found in many matrices both of animal and vegetable origin, were detected in all samples. Their relative ratio is affected by aging. Based on the literature data [3], the ratio C16:0/C18 in the samples 151 A and 151 B is compatible with the presence of animal fat, although in the samples 151 C and 187 is not clearly ascribable to animal or vegetable fat.

Table 2: Table 2: Chemical profile of the polar fraction by GC-MS analysis after sugar analysis and chemical derivatization with TMS. Identification was done by comparison of the retention time and mass spectra to those of standards in the instrument manufacture database NIST MS Search 2.0.

Sample	RT(min)	Name
151 A	6.132	Cholest-5-en-19-al,3 β -hydroxy-,cyclic ethylene mercaptal, acetate
	17.77	10,18-bisnorabieta-8,11,13-triene
	22.12	Diglycidyl bisphenol A
	23.225	Hexestrol, di-TMS
151 B	6.383	Cholest-5-en-19-al,3 β -hydroxy-,cyclic ethylene mercaptal, acetate
	10.434	Dianhydro-2-deoxy- β -d-ribo-hexopyranose
	13.563	2-Methyl-7-hydroxy-8-allyl-isoflavone

	14.097	3-Heptadecenal
	21.473	Oxalic acid, hexadecyl isohexylester
	17.77	10,18-bisnorabieta-8,11,13-triene
	22.12	Diglycidyl bisphenol A
151 C	17.04	Stephaboline
	17.77	10,18-bisnorabieta-8,11,13-triene
	22.12	Diglycidyl bisphenol A
187	17.77	10,18-bisnorabieta-8,11,13-triene
	22.12	Diglycidyl bisphenol A
	21.284	1-Dodecanol, 3,7,11-trimethyl-
	21.356	7-Hexadecenal, (Z)-
	21.452	Octadecane, 1-(ethenyloxy)-
	21.555	1-Dodecanol, 3,7,11-trimethyl-
	21.808	7-Hexadecenal, (Z)-
	22.253	1-Dodecanol, 3,7,11-trimethyl-

Table 3: FAMES profile obtained by GC-MS analysis of the lipidic fractions. Identification was done by comparison of the retention time and mass spectra to those of standards in the instrument manufacture database NIST MS Search 2.0. *(C:N) indicates the number of carbon atoms (C) and double bonds (N) in the fatty acid side chains.

RT (min)	Name	C:N	151 A	151 B	151C	187
12.88	Myristic acid	C14:0	1.60%	1.93%		9.25%
14.71	Pentadecanoic acid	C15:0		3.44%		3.45%
16.25	Palmitic acid	C16:0	35.73%	28.17%	26.25%	45.52%
17.52	Margaric acid	C17:0	4.57%	4.63%		
18.36	16-octadecenoic acid	C18:1	43.30%	49.80%	10.90%	
18.63	Stearic acid	C18:0	14.80%	12.03%	62.85%	41.78%
	Palmitic/Stearic	C16:0/C18:0	2.41	2.34	0.42	1.09

LC-MS/MS Analysis of proteinaceous binders

A classical proteomic approach in heterogeneous phase was carried out to investigate the presence of proteinaceous binders. Proteins were initially, identified searching SwissProt database, with all entries as taxonomy restriction, with MS/MS Ion search Mascot software (Matrix Science). When collagen was detected the sample was reanalyzed against a homemade collagenous database, "COLLE" (60 sequences; 88859 residues). Deamidation of Gln and Asn, oxidation on Met, hydroxylation (K) and hydroxylation (P) were set as variable modifications were no fixed modifications were inserted. In the samples 151 A and 151 B collagen chains of type I of *Bos taurus* were detected, supporting the hypothesis of lipids from an animal origin (Tables 1-2). No significant proteins were detected in the samples 151C and 187.

Proteomic data were also used for a different type of study. According to the literature, there are some bacteria capable of producing pigment with different varieties of colors [62,63]. For example, the bacteria *Flavobacterium sp* produces a yellow-creamy color although the *Chromobacterium violaceum* a red color. We thought that these bacteria could in case of a historical painting be produced using the sugars coming from the resins or the natural gums used for the color protection in the wall surface, thus providing another color source. Therefore, to further validate that the colors observed by spectroscopy are natural colors and not a product of a bacteria production, we constructed a protein database with the protein sequences of bacteria that are reported to produce intense colors (18191sequences; 6393558 residues), that are listed in the table 3.

Table 1 : Identification of collagen in the sample 151 A. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications. Only identification of proteins with at least two peptides with individual ion score above the significance threshold (>20), were considered as significative.

Protein Name (Uniprot Entry)	Protein Sequence Coverage %	m/z	Peptide
Collagen alpha-1(I) chain (P02453)	7	314.677	R.GLPGER.G
		322.673	R.GLPGER.G + Hydroxylation (P) (P)
		392.223	R.GAAGLPGPK.G + Hydroxylation (P) (P)
		426.218	R.GFSGLDGAK.G
		449.756	R.GVVGLPGQR.G + Hydroxylation (P) (P)
		450.251	R.GVVGLPGQR.G + Deamidated (NQ); Hydroxylation (P) (P)
		553.286	R.GVQGPPGPAGPR.G + Hydroxylation (P) (P)

		553.779	R.GVQGPPGPAGPR.G + Deamidated (NQ); Hydroxylation (P) (P)
		589.776	R.GQAGVMGFPGPK.G + Oxidation (M); Deamidated (NQ); Hydroxylation (K) (K)
		781.888	K.DGLNGLPGPIGPPGPR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
		941.451	K.GDTGAKGEPGPAGVQPPGPAGEEGKRGA R.G + Deamidated (NQ); 4 Hydroxylation (P) (P)
Collagen alpha-2(I) chain (P02465)	5%	314.677	R.GLPGER.G
		322.673	R.GLPGER.G + Hydroxylation (P) (P)
		379.696	R.GLPGADGR.A + Hydroxylation (P) (P)
		393.220	R.GATGPAGVR.G
		420.739	R.GVVGPPQGAR.G
		421.234	R.GVVGPPQGAR.G + Deamidated (NQ)
		434.735	R.VGAPGPAGAR.G + Hydroxylation (P) (P)
		459.726	R.AGVMGPAGSR.G + Oxidation (M)
		596.838	R.IGQPGAVGPAGIR.G
		634.339	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)

Table 2: Identification of collagen in the sample 151 B. MS/MS raw data were searched by Mascot MS/MS Ion search software using COLLE database and considering deamidation on Gln and Asn, oxidation on Met, hydroxylation of proline and lysine as variable modifications. Only identification of proteins with at least two peptides with individual ion score above the significance threshold (>20), were considered as significant.

Protein Name (Uniprot Entry)	Protein Sequence Coverage %	<i>m/z</i>	Peptide
Collagen alpha-1(I) chain (P02453)	7%	449.760	R.GVVGLPGQR.G + Hydroxylation (P) (P)
		450.254	R.GVVGLPGQR.G + Deamidated (NQ); Hydroxylation (P) (P)
		544.777	R.GFPGADGVAGPK.G + Hydroxylation (P) (P)
		553.788	R.GVQPPGAGPR.G + Deamidated (NQ); Hydroxylation (P) (P)
		589.783	R.GQAGVMGFPGPK.G + Oxidation (M); Deamidated (NQ); Hydroxylation (K) (K)
		629.800	K.GLTGSPGSPGDGK.T + 2 Hydroxylation (P) (P)
		730.351	R.GSAGPPGATGFPGAAGR.V + 2 Hydroxylation (P) (P)
		781.900	K.DGLNGLPGIPGPPGPR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
		793.884	K.GANGAPGIAGAPGFPGAR.G + Deamidated (NQ); 3 Hydroxylation (P) (P)
Collagen alpha-2(I) chain (P02465)	4%	367.185	K.GPSGDPGKAGEK.G
		596.845	R.IGQPGAVGPAGIR.G
		597.334	R.IGQPGAVGPAGIR.G + Deamidated (NQ)
		601.296	R.GEPGNIGFPGPK.G + Hydroxylation (P) (P); Hydroxylation (K) (K)
		604.844	R.IGQPGAVGPAGIR.G + Hydroxylation (P) (P)
		634.342	R.GIPGPVGAAGATGAR.G + Hydroxylation (P) (P)
		714.369	R.GIPGEFGLPGPAGAR.G + 2 Hydroxylation (P) (P)

Table 3: The type of bacteria that were tested against pigment production in the wall surface [62].

Microorganisms (Bacteria)	Pigments/Molecule	Colour
Agrobacterium aurantiacum	Astaxanthin	Pink-red
Paracoccus carotinifaciens	Astaxanthin	Pink-red
Bradyrhizobium specie	Canthaxanthin	Dark- red
Flavobacterium specie, Paracoccus zeaxanthinifaciens	Zeaxanthin	Yellow
Achromobacter	Zeaxanthin	Creamy
Bacillus	Zeaxanthin	Brown
Brevibacterium specie	Zeaxanthin	Orange Yellow
Corynebacterium insidiosum	Indigoidine	Blue
Rugamonas rubra, Streptovorticillium rubrireticuli, Vibrio gaogenes, Alteromonas rubra	Prodigiosin	Red
Rhodococcus maris		Bluish-Red
Xanthophyllomyces dendrorhous	Astaxanthin	Pink-Red
Haloferax alexandrinus	Canthaxanthin	Dark Red
Staphylococcus aureus	Staphyloxanthin, Zeaxanthin	Golden Yellow
Chromobacterium violaceum	Violacein	Purple
Serratia marcescens, Serratia rubidaea,	Prodigiosin	Red
Pseudomonas aeruginosa	Pyocyanin	Blue-Green
Xanthomonas oryzae	Xanthomonadin	Yellow
Janthinobacterium lividum	Violacein	Purple

Conclusions

Raman and SEM spectroscopy analyses allowed to characterize the inorganic components of the different layers of the wall painting of the Domus in S.M. Capua Vetere.

The blue shade identified in the samples as “Egyptian *Blue*” [26, 27], is a concise pigment based on Cuprorivaite. The green color is Glauconite [36] known as “Green Earth”. The elemental analysis of the *yellow* pigments indicates that they are highly heterogeneous, most likely corresponding to the mineral limonite (mixture of hydroxide/oxides of iron, zinc and titan), and Raman analysis revealed one minor component, assigned as Fe(OH)₃. The Raman results for the *red* shade confirms the presence of HgS in the “151 A” sample, while the “151 B” and the “151 C” shows the co-presence of Goethite and Haematite. The Hematite is visible only in the Raman analysis, and it is probably caused by a partial “Goethite-Hematite” conversion, possibly due to the laser exposure [31]. It is worth noting that most likely goethite itself is a product of oxidation of an initial hematite, so accidentally the laser-induced Raman assignment can correspond to the actual starting pigment. The complex of mortars preparation (that goes from three to four

layers), and the variety of used shades, result of raw material blends, as shown from the optical microscopy results and by the binder abundance suggests that the villa is an aristocratic one.

These observations were confirmed by the analysis of the organic component. In the preparation of the wall paintings, resins and animal glue have been extensively used, so they can be classified as fresco-secco paintings. In the Roman period this technique was less common in comparison to fresco where the painting preparation doesn't include the use of a proteinaceous binder but instead the use of emulsified beeswax. Interestingly, a screening of proteomic data against a database of bacteria that are able to produce color we confirmed that the observed colors by spectroscopy are not related to microbial attack or to degradation processes related to bacteria.

References

- [1] G. Giorgi, Overview of Mass Spectrometric Based Techniques Applied in the Cultural Heritage Field, in: *Org. Mass Spectrom. Art Archaeol.*, John Wiley & Sons, Ltd, 2009: pp. 37–74. <https://doi.org/10.1002/9780470741917.ch2>.
- [2] I. Degano, F. Modugno, I. Bonaduce, E. Ribechini, M.P. Colombini, Recent Advances in Analytical Pyrolysis to Investigate Organic Materials in Heritage Science, *Angew. Chemie Int. Ed.* 57 (2018) 7313–7323. <https://doi.org/10.1002/anie.201713404>.
- [3] J. Blaško, R. Kubinec, B. Husová, P. Přikryl, V. Pacáková, K. Štulík, J. Hradilová, Gas chromatography/mass spectrometry of oils and oil binders in paintings, *J. Sep. Sci.* 31 (2008) 1067–1073. <https://doi.org/10.1002/jssc.200700449>.
- [4] S. Vahur, A. Teearu, P. Peets, L. Joosu, I. Leito, ATR-FT-IR spectral collection of conservation materials in the extended region of 4000–80 cm^{-1} , *Anal. Bioanal. Chem.* 408 (2016) 3373–9. <https://doi.org/10.1007/s00216-016-9411-5>.
- [5] N. Salvadó, S. Butí, M.J. Tobin, E. Pantos, A.J.N.W. Prag, T. Pradell, Advantages of the use of SR-FT-IR microspectroscopy: applications to cultural heritage., *Anal. Chem.* 77 (2005) 3444–51. <https://doi.org/10.1021/ac050126k>.
- [6] D. Lazidou, D. Lampakis, I. Karapanagiotis, C. Panayiotou, Investigation of the Cross-Section Stratifications of Icons Using Micro-Raman and Micro-Fourier Transform Infrared (FT-IR) Spectroscopy, *Appl. Spectrosc.* 72 (2018) 1258–1271. <https://doi.org/10.1177/0003702818777772>.
- [7] C. Daher, L. Bellot-Gurlet, A.-S. Le Hô, C. Paris, M. Regert, Advanced discriminating criteria for natural organic substances of cultural heritage interest: spectral decomposition and multivariate analyses of FT-Raman and FT-IR signatures., *Talanta*. 115 (2013) 540–7. <https://doi.org/10.1016/j.talanta.2013.06.014>.
- [8] M. I. M., *Atti e Memorie della Società Magna Grecia, 1926–30 (Vols. I–IV)*. Rome: Palazzo Taverna. Annual Subscription, 100 lire. - *L'opera della Società Magna Grecia nei primi 10 anni*. By U. Zanotti-Bianco. From *Annales Institutorum*, 1931. Rome: Palazzo Ricci, 19, *J. Hell. Stud.* (1932). <https://doi.org/10.2307/627206>.
- [9] P. Zancani Montuoro, *Heraion alla force del Sele: I: Altre metope del primo thesauros*, *Atti e Mem. Della Soc. Magna Grecia*. (1964).
- [10] P. Cennamo, M. Rosaria, B. Lumaga, G. Fatigati, A. Amoresano, A. Carpentieri, CONSERVATION SCIENCE A MULTIDISCIPLINARY ASSESSMENT TO INVESTIGATE A XXII, 11 (2020) 25–38.
- [11] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage: From bones to paintings ☆, *Microchem. J.* 126 (2016) 341–348.

<https://doi.org/10.1016/j.microc.2015.12.024>.

[12] S. Rossi, M. Petrelli, D. Morgavi, F.P. Vetere, R.R. Almeev, R.L. Astbury, D. Perugini, Role of magma mixing in the pre-eruptive dynamics of the Aeolian Islands volcanoes (Southern Tyrrhenian Sea, Italy), *Lithos*. 324–325 (2019) 165–179. <https://doi.org/10.1016/j.lithos.2018.11.004>.

[13] B. Jananee, V. Thangam, A. Rajalakshmi, Investigation of soils by thermal and spectroscopic analysis, *Chem. Eng. Commun.* (2019). <https://doi.org/10.1080/00986445.2019.1680370>.

[14] R.J. Cox, H.L. Peterson, J. Young, C. Cusik, E.O. Espinoza, The forensic analysis of soil organic by FTIR, *Forensic Sci. Int.* 108 (2000) 107–116. [https://doi.org/10.1016/S0379-0738\(99\)00203-0](https://doi.org/10.1016/S0379-0738(99)00203-0).

[15] M. Cavagna, R. Dell'Anna, F. Monti, F. Rossi, S. Torriani, Use of ATR-FTIR microspectroscopy to monitor autolysis of *Saccharomyces cerevisiae* cells in a base wine, *J. Agric. Food Chem.* 58 (2010) 39–45. <https://doi.org/10.1021/jf902369s>.

[16] S.H. Gordon, S.C. Whatel, B.C. Wheeler, C. James, Multivariate FTIR analysis of substrates for protein, polysaccharide, lipid and microbe content: Potential for solid-state fermentations, *Biotechnol. Adv.* 11 (1993) 665–675. [https://doi.org/10.1016/0734-9750\(93\)90035-L](https://doi.org/10.1016/0734-9750(93)90035-L).

[17] B. Wiyono, S. Tachibana, D. Tinambunan, Chemical Compositions of Pine Resin, Rosin and Turpentine Oil from West Java, Indones. *J. For. Res.* (2006). <https://doi.org/10.20886/ijfr.2006.3.1.7-17>.

[18] R.P. Evershed, C. Heron, L. John Goad, Analysis of organic residues of archaeological origin by high-temperature gas chromatography and gas chromatography - Mass spectrometry, *Analyst.* (1990). <https://doi.org/10.1039/an9901501339>.

[19] M.D. Petit-Dominguez, J. Martinez-Maganto, MCF fast derivatization procedure for the identification of resinous deposit components from the inner walls of roman age amphorae by GC-MS, *Talanta.* (2000). [https://doi.org/10.1016/S0039-9140\(99\)00330-6](https://doi.org/10.1016/S0039-9140(99)00330-6).

[20] C. Azemard, M. Menager, C. Vieillescazes, Analysis of diterpenic compounds by GC-MS/MS: contribution to the identification of main conifer resins, *Anal. Bioanal. Chem.* 408 (2016) 6599–6612. <https://doi.org/10.1007/s00216-016-9772-9>.

[21] J. Font, N. Salvadó, S. Butí, J. Enrich, Fourier transform infrared spectroscopy as a suitable technique in the study of the materials used in waterproofing of archaeological amphorae, *Anal. Chim. Acta.* 598 (2007) 119–127. <https://doi.org/10.1016/j.aca.2007.07.021>.

[22] I. Jerković, Z. Marijanović, M. Gugić, M. Roje, Chemical profile of the organic residue from ancient amphora found in the adriatic sea determined by direct GC and GC-MS analysis, *Molecules.* 16 (2011) 7936–7948. <https://doi.org/10.3390/molecules16097936>.

[23] F.J. Moreno, A. Montilla, M. Villamiel, N. Corzo, A. Olano, Analysis, structural characterization, and bioactivity of oligosaccharides derived from lactose, *Electrophoresis.* 35 (2014) 1519–1534. <https://doi.org/10.1002/elps.201300567>.

[24] G.J. Dutton, Glucuronide-Forming Enzymes, in: *Concepts Biochem. Pharmacol.*, Springer Berlin Heidelberg, 1971: pp. 378–400. https://doi.org/10.1007/978-3-642-65177-9_24.

[25] A. Goyon, J.Z. Cai, K. Kraehenbuehl, C. Hartmann, B. Shao, P. Mottier, Determination of steroid hormones in bovine milk by LC-MS/MS and their levels in Swiss Holstein cow milk, *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* 33 (2016) 804–816. <https://doi.org/10.1080/19440049.2016.1175186>.

[26] J. Orsavova, L. Misurcova, J. Vavra Ambrozova, R. Vicha, J. Mlcek, Fatty acids composition of vegetable oils and its contribution to dietary energy intake and dependence of cardiovascular mortality on dietary intake of fatty acids, *Int. J. Mol. Sci.* 16 (2015) 12871–12890. <https://doi.org/10.3390/ijms160612871>.

[27] J. Pollier, A. Goossens, Oleanolic acid, *Phytochemistry.* 77 (2012) 10–15. <https://doi.org/10.1016/j.phytochem.2011.12.022>.

[28] M.A. Lozano-Grande, S. Gorinstein, E. Espitia-Rangel, G. Dávila-Ortiz, A.L. Martínez-Ayala, Plant Sources, Extraction Methods, and Uses of Squalene, *Int. J. Agron.* 2018 (2018). <https://doi.org/10.1155/2018/1829160>.

- [29] S.K. Kim, F. Karadeniz, Biological Importance and Applications of Squalene and Squalane, in: *Adv. Food Nutr. Res.*, Academic Press Inc., 2012: pp. 223–233. <https://doi.org/10.1016/B978-0-12-416003-3.00014-7>.
- [30] X. Zhang, S. Zhang, Y. Yang, D. Wang, H. Gao, Natural barrigenol-like triterpenoids: A comprehensive review of their contributions to medicinal chemistry, *Phytochemistry*. 161 (2019) 41–74. <https://doi.org/10.1016/j.phytochem.2019.01.017>.
- [31] P.D. Uchil, A. Nagarajan, P. Kumar, β -galactosidase, *Cold Spring Harb. Protoc.* 2017 (2017) 774–779. <https://doi.org/10.1101/pdb.top096198>.
- [32] M.P. Colombini, G. Giachi, F. Modugno, E. Ribechini, Characterisation of organic residues in pottery vessels of the Roman age from Antinoe (Egypt), in: *Microchem. J.*, 2005: pp. 83–90. <https://doi.org/10.1016/j.microc.2004.05.004>.
- [33] J. Cuní, P. Cuní, B. Eisen, R. Savizky, J. Bové, Characterization of the binding medium used in Roman encaustic paintings on wall and wood, *Anal. Methods*. 4 (2012) 659–669. <https://doi.org/10.1039/c2ay05635f>.
- [34] A. Pichot, Études de lettres Theatra et spectacula Théâtres et amphithéâtres : outils de romanisation en Maurétanie ?, in: [Http://Journals.Openedition.Org/Edl](http://Journals.Openedition.Org/Edl), Faculté des lettres de l'Université de Lausanne, 2011: pp. 171–192. <https://doi.org/10.4000/edl.113>.
- [35] G. Renda, S. Gigli, A. Amato, S. Venticinque, B. Di Martino, F.R. Cappa, Mobile devices for the visit of anfiteatro campano in Santa Maria Capua Vetere, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Springer, Berlin, Heidelberg, 2012: pp. 281–290. https://doi.org/10.1007/978-3-642-34234-9_28.
- [36] M. Gelzo, M. Grimaldi, A. Vergara, V. Severino, A. Chambery, A. Dello Russo, C. Piccioli, G. Corso, P. Arcari, Comparison of binder compositions in Pompeian wall painting styles from Insula Occidentalis, *Coke Chem.* 8 (2014) 65. <https://doi.org/10.1186/s13065-014-0065-0>.
- [37] M. Gelzo, M. Grimaldi, A. Vergara, V. Severino, A. Chambery, A. Dello Russo, C. Piccioli, G. Corso, P. Arcari, Comparison of binder compositions in Pompeian wall painting styles from Insula Occidentalis, *Chem. Cent. J.* 8 (2014) 65. <https://doi.org/10.1186/s13065-014-0065-0>.
- [38] G. Corso, M. Gelzo, A. Chambery, V. Severino, A. Di Maro, F.S. Lomoriello, O. D'Apolito, A. Dello Russo, P. Gargiulo, C. Piccioli, P. Arcari, Characterization of pigments and ligands in a wall painting fragment from Litternum archaeological park (Italy), *J. Sep. Sci.* 35 (2012) 2986–2993. <https://doi.org/10.1002/jssc.201200490>.
- [39] A. Casoli, S. Santoro, Organic materials in the wall paintings in Pompei: A case study of Insula del Centenario, *Chem. Cent. J.* 6 (2012) 107. <https://doi.org/10.1186/1752-153X-6-107>.
- [40] J. Cuní, What do we know of Roman wall painting technique? Potential confounding factors in ancient paint media analysis, *Herit. Sci.* 4 (2016) 44. <https://doi.org/10.1186/s40494-016-0111-4>.
- [41] L. Birolo, A. Tomeo, M. Trifuoggi, F. Auriemma, L. Paduano, A. Amoresano, R. Vinciguerra, C. De Rosa, L. Ferrara, A. Giarra, A. Luchini, C. De Maio, G. Greco, A. Vergara, A hypothesis on different technological solutions for outdoor and indoor Roman wall paintings, *Archaeol. Anthropol. Sci.* 9 (2017) 591–602. <https://doi.org/10.1007/s12520-016-0408-y>.
- [42] S. Gunasekaran, G. Anbalagan, S. Pandi, Raman and infrared spectra of carbonates of calcite structure, *J. Raman Spectrosc.* 37 (2006) 892–899. <https://doi.org/10.1002/jrs.1518>.
- [43] H.M. Mahmoud, A MULTI-ANALYTICAL APPROACH FOR CHARACTERIZING PIGMENTS FROM THE TOMB OF DJEHUTYEMHAB (TT 194), (2013).
- [44] H.H. Marey Mahmoud, M. Mahmoud, H. H., A Preliminary Investigation Of Ancient Pigments From The Mortuary Temple Of Seti I, El-Qurna (Luxor, Egypt), *MAA.* 11 (2011) 99–106. <https://ui.adsabs.harvard.edu/abs/2011MAA....11...99M/abstract> (accessed December 5, 2020).
- [45] F. Ospitali, D. Bersani, G. Di Lonardo, P.P. Lottici, “Green-earths”: vibrational and elemental characterisation of glauconites, celadonites and historical pigments, (2007) 40–41.

- [46] R.J.H. Clark, P.J. Gibbs, K.R. Seddon, N.M. Brovenko, Y.A. Petrosyan, Non-destructive in situ identification of cinnabar on ancient Chinese manuscripts, *J. Raman Spectrosc.* 28 (1997) 91–94. [https://doi.org/10.1002/\(sici\)1097-4555\(199702\)28:2/3<91::aid-jrs67>3.0.co;2-x](https://doi.org/10.1002/(sici)1097-4555(199702)28:2/3<91::aid-jrs67>3.0.co;2-x).
- [47] P. Baraldi, C. Baraldi, R. Curina, L. Tassi, P. Zannini, A micro-Raman archaeometric approach to Roman wall paintings, *Vib. Spectrosc.* 43 (2007) 420–426. <https://doi.org/10.1016/j.vibspec.2006.04.029>.
- [48] E.A. Lalla, G. Lopez-Reyes, A. Sansano, A. Sanz-Arranz, J. Martínez-Frías, J. Medina, F. Rull-Pérez, Raman-IR vibrational and XRD characterization of ancient and modern mineralogy from volcanic eruption in Tenerife Island: Implication for Mars, *Geosci. Front.* 7 (2016) 673–681. <https://doi.org/10.1016/j.gsf.2015.07.009>.
- [49] M. Alfè, V. Gargiulo, R. Di Capua, F. Chiarella, J.N. Rouzaud, A. Vergara, A. Ciajolo, Wet chemical method for making graphene-like films from carbon black, *ACS Appl. Mater. Interfaces.* 4 (2012) 4491–4498. <https://doi.org/10.1021/am301197q>.
- [50] D.C. Smith, M. Bouchard, M. Lorblanchet, An initial Raman microscopic investigation of prehistoric rock art in caves of the Quercy District, S. W. France, *J. Raman Spectrosc.* 30 (1999) 347–354. [https://doi.org/10.1002/\(SICI\)1097-4555\(199904\)30:4<347::AID-JRS379>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-4555(199904)30:4<347::AID-JRS379>3.0.CO;2-A).
- [51] D.L.A. De Faria, S. Venâncio Silva, M.T. De Oliveira, Raman microspectroscopy of some iron oxides and oxyhydroxides, *J. Raman Spectrosc.* 28 (1997) 873–878. [https://doi.org/10.1002/\(sici\)1097-4555\(199711\)28:11<873::aid-jrs177>3.0.co;2-b](https://doi.org/10.1002/(sici)1097-4555(199711)28:11<873::aid-jrs177>3.0.co;2-b).
- [52] S. Das, M.J. Hendry, Application of Raman spectroscopy to identify iron minerals commonly found in mine wastes, *Chem. Geol.* 290 (2011) 101–108. <https://doi.org/10.1016/j.chemgeo.2011.09.001>.
- [53] I.M. Bell, R.J.H. Clark, P.J. Gibbs, Raman spectroscopic library of natural and synthetic pigments (pre- \approx 1850 AD), *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 53 (1997) 2159–2179. [https://doi.org/10.1016/S1386-1425\(97\)00140-6](https://doi.org/10.1016/S1386-1425(97)00140-6).
- [54] J. Jehlička, H.G.M. Edwards, K. Osterrothová, J. Novotná, L. Nedbalová, J. Kopecký, I. Němec, A. Oren, Potential and limits of Raman spectroscopy for carotenoid detection in microorganisms: Implications for astrobiology, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 372 (2014). <https://doi.org/10.1098/rsta.2014.0199>.
- [55] Ugo Zezza, *Petrografia microscopica*, La Goliardica Pavese, 1976.
- [56] A.P.D. Peccerillo, *Introduzione alla petrografia ottica*, Morlacchi Editore, 2003. <https://www.amazon.it/Introduzione-alla-petrografia-ottica-CD-ROM/dp/8888778276> (accessed December 5, 2020).
- [57] N. Szczepańska, P. Kubica, B. Kudłak, J. Namieśnik, A. Wasik, Stabilities of bisphenol A diglycidyl ether, bisphenol F diglycidyl ether, and their derivatives under controlled conditions analyzed using liquid chromatography coupled with tandem mass spectrometry, *Anal. Bioanal. Chem.* 411 (2019) 6387–6398. <https://doi.org/10.1007/s00216-019-02016-5>.
- [58] V. Beltran, N. Salvadó, S. Butí, T. Pradell, Ageing of resin from Pinus species assessed by infrared spectroscopy, *Anal. Bioanal. Chem.* 408 (2016) 4073–4082. <https://doi.org/10.1007/s00216-016-9496-x>.
- [59] G. Kaklamanos, G. Theodoridis, T. Dabalís, Determination of anabolic steroids in bovine urine by liquid chromatography-tandem mass spectrometry, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 877 (2009) 2330–2336. <https://doi.org/10.1016/j.jchromb.2009.03.033>.
- [60] H. Zhang, J.M. Yue, Hasubanan type alkaloids from *Stephania longa*, *J. Nat. Prod.* 68 (2005) 1201–1207. <https://doi.org/10.1021/np0500833>.
- [61] P. Ranchana, M. Ganga, M. Jawaharlal, M. Kannan, Investigation of Volatile Compounds from the Concrete of *Jasminum auriculatum* Flowers, *Int.J.Curr.Microbiol.App.Sci.* 6 (2017) 1525–1531. <https://doi.org/10.20546/ijcmas.2017.611.180>.
- [62] H.M. Usman, N. Abdulkadir, M. Gani, H.M. Maiturare, Bacterial pigments and its significance, (2017). <https://doi.org/10.15406/mojbb.2017.04.00073>.
- [63] P. Sabbagh, A. Ebrahimzadeh Namvar, The eminence status of bacterial pigments under different

- aspects, *Microbiol. Medica*. 32 (2017). <https://doi.org/10.4081/mm.2017.7247>.
- [64] Il Palazzo - Sito web non ufficiale della Reggia di Caserta, (n.d.). <https://www.reggiadicasertaunofficial.it/it/reggia/il-palazzo/> (accessed November 24, 2020).
- [65] 18th-Century Royal Palace at Caserta with the Park, the Aqueduct of Vanvitelli, and the San Leucio Complex - Documents - UNESCO World Heritage Centre, (n.d.). <https://whc.unesco.org/en/list/549/documents/%23ABevaluation> (accessed November 24, 2020).
- [66] Dictionnaire amoureux de Versailles - Franck FERRAND - Google Libri, (n.d.). https://books.google.it/books?id=qTdIAQAAQBAJ&redir_esc=y (accessed November 24, 2020).
- [67] J. Mazurek, M. Svoboda, M. Schilling, GC/MS Characterization of Beeswax, Protein, Gum, Resin, and Oil in Romano-Egyptian Paintings, *Heritage*. 2 (2019) 1960–1985. <https://doi.org/10.3390/heritage2030119>.
- [68] Rapid differentiation of beeswaxes using EGA-MS and Py-GC / MS, (n.d.) 8862.
- [69] M.P. Colombini, G. Giachi, F. Modugno, E. Ribechini, Characterisation of organic residues in pottery vessels of the Roman age from Antinoe (Egypt), *Microchem. J.* 79 (2005) 83–90. <https://doi.org/10.1016/j.microc.2004.05.004>.
- [70] I. Jerković, Z. Marijanović, M. Gugić, M. Roje, Chemical profile of the organic residue from ancient amphora found in the adriatic sea determined by direct GC and GC-MS analysis, *Molecules*. 16 (2011) 7936–7948. <https://doi.org/10.3390/molecules16097936>.
- [71] Y. Zhang, G. Wu, C. Chang, Y. Lv, W. Lai, H. Zhang, X. Wang, Q. Jin, Determination of Origin of Commercial Flavored Rapeseed Oil by the Pattern of Volatile Compounds Obtained via GC–MS and Flash GC Electronic Nose, *Eur. J. Lipid Sci. Technol.* 122 (2020) 1–7. <https://doi.org/10.1002/ejlt.201900332>.
- [72] A. Lluveras-Tenorio, J. Mazurek, A. Restivo, M.P. Colombini, I. Bonaduce, The Development of a New Analytical Model for the Identification of Saccharide Binders in Paint Samples, *PLoS One*. 7 (2012) e49383. <https://doi.org/10.1371/journal.pone.0049383>.
- [73] S. Bogdanov, B.P. Science, Beeswax : Uses and Trade, *The Beeswax*. (2009) 1–16.
- [74] C. Brøns, K.L. Rasmussen, M.M. Di Crescenzo, R. Stacey, A. Lluveras-Tenorio, Painting the Palace of Apries I: Ancient binding media and coatings of the reliefs from the Palace of Apries, Lower Egypt, *Herit. Sci.* 6 (2018) 1–20. <https://doi.org/10.1186/s40494-018-0170-9>.

5.5 Manuscript 5: Molecular characterization of protective layers in statues from the monumental staircase in the Royal Palace of Caserta

The analyses herein reported are part of a study in progress with the Laboratorio di Restauro of the Royal Palace of Caserta intended to establish an active collaboration between Academia and Museums in the perspective of a full integration of research in the field of cultural heritage and the actual conservation and restoration activities. The data will be integrated with other analyses in a common manuscript focused on the diagnosis of different elements of the decoration of the immediate external area of the Royal Palace in Caserta.

Introduction

The Royal Palace of Caserta was built as the new kingdom of the King Carlo di Borbone. The design of this impressive building with more than 1000 rooms, including a chapel library and a theater was started in 1752 by the architect Luigi Vanvitelli and was finished in 1845 by his son Carlo Vanvitelli. The project included not only the construction of the building, but also of pools, botanical gardens, and cascades that are aligned in a telescopic view and had as purpose to relax the royals but also to ameliorate the botanical research. The fountains stretch for over 3km from the Palace; among them, the Diana and Atteone, the Delfini fountain and the Eolo fountain that were inspired from the latin poem “Eneide” written by Virgil between 29 and 19 BC. Inside the royal palace a monumental marble staircase, guarded by two carved polished lions, leads up to the Palatine chapel [1-2]. The lions, as can be deduced from the declaration of Vanvitelli's drawings, symbolize the strength of reason and of the weapons that ensure the king the possession of his kingdoms. The architects are Tommaso Solari and Paolo Persico [64–66].



Figure 1 : Lion's statue in the royal palace of Caserta © thephraser.com

Aim of the study

Aim of this study was the molecular characterization of lion's outer part by mass spectrometry (GC-MS) in order to investigate the hedging techniques of that artistic period as also to provide clues on the mechanisms of biological or chemical-physical degradation, as preliminary analysis requested for the monument restoration expected in 2020-2021.

Materials and methods

Samples

Three samples (A,B,C) were collected from the left lion in the entrance of the royal palace in Caserta. Sample A was from lion's foot, sample B from the transparent pouring and sample C that demonstrates a deposit accumulated material near the lion's foot..



Figure 2: The sampling that was performed in the royal palace of Caserta

GC-MS analysis

Samples were treated and analyzed as reported in [10].

Results

A multistep protocol was applied to the samples for the fractionation of organic molecules based on polarity. Polar fractions, where sugars were expected, was subjected to sugar analysis and GC-MS analysis after chemical derivatization with TMS, whereas the unpolar fraction, where lipids are expected, to transesterification and then to GC-MS analysis. Figure 1 and table 1 demonstrate the chemical compounds of the unpolar fractions that are divided into monocarboxylic and dicarboxylic acids, compounds with abietanic skeletons as also long chain fatty acids. In particular, the sample from the lion's foot contains fatty acids and triglycerides that can be attributed to plant wax [67,68]. However, the absence of very long fatty esters, typical of wax (C_{40} - C_{48}), suggests that the wax has been mixed with other materials, such as resins and plant extracts [67]. Moreover, several tricyclic diterpenoid acids are identified. Among them, methyl abietate (MA) methyldehydroabietic acid (MDA) as also 7-oxo-dehydroabietic acid (7ODA), are characteristic compounds, found in resins from the *Pinaceae* family [69]. More specifically, 7-oxo-dehydroabietic acid is an oxidation product [70]. The mix of the bees wax with natural resins is confirmed from the chemical profiling of the transparent pouring that demonstrated the same kind of compounds as before. A completely different chemical profile is presented in the case of the deposit-accumulated material. Extremely interesting was the identification of palmitic, vaccenic, stearic, arachidic, behenic, lignoceric, erucid, cerotic and montanic acid that unambiguously demonstrates the use of rapeseed oil [71]. The determination and specification of rapeseed oil was mainly based on the presence of erucid acid, quite characteristic of the *Brassicaceae* family. A semiquantitative analysis of the ten major fatty acids was carried out, on the basis of relative area percentages. As shown in table 4, erucid acid in the C sample is 4 times higher than palmitic and stearic acid, defining the presence of rapeseed oil, according to literature data [71]. Furthermore, in the polar fraction several sugars were detected (figure 2), the relative abundance of which is presented in table 2. The presence of arabinose and galactose, that are the main constituents of Arabic gum, suggests the use of natural gum [8]. However, the abundance of glycerol that is a plasticizer agent and whose detection is not compatible with ancient decoration techniques, suggests that the gum might have been added at a later stage during conservation intervention. Moreover, glucose and xylose were identified, which may originate from degraded monosaccharides migrating from the varnish layer. This could be also confirmed by the different percentages of xylose and glucose in the three samples. As it has previously reported, ratios of xylose/arabinose >1 suggests that external contamination must be considered as a source of saccharide material. [72], and in the transparent and the deposit material this ratio is two and four, respectively.

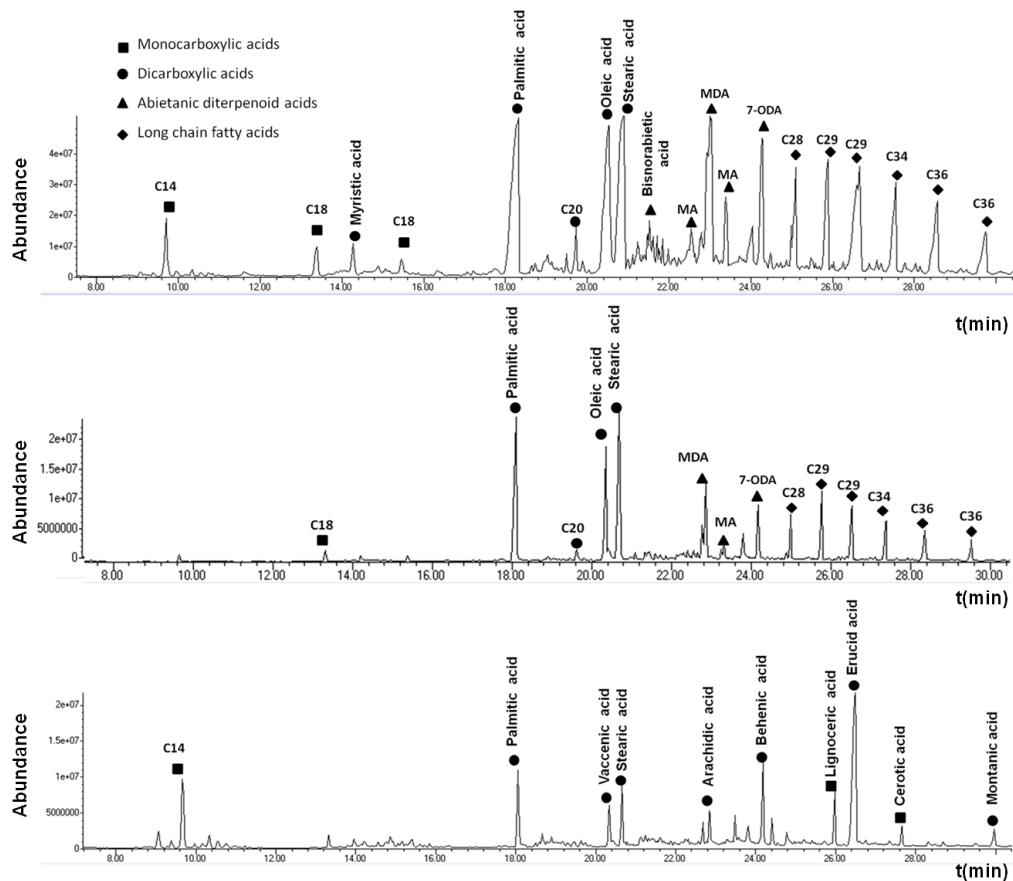


Fig. 1. Total Ion Current (TIC) showing the profile of the methyl ester: fatty acids, mono and dicarboxylic acids, and abietanic acids identified in the three samples. **A:** Lion's foot; **B:** transparent pouring; **C:** Deposit accumulated material. Identification was based on the direct match with the spectra of NIST library. A correlation match index higher than 95% was considered for a confident identification.

Table 1: Methyl ester fatty acids, mono and dicarboxylic acids, and abietanic acids identified by GC-MS analysis. **A:** Lion's foot; **B:** transparent pouring; **C:** Deposit accumulated material.

Sample	Retention time (min)	Chemical name
A	9.71	Phenol, 2,4-bis(1,1-dimethylethyl)
	13.39	Benzene, 1,2,4,5-tetrakis(1-methylethyl)
	15.47	Phenol, 2-(1,1-dimethylethyl)-4-(1,1,3,3-tetramethylbutyl)
	21.50	10,18 Bisnorabieta, 1,11-13 triene
	21.80	Methyl abietate
	23.02	Methyldehydroabietic acid
	23.37	Methyl abietate
	24.25	7-oxodehydroabietic acid

	25.08	Octacosane
	26.60	Nonacosane
	27.50	Nonacosane
	28.54	Tetratriacontane
	29.75	Hexatriacontane
B	15.57	Phenol, 2-(1,1-dimethylethyl)-4-(1,1,3,3-tetramethylbutyl)-
	23.02	Methyldehydroabietic acid
	23.37	Methyl abietate
	24.25	7-oxodehydroabietic acid
	25.08	Octacosane
	26.60	Nonacosane
	27.50	Nonacosane
	28.54	Tetratriacontane
	29.75	Hexatriacontane
C	9.69	Phenol, 2,4-bis-1,1-dimethylethyl

Table 2: Oil-specific chemical markers identified by GC-MS analysis. *(C:N) indicates the number of carbon atoms (C) and double bonds (N) in the fatty acid side chains.

Sample	Retention time (min)	Fatty acids methyl esters	C:N	%
Leone	14.28	Myristic acid	14:0	2.23
	18.31	Palmitic acid	16:0	36.14
	20.56	Oleic acid	18:1	28.81
	20.76	Stearic acid	18:0	32.82
Colatura	18.31	Palmitic acid	16:0	37.53
	20.56	Oleic acid	18:1	20.72
	20.76	Stearic acid	18:0	41.75
Deposit	18.08	Palmitic acid	16:0	12.69
	20.33	Vaccenic acid	18:1	6.85
	20.66	Stearic acid	18:0	11.13
	22.68	Arachidic acid	20:0	3.65
	24.42	Behenic acid	22:0	6.43
	25.98	Lignoceric acid	24:0	5.18
	26.49	Erucid acid	22:1	47.35
	27.64	Cerotic acid	26:0	3.10
	29.98	Montanic acid	28:0	3.62

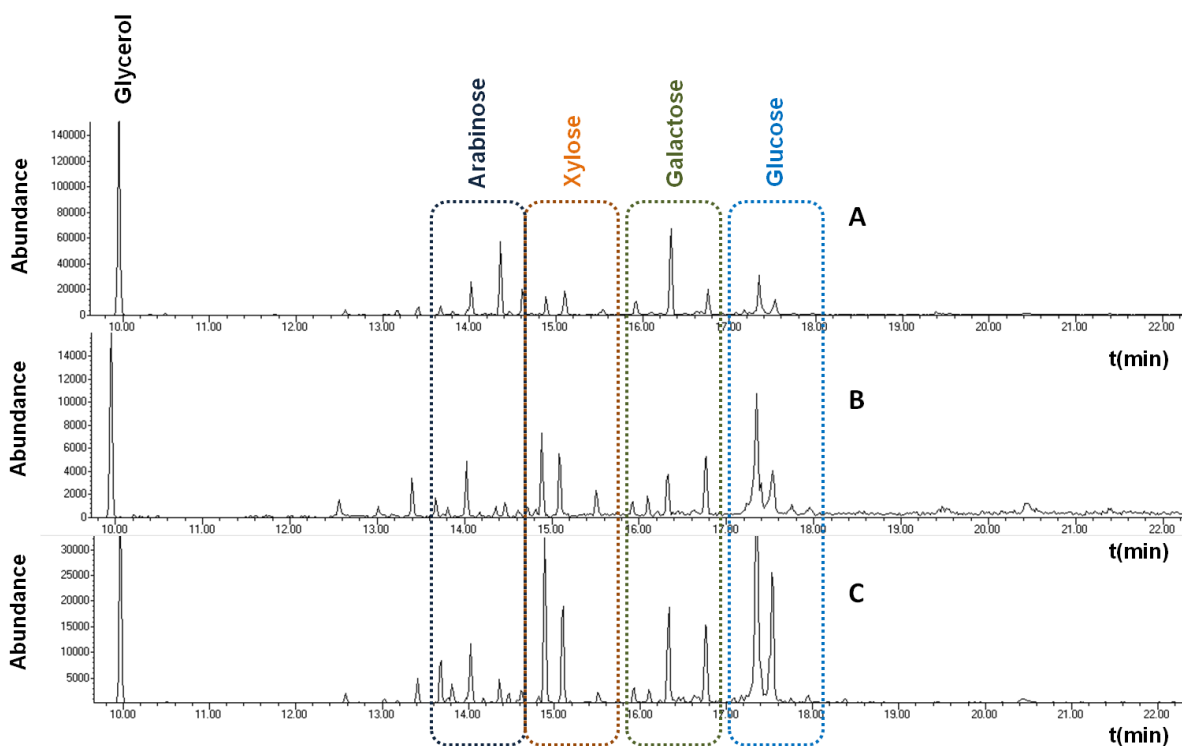


Figure 2: Total Ion Current (TIC) showing the profile of monosaccharides. A: Lion's foot; B: transparent pouring; C: Deposit material. Annotation of the sugars was performed by comparing the retention times and the fragmentation spectra of the sugars with analytical standards.

Table 3: Monosaccharides identified in the different samples of the Lion statue. The identification was based on the retention time and the fragmentation spectra of analytical standards. The percentage is the relative abundance of the sugar.

Compound	Retention time (min)	Transparent pouring (%)	Depositing material (%)	Lion's foot (%)
Glycerol	9.958	22.52	14.23	33.39
Arabinose	14.031	6.92	4.51	6.96
Arabinose	14.367	2.67	2.09	12.87
Xylose	14.892	9.40	12.72	3.21
Xylose	15.097	8.76	7.31	5.59
Xylose	15.516	5.01	1.70	1.32
Galactose	16.334	5.71	7.06	18.44
Galactose	16.771	8.45	7.19	5.41
Glucose	17.351	21.24	28.87	9.74
Glucose	17.536	9.32	14.32	3.08

Conclusions

GC-MS analysis suggests that the varnish used for the decoration of the lion in the Royal palace of Caserta consisted of waxes mixed with resins and rapeseed oil. Interesting is the presence of natural gum, that, however, will need further investigation to be confirmed. Furthermore, the simultaneous presence of abietanic acids, retene and methyl-dehydroabietate seems to indicate the presence of a product of destructive distillation of wood of the Pinaceae family. Also, the prevalence of non-oxygenated species as methyl abietate (MA) and methyl dehydro abietic acid (MDA) rather than oxygenated such as 7-oxo dehydro abietic acid, suggests that the varnish has been protected from oxidative processes. This protective action might be probably due to the use of wax. The mixture of waxes with oils and resins in decoration, in order to adjust the hardness of the wax, is well known [69,73,74].

References

- [1] G. Giorgi, Overview of Mass Spectrometric Based Techniques Applied in the Cultural Heritage Field, in: *Org. Mass Spectrom. Art Archaeol.*, John Wiley & Sons, Ltd, 2009: pp. 37–74. <https://doi.org/10.1002/9780470741917.ch2>.
- [2] I. Degano, F. Modugno, I. Bonaduce, E. Ribechini, M.P. Colombini, Recent Advances in Analytical Pyrolysis to Investigate Organic Materials in Heritage Science, *Angew. Chemie Int. Ed.* 57 (2018) 7313–7323. <https://doi.org/10.1002/anie.201713404>.
- [3] J. Blaško, R. Kubinec, B. Husová, P. Přikryl, V. Pacáková, K. Štulík, J. Hradilová, Gas chromatography/mass spectrometry of oils and oil binders in paintings, *J. Sep. Sci.* 31 (2008) 1067–1073. <https://doi.org/10.1002/jssc.200700449>.
- [4] S. Vahur, A. Teearu, P. Peets, L. Joosu, I. Leito, ATR-FT-IR spectral collection of conservation materials in the extended region of 4000–80 cm^{-1} , *Anal. Bioanal. Chem.* 408 (2016) 3373–9. <https://doi.org/10.1007/s00216-016-9411-5>.
- [5] N. Salvadó, S. Butí, M.J. Tobin, E. Pantos, A.J.N.W. Prag, T. Pradell, Advantages of the use of SR-FT-IR microspectroscopy: applications to cultural heritage., *Anal. Chem.* 77 (2005) 3444–51. <https://doi.org/10.1021/ac050126k>.
- [6] D. Lazidou, D. Lampakis, I. Karapanagiotis, C. Panayiotou, Investigation of the Cross-Section Stratifications of Icons Using Micro-Raman and Micro-Fourier Transform Infrared (FT-IR) Spectroscopy, *Appl. Spectrosc.* 72 (2018) 1258–1271. <https://doi.org/10.1177/0003702818777772>.
- [7] C. Daher, L. Bellot-Gurlet, A.-S. Le Hô, C. Paris, M. Regert, Advanced discriminating criteria for natural organic substances of cultural heritage interest: spectral decomposition and multivariate analyses of FT-Raman and FT-IR signatures., *Talanta.* 115 (2013) 540–7. <https://doi.org/10.1016/j.talanta.2013.06.014>.
- [8] M. I. M., *Atti e Memorie della Società Magna Grecia, 1926–30 (Vols. I–IV)*. Rome: Palazzo Taverna. Annual Subscription, 100 lire. - *L'opera della Società Magna Grecia nei primi 10 anni*. By U. Zanotti-Bianco. From *Annales Institutorum*, 1931. Rome: Palazzo Ricci, 19, *J. Hell. Stud.* (1932). <https://doi.org/10.2307/627206>.
- [9] P. Zancani Montuoro, *Heraion alla force del Sele: I: Altre metope del primo thesauros*, *Atti e Mem. Della Soc. Magna Grecia.* (1964).
- [10] P. Cennamo, M. Rosaria, B. Lumaga, G. Fatigati, A. Amoresano, A. Carpentieri, CONSERVATION SCIENCE A MULTIDISCIPLINARY ASSESSMENT TO INVESTIGATE A XXII, 11 (2020) 25–38.
- [11] R. Vinciguerra, A. De Chiaro, P. Pucci, G. Marino, L. Birolo, Proteomic strategies for cultural heritage:

From bones to paintings ☆, *Microchem. J.* 126 (2016) 341–348. <https://doi.org/10.1016/j.microc.2015.12.024>.

[12] S. Rossi, M. Petrelli, D. Morgavi, F.P. Vetere, R.R. Almeev, R.L. Astbury, D. Perugini, Role of magma mixing in the pre-eruptive dynamics of the Aeolian Islands volcanoes (Southern Tyrrhenian Sea, Italy), *Lithos.* 324–325 (2019) 165–179. <https://doi.org/10.1016/j.lithos.2018.11.004>.

[13] B. Jananee, V. Thangam, A. Rajalakshmi, Investigation of soils by thermal and spectroscopic analysis, *Chem. Eng. Commun.* (2019). <https://doi.org/10.1080/00986445.2019.1680370>.

[14] R.J. Cox, H.L. Peterson, J. Young, C. Cusik, E.O. Espinoza, The forensic analysis of soil organic by FTIR, *Forensic Sci. Int.* 108 (2000) 107–116. [https://doi.org/10.1016/S0379-0738\(99\)00203-0](https://doi.org/10.1016/S0379-0738(99)00203-0).

[15] M. Cavagna, R. Dell'Anna, F. Monti, F. Rossi, S. Torriani, Use of ATR-FTIR microspectroscopy to monitor autolysis of *Saccharomyces cerevisiae* cells in a base wine, *J. Agric. Food Chem.* 58 (2010) 39–45. <https://doi.org/10.1021/jf902369s>.

[16] S.H. Gordon, S.C. Whatel, B.C. Wheeler, C. James, Multivariate FTIR analysis of substrates for protein, polysaccharide, lipid and microbe content: Potential for solid-state fermentations, *Biotechnol. Adv.* 11 (1993) 665–675. [https://doi.org/10.1016/0734-9750\(93\)90035-L](https://doi.org/10.1016/0734-9750(93)90035-L).

[17] B. Wiyono, S. Tachibana, D. Tinambunan, Chemical Compositions of Pine Resin, Rosin and Turpentine Oil from West Java, Indones. *J. For. Res.* (2006). <https://doi.org/10.20886/ijfr.2006.3.1.7-17>.

[18] R.P. Evershed, C. Heron, L. John Goad, Analysis of organic residues of archaeological origin by high-temperature gas chromatography and gas chromatography - Mass spectrometry, *Analyst.* (1990). <https://doi.org/10.1039/an9901501339>.

[19] M.D. Petit-Dominguez, J. Martinez-Maganto, MCF fast derivatization procedure for the identification of resinous deposit components from the inner walls of roman age amphorae by GC-MS, *Talanta.* (2000). [https://doi.org/10.1016/S0039-9140\(99\)00330-6](https://doi.org/10.1016/S0039-9140(99)00330-6).

[20] C. Azemard, M. Menager, C. Vieillescazes, Analysis of diterpenic compounds by GC-MS/MS: contribution to the identification of main conifer resins, *Anal. Bioanal. Chem.* 408 (2016) 6599–6612. <https://doi.org/10.1007/s00216-016-9772-9>.

[21] J. Font, N. Salvadó, S. Butí, J. Enrich, Fourier transform infrared spectroscopy as a suitable technique in the study of the materials used in waterproofing of archaeological amphorae, *Anal. Chim. Acta.* 598 (2007) 119–127. <https://doi.org/10.1016/j.aca.2007.07.021>.

[22] I. Jerković, Z. Marijanović, M. Gugić, M. Roje, Chemical profile of the organic residue from ancient amphora found in the adriatic sea determined by direct GC and GC-MS analysis, *Molecules.* 16 (2011) 7936–7948. <https://doi.org/10.3390/molecules16097936>.

[23] F.J. Moreno, A. Montilla, M. Villamiel, N. Corzo, A. Olano, Analysis, structural characterization, and bioactivity of oligosaccharides derived from lactose, *Electrophoresis.* 35 (2014) 1519–1534. <https://doi.org/10.1002/elps.201300567>.

[24] G.J. Dutton, Glucuronide-Forming Enzymes, in: *Concepts Biochem. Pharmacol.*, Springer Berlin Heidelberg, 1971: pp. 378–400. https://doi.org/10.1007/978-3-642-65177-9_24.

[25] A. Goyon, J.Z. Cai, K. Kraehenbuehl, C. Hartmann, B. Shao, P. Mottier, Determination of steroid hormones in bovine milk by LC-MS/MS and their levels in Swiss Holstein cow milk, *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* 33 (2016) 804–816. <https://doi.org/10.1080/19440049.2016.1175186>.

[26] J. Orsavova, L. Misurcova, J. Vavra Ambrozova, R. Vicha, J. Mlcek, Fatty acids composition of vegetable oils and its contribution to dietary energy intake and dependence of cardiovascular mortality on dietary intake of fatty acids, *Int. J. Mol. Sci.* 16 (2015) 12871–12890. <https://doi.org/10.3390/ijms160612871>.

[27] J. Pollier, A. Goossens, Oleanolic acid, *Phytochemistry.* 77 (2012) 10–15. <https://doi.org/10.1016/j.phytochem.2011.12.022>.

[28] M.A. Lozano-Grande, S. Gorinstein, E. Espitia-Rangel, G. Dávila-Ortiz, A.L. Martínez-Ayala, Plant Sources, Extraction Methods, and Uses of Squalene, *Int. J. Agron.* 2018 (2018).

<https://doi.org/10.1155/2018/1829160>.

[29] S.K. Kim, F. Karadeniz, Biological Importance and Applications of Squalene and Squalane, in: *Adv. Food Nutr. Res.*, Academic Press Inc., 2012: pp. 223–233. <https://doi.org/10.1016/B978-0-12-416003-3.00014-7>.

[30] X. Zhang, S. Zhang, Y. Yang, D. Wang, H. Gao, Natural barrigenol-like triterpenoids: A comprehensive review of their contributions to medicinal chemistry, *Phytochemistry*. 161 (2019) 41–74. <https://doi.org/10.1016/j.phytochem.2019.01.017>.

[31] P.D. Uchil, A. Nagarajan, P. Kumar, β -galactosidase, *Cold Spring Harb. Protoc.* 2017 (2017) 774–779. <https://doi.org/10.1101/pdb.top096198>.

[32] M.P. Colombini, G. Giachi, F. Modugno, E. Ribechini, Characterisation of organic residues in pottery vessels of the Roman age from Antinoe (Egypt), in: *Microchem. J.*, 2005: pp. 83–90. <https://doi.org/10.1016/j.microc.2004.05.004>.

[33] J. Cuní, P. Cuní, B. Eisen, R. Savizky, J. Bové, Characterization of the binding medium used in Roman encaustic paintings on wall and wood, *Anal. Methods*. 4 (2012) 659–669. <https://doi.org/10.1039/c2ay05635f>.

[34] A. Pichot, Études de lettres Theatra et spectacula Théâtres et amphithéâtres : outils de romanisation en Maurétanie ?, in: [Http://Journals.Openedition.Org/Edl](http://Journals.Openedition.Org/Edl), Faculté des lettres de l'Université de Lausanne, 2011: pp. 171–192. <https://doi.org/10.4000/edl.113>.

[35] G. Renda, S. Gigli, A. Amato, S. Venticinque, B. Di Martino, F.R. Cappa, Mobile devices for the visit of anfiteatro campano in Santa Maria Capua Vetere, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Springer, Berlin, Heidelberg, 2012: pp. 281–290. https://doi.org/10.1007/978-3-642-34234-9_28.

[36] M. Gelzo, M. Grimaldi, A. Vergara, V. Severino, A. Chambery, A. Dello Russo, C. Piccioli, G. Corso, P. Arcari, Comparison of binder compositions in Pompeian wall painting styles from Insula Occidentalis, *Coke Chem.* 8 (2014) 65. <https://doi.org/10.1186/s13065-014-0065-0>.

[37] M. Gelzo, M. Grimaldi, A. Vergara, V. Severino, A. Chambery, A. Dello Russo, C. Piccioli, G. Corso, P. Arcari, Comparison of binder compositions in Pompeian wall painting styles from Insula Occidentalis, *Chem. Cent. J.* 8 (2014) 65. <https://doi.org/10.1186/s13065-014-0065-0>.

[38] G. Corso, M. Gelzo, A. Chambery, V. Severino, A. Di Maro, F.S. Lomoriello, O. D'Apolito, A. Dello Russo, P. Gargiulo, C. Piccioli, P. Arcari, Characterization of pigments and ligands in a wall painting fragment from Litternum archaeological park (Italy), *J. Sep. Sci.* 35 (2012) 2986–2993. <https://doi.org/10.1002/jssc.201200490>.

[39] A. Casoli, S. Santoro, Organic materials in the wall paintings in Pompei: A case study of Insula del Centenario, *Chem. Cent. J.* 6 (2012) 107. <https://doi.org/10.1186/1752-153X-6-107>.

[40] J. Cuní, What do we know of Roman wall painting technique? Potential confounding factors in ancient paint media analysis, *Herit. Sci.* 4 (2016) 44. <https://doi.org/10.1186/s40494-016-0111-4>.

[41] L. Birolo, A. Tomeo, M. Trifuoggi, F. Auriemma, L. Paduano, A. Amoresano, R. Vinciguerra, C. De Rosa, L. Ferrara, A. Giarra, A. Luchini, C. De Maio, G. Greco, A. Vergara, A hypothesis on different technological solutions for outdoor and indoor Roman wall paintings, *Archaeol. Anthropol. Sci.* 9 (2017) 591–602. <https://doi.org/10.1007/s12520-016-0408-y>.

[42] S. Gunasekaran, G. Anbalagan, S. Pandi, Raman and infrared spectra of carbonates of calcite structure, *J. Raman Spectrosc.* 37 (2006) 892–899. <https://doi.org/10.1002/jrs.1518>.

[43] H.M. Mahmoud, A MULTI-ANALYTICAL APPROACH FOR CHARACTERIZING PIGMENTS FROM THE TOMB OF DJEHUTYEMHAB (TT 194), (2013).

[44] H.H. Marey Mahmoud, M. Mahmoud, H. H., A Preliminary Investigation Of Ancient Pigments From The Mortuary Temple Of Seti I, El-Qurna (Luxor, Egypt), *MAA.* 11 (2011) 99–106. <https://ui.adsabs.harvard.edu/abs/2011MAA....11...99M/abstract> (accessed December 5, 2020).

[45] F. Ospitali, D. Bersani, G. Di Lonardo, P.P. Lottici, “Green-earths”: vibrational and elemental

characterisation of glauconites, celadonites and historical pigments, (2007) 40–41.

[46] R.J.H. Clark, P.J. Gibbs, K.R. Seddon, N.M. Brovenko, Y.A. Petrosyan, Non-destructive in situ identification of cinnabar on ancient Chinese manuscripts, *J. Raman Spectrosc.* 28 (1997) 91–94. [https://doi.org/10.1002/\(sici\)1097-4555\(199702\)28:2/3<91::aid-jrs67>3.0.co;2-x](https://doi.org/10.1002/(sici)1097-4555(199702)28:2/3<91::aid-jrs67>3.0.co;2-x).

[47] P. Baraldi, C. Baraldi, R. Curina, L. Tassi, P. Zannini, A micro-Raman archaeometric approach to Roman wall paintings, *Vib. Spectrosc.* 43 (2007) 420–426. <https://doi.org/10.1016/j.vibspec.2006.04.029>.

[48] E.A. Lalla, G. Lopez-Reyes, A. Sansano, A. Sanz-Arranz, J. Martínez-Frías, J. Medina, F. Rull-Pérez, Raman-IR vibrational and XRD characterization of ancient and modern mineralogy from volcanic eruption in Tenerife Island: Implication for Mars, *Geosci. Front.* 7 (2016) 673–681. <https://doi.org/10.1016/j.gsf.2015.07.009>.

[49] M. Alfè, V. Gargiulo, R. Di Capua, F. Chiarella, J.N. Rouzaud, A. Vergara, A. Ciajolo, Wet chemical method for making graphene-like films from carbon black, *ACS Appl. Mater. Interfaces.* 4 (2012) 4491–4498. <https://doi.org/10.1021/am301197q>.

[50] D.C. Smith, M. Bouchard, M. Lorblanchet, An initial Raman microscopic investigation of prehistoric rock art in caves of the Quercy District, S. W. France, *J. Raman Spectrosc.* 30 (1999) 347–354. [https://doi.org/10.1002/\(SICI\)1097-4555\(199904\)30:4<347::AID-JRS379>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-4555(199904)30:4<347::AID-JRS379>3.0.CO;2-A).

[51] D.L.A. De Faria, S. Venâncio Silva, M.T. De Oliveira, Raman microspectroscopy of some iron oxides and oxyhydroxides, *J. Raman Spectrosc.* 28 (1997) 873–878. [https://doi.org/10.1002/\(sici\)1097-4555\(199711\)28:11<873::aid-jrs177>3.0.co;2-b](https://doi.org/10.1002/(sici)1097-4555(199711)28:11<873::aid-jrs177>3.0.co;2-b).

[52] S. Das, M.J. Hendry, Application of Raman spectroscopy to identify iron minerals commonly found in mine wastes, *Chem. Geol.* 290 (2011) 101–108. <https://doi.org/10.1016/j.chemgeo.2011.09.001>.

[53] I.M. Bell, R.J.H. Clark, P.J. Gibbs, Raman spectroscopic library of natural and synthetic pigments (pre- \approx 1850 AD), *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 53 (1997) 2159–2179. [https://doi.org/10.1016/S1386-1425\(97\)00140-6](https://doi.org/10.1016/S1386-1425(97)00140-6).

[54] J. Jehlička, H.G.M. Edwards, K. Osterrothová, J. Novotná, L. Nedbalová, J. Kopecký, I. Němec, A. Oren, Potential and limits of Raman spectroscopy for carotenoid detection in microorganisms: Implications for astrobiology, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 372 (2014). <https://doi.org/10.1098/rsta.2014.0199>.

[55] Ugo Zezza, *Petrografia microscopica*, La Goliardica Pavese, 1976.

[56] A.P.D. Peccerillo, *Introduzione alla petrografia ottica*, Morlacchi Editore, 2003. <https://www.amazon.it/Introduzione-alla-petrografia-ottica-CD-ROM/dp/8888778276> (accessed December 5, 2020).

[57] N. Szczepańska, P. Kubica, B. Kudłak, J. Namieśnik, A. Wasik, Stabilities of bisphenol A diglycidyl ether, bisphenol F diglycidyl ether, and their derivatives under controlled conditions analyzed using liquid chromatography coupled with tandem mass spectrometry, *Anal. Bioanal. Chem.* 411 (2019) 6387–6398. <https://doi.org/10.1007/s00216-019-02016-5>.

[58] V. Beltran, N. Salvadó, S. Butí, T. Pradell, Ageing of resin from Pinus species assessed by infrared spectroscopy, *Anal. Bioanal. Chem.* 408 (2016) 4073–4082. <https://doi.org/10.1007/s00216-016-9496-x>.

[59] G. Kaklamanos, G. Theodoridis, T. Dabalís, Determination of anabolic steroids in bovine urine by liquid chromatography-tandem mass spectrometry, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 877 (2009) 2330–2336. <https://doi.org/10.1016/j.jchromb.2009.03.033>.

[60] H. Zhang, J.M. Yue, Hasubanan type alkaloids from *Stephania longa*, *J. Nat. Prod.* 68 (2005) 1201–1207. <https://doi.org/10.1021/np0500833>.

[61] P. Ranchana, M. Ganga, M. Jawaharlal, M. Kannan, Investigation of Volatile Compounds from the Concrete of *Jasminum auriculatum* Flowers, *Int.J.Curr.Microbiol.App.Sci.* 6 (2017) 1525–1531. <https://doi.org/10.20546/ijcmas.2017.611.180>.

[62] H.M. Usman, N. Abdulkadir, M. Gani, H.M. Maiturare, Bacterial pigments and its significance, (2017). <https://doi.org/10.15406/mojbb.2017.04.00073>.

- [63] P. Sabbagh, A. Ebrahimzadeh Namvar, The eminence status of bacterial pigments under different aspects, *Microbiol. Medica*. 32 (2017). <https://doi.org/10.4081/mm.2017.7247>.
- [64] Il Palazzo - Sito web non ufficiale della Reggia di Caserta, (n.d.). <https://www.reggiadicasertaunofficial.it/it/reggia/il-palazzo/> (accessed November 24, 2020).
- [65] 18th-Century Royal Palace at Caserta with the Park, the Aqueduct of Vanvitelli, and the San Leucio Complex - Documents - UNESCO World Heritage Centre, (n.d.). <https://whc.unesco.org/en/list/549/documents/%23ABevaluation> (accessed November 24, 2020).
- [66] Dictionnaire amoureux de Versailles - Franck FERRAND - Google Libri, (n.d.). https://books.google.it/books?id=qTdlAQAAQBAJ&redir_esc=y (accessed November 24, 2020).
- [67] J. Mazurek, M. Svoboda, M. Schilling, GC/MS Characterization of Beeswax, Protein, Gum, Resin, and Oil in Romano-Egyptian Paintings, *Heritage*. 2 (2019) 1960–1985. <https://doi.org/10.3390/heritage2030119>.
- [68] Rapid differentiation of beeswaxes using EGA-MS and Py-GC / MS, (n.d.) 8862.
- [69] M.P. Colombini, G. Giachi, F. Modugno, E. Ribechini, Characterisation of organic residues in pottery vessels of the Roman age from Antinoe (Egypt), *Microchem. J.* 79 (2005) 83–90. <https://doi.org/10.1016/j.microc.2004.05.004>.
- [70] I. Jerković, Z. Marijanović, M. Gugić, M. Roje, Chemical profile of the organic residue from ancient amphora found in the adriatic sea determined by direct GC and GC-MS analysis, *Molecules*. 16 (2011) 7936–7948. <https://doi.org/10.3390/molecules16097936>.
- [71] Y. Zhang, G. Wu, C. Chang, Y. Lv, W. Lai, H. Zhang, X. Wang, Q. Jin, Determination of Origin of Commercial Flavored Rapeseed Oil by the Pattern of Volatile Compounds Obtained via GC–MS and Flash GC Electronic Nose, *Eur. J. Lipid Sci. Technol.* 122 (2020) 1–7. <https://doi.org/10.1002/ejlt.201900332>.
- [72] A. Lluveras-Tenorio, J. Mazurek, A. Restivo, M.P. Colombini, I. Bonaduce, The Development of a New Analytical Model for the Identification of Saccharide Binders in Paint Samples, *PLoS One*. 7 (2012) e49383. <https://doi.org/10.1371/journal.pone.0049383>.
- [73] S. Bogdanov, B.P. Science, Beeswax : Uses and Trade, *The Beeswax*. (2009) 1–16.
- [74] C. Brøns, K.L. Rasmussen, M.M. Di Crescenzo, R. Stacey, A. Lluveras-Tenorio, Painting the Palace of Apries I: Ancient binding media and coatings of the reliefs from the Palace of Apries, Lower Egypt, *Herit. Sci.* 6 (2018) 1–20. <https://doi.org/10.1186/s40494-018-0170-9>.
- [8] A. Lluveras-Tenorio, J. Mazurek, A. Restivo, M.P. Colombini, I. Bonaduce, Analysis of plant gums and saccharide materials in paint samples: Comparison of GC-MS analytical procedures and databases, *Chem. Cent. J.* 6 (2012) 115. <https://doi.org/10.1186/1752-153X-6-115>.

6. Final conclusions

Conservation of cultural heritage is essential for the future generations as the diverse cultures; habits of the past and the ecosystem's evolution over the years give essential information about where we come from and who we are. Mass spectrometry, as one of the latest developed analytical techniques, has significantly contributed and contributes to achieving that aim nowadays. This Ph.D. project aimed to contribute to this scope by developing proteomic analytical strategies using mass spectrometry in archaeological objects and works of art. These strategies were used as such or combined with data obtained by other analytical approaches for a more in-depth insight of material origin and use as well as for the study of chemical degradation mechanisms taking place over the years. Each project, involved in this work, aimed to open a window into the past, revealing important data for each specific case of study, also trying to face the different types of challenges in the protein identification and protein characterization, encountered in the palaeoproteomic field.

The development of the micro-invasive kit for the protein identification and characterization in ancient and historical objects, described in chapter 4, addressed the challenge of micro-sampling that is crucial in the world of diagnosis of cultural heritage objects. Given the value and sensitivity of works of art, sampling needs to be as less invasive as possible, because in most cases, it is carried out on small pieces of the archaeological object. Taking advantage of some fungal proteins with adhesive proteins that turn a hydrophobic surface to hydrophilic and vice versa, forming stable films, we functionalized cellulose acetate sheets with hydrophobins, immobilizing latter trypsin. The developed kit includes a 10 min protocol for *in-situ* protein hydrolysis, is user-friendly, and most importantly, does not require the removal of the ancient object, but instead, sampling can be performed *on-site*. Moreover, because of its simplicity, can be carried out also by restorers without any need of specific knowledge in protein chemistry or mass spectrometry. The amount of extracted peptides is sufficient not only for protein identification but also for the study of protein chemical modifications by standard MS techniques. That allows characterizing the sample for the state of degradation. Furthermore, we enhanced the protocol by introducing a deglycosylation step for the removal of the sugar layer that is sometimes present in archaeological/artistic samples, facilitating the access of trypsin to proteins. The proposed functionalized film is still amenable to further improvements using for instance, more flexible inert support materials, giving future perspectives and enhancements of this methodology. This project was described by the European Union as one of the most innovative EU-funded projects with a "remarkable" level of market capability, ranking it among the most advanced technology development processes (<https://www.innoradar.eu/innovation/36157>).

Other major challenges in palaeoproteomics and tandem mass spectrometry identification are the right choices of bioinformatic tools and protein databases to assign potential peptide sequences to MS/MS spectra. Chapter 3 dealt with these challenges during the analysis of collagenous materials such as ancient human bones from the eruptive Vesuvius area and animal glues that are used for restoration purposes. For instance, the collagen identification of animal glues required apart from using standard proteomic tools such as MASCOT, the use of protein alignment tools such as BLAST for confident identification of the actual collagen origin in a quality control perspective of the analysis. Difficult attribution originates from the extremely high similarity of collagen sequences among the species that makes tricky any species discrimination. Attention was given to what really means a unique peptide during collagen discrimination between two animal species (i.e., bovine and porcine collagen, for instance). A standard MS/MS protein search is often based on the selection of unique peptides that identify the protein family (i.e. COL1A1 of mammals) rather than the single type of collagen (i.e. bovine, porcine etc.), something that could lead to a wrong protein assignment. For this reason, further validation of the assigned as unique peptides is needed. Moreover, to restrict the search space, we created a handmade database, collecting collagen sequences of animal species that could be reasonably found in animal glues. On the other side, the protein identification in case of the ancient human bones was easier since only proteins of *Homo sapiens* were identified, yet it was to the best of our knowledge the first proteomic study of burned bones in the literature. Interestingly, the absence of non-collageneous proteins in the bones from Pompeii in comparison to those from Herculaneum, suggest a different bone preservation that may reflect on the different temperature and pressure that the victims were exposed to during the Vesuvius eruption, clearly demonstrating that bone diagenetic alterations are written in bone proteins.

A further challenge in protein characterization is the identification of chemical modifications. In the study of both ancient human bones and animal glues we used in-house java-based bioinformatic tools and standard proteomic tools for the characterization of collagen chemical modifications. First of all, we studied degradation phenomena, commonly observed in aged proteins, such as oxidation of Met, deamidation of Gln and Asn, and backbone cleavage of the polipeptidic chains. Both cases demonstrated that collagen degradation mechanisms are quite complex also because of the collagen structure's peculiarity. The repetitive sequence and the high presence of hydroxyprolines seem to play a key role in collagen's stability and, thus, to its susceptibility to mortem and post-mortem chemical modifications. To have a deeper insight into collagen chemical modifications, each study was carried out by calculating the modification both as total collageneous content and in each single aminoacidic position. Many results were obtained, some of which never discussed before.

From the global evaluation of deamidation and the calculation of deamidation at single deamidation sites along collagen sequences, we discovered in ancient human bones, that there are some zones where deamidation is more pronounced than others, suggesting that three-dimensional arrangement might affect the local deamidation level. Also, in the same project, the fewer proteins were found the more deamidated the sample was found to be. In terms of accuracy, other archeological bones that come from a natural death were analyzed, and a comparative study was performed. The three-dimensional effect of collagen on deamidation level can also be observed in animal glues analysis. However, in this case study, the challenges arise from the observation that rather frequently mixtures of collagen were present, impairing an actual and confident evaluation and distribution of deamidation levels.

To further whether if the three-dimensional arrangement of collagen affects collagen deamidation, complementary studies were performed. In the case of the ancient human bones, a mass screening of 56 archaeological human bones from the Vesuvius area was performed in search for the most frequently detected and/or deamidated peptides and the results were analyzed on a statistical basis in order to select the best peptides to be later used in the quantification of deamidation. By this way we would be able to better evaluate deamidation levels in comparative analyses. Collaboration with Prof. Matthew Collins in the University of Cambridge is in progress to extend this approach to more than 100 collagenous materials, others than bones, such as animal glues and parchment, extracted and analyzed with different instrumentation provide a common tool for collagen materials.

Backbone cleavage of the polypeptide chain is also considered as a common degradation feature in ancient proteins, and for this reason evaluation in the basis of detection of the semitryptic peptides along the several types of collagen sequences was performed. In case of human bones, it seems that collagen's three dimensional arrangement may also affect collagen spontaneous hydrolysis since in some zones a high frequency of cleavage sites was detected. Additionally, in case of animal glues, the more vigorous collagen extraction of strong glues seems to be reflected in the collagen hydrolysis in comparison to hide glues that involve "softer" treatment of the material.

Moreover, a big challenge in protein characterization is the discovery of unexpected chemical modifications that requires an untargeted approach. In case of ancient human bones from the Vesuvius area, the data-dependent peptide algorithm of MaxQuant was used for an unbiased discovery of chemical modifications that were ranked by their occurrence within the dataset. Diagenetic increase of AGEs correlates with oxidative conditions, and extensive glycation products were observed in the samples from the volcanic area, always more pronounced in samples from Pompeii, likely originating from reactions with the sugars originally

present in the extracellular matrix. However, because of the high instability of the formed imidazolones during the years and under extremely high pH makes their identification more tricky. In addition, we observed oxidative processes to occur on prolines and hydroxyprolines, whose detection will need further development of dedicated bioinformatic tools. The high intensities of hydroxyprolines indeed mask the identification of other modifications. The integration of Hyp as a separate amino acid in the protein databases might be a solution to the above challenge bringing more light in this poorly considered degradation phenomenon. It is also worth to be mentioned that all the aforementioned discovered modifications appear to be more pronounced in specific regions. This suggests a strong three-dimensional effect in directing chemical modification events, which is completely in line with deamidation results. These data, including deamidation (N,Q), oxidation(M), and backbone cleavage data demonstrated that molecular differences exist and can be seen as a perspective on the chemist's approach to read through the processes that alter proteins in bone during burial, the history written in the molecules, a sort of "chemical life history tracer".

The last chapter of this Ph.D thesis explores the value of integrating mass spectrometry analyses to spectroscopic and spectrometric analyses for the molecular identification and characterization of historical paintings, vessels, marble sculptures and an Egyptian wooden sarcophagus. Cultural heritage objects comprise a wide variety of materials (metals, ceramics, glass, various igneous and sedimentary rocks, textile, leather, wood, horn, parchment, paper, etc.) and usually exhibit a fairly complex three-dimensional structure and a heterogeneous chemical composition. For this reason a comprehensive study involves a multidisciplinary approach with the combination of different micro-destructive (as much as possible) techniques, each of which may reply to a different question regarding material's chemical composition, yet all complete each other, revealing important information on the state of material and the reason of its use in the past, contributing therefore to create our cultural heritage. A very much felt problem that has been highlighted in this chapter concern the sample contamination that makes extremely tough the molecule assignment and for this reason the use of controls and analytical standards is required. The major part of the molecules can derive from both natural and biological sources and the scientist has to be critical in the interpretation of the obtained data. In some cases, the existence of more than one scenario is plausible.

7.PhD Course Activity Summary

Attended courses

“Recombinant Production of Natural and Mutant Proteins” 8hours, by Prof. Angela Duilio”, Department of chemical sciences, March 2018

“Mass Spectrometry”, 16 hours, by Prof. Piero Pucci, Department of chemical sciences, July 2018

“After your PhD what? Grant writing, job market survey and networking” 16 hours , by Prof. Enrico Cappelini, Department of chemical sciences, February 2019

“Innovation, Entrepreneurship and Career Development, 16 hours, by Prof. Enrico Cappelini, Crete, Greece, September-October2019

Glucocchemistry, 16 hours, by Prof. Emiliano bedini, Department of chemical sciences, January 2020

“Crystallographic studies to delineate the mode of action of anticancer metallodrugs” , 16 hours, by Prof. Antonio Merlino, online, June 2020

“Synthesis, structure and applications of natural and modified oligonucleotides” 16 hours, by Prof. Daniela Montesarchio, online, June 2020

Compulsory training on safety in the work environment for workers particularly exposed to risk described by the “ART.37-comma 1 lett B of the LGS 81/2008” and the point 4 of the “accordo stato regione del 21/12/2011”by Prof Ettore Nardi, Online compulsory PhD course, 16 hours, July 2020

Attended seminars

“Supported metal particles in catalysis: Role of metal-oxide interface” by Prof. Gianfranco Pacchioni, Department of chemical sciences, 15 February 2018

“Chimica alla nascita nella ricerca e diagnostica clinica”,by Prof. Sara Boenzi, Department of chemical sciences, 9 April 2018

“Gli inquinanti organici persistenti”, by Prof. Anna Andolfi, Department of chemical sciences, 14 September 2018

“Introduction to University Pedagogy English Spoken”, by PhD School of SCIENCE, University of Copenhagen, 8 March 2019.

“Applicazione di LCA al settore della chimica industriale: alcuni esempi di successo”, organized by Società Chimica Italiana, held online in date 16/9/2020, by Prof. Stefania Albonetti from the University of Bologna, Italy

‘SP3 meets iST: Automated sample preparation with SP3 for reproducible proteomics in biology and the clinic’ which was held on October 6th, 2020, by PreOmics GmbH

Synthetic Carbohydrate Materials", Dr.ssa Martina Delbianco del Max Planck Institute of Colloids and Interfaces ,23 ottobre 2020

“Sintesi organica e biologia sintetica: “thinking outside the box” in opto-elettronica”, organized by Società Chimica Italiana, held online in date 20/11/2020, by Prof. Gianluca Maria Farinola, from Università degli Studi di Bari Aldo Moro

"Plasmonic-Enhanced Raman Scattering and its Applications in Molecular Detection, Biodiagnostic and Cultural Heritage" held online in date il 27 novembre 2020 by Prof. Santiago Sánchez-Cortés , from IEM Institute of Structure of Matter - Department of Vibrational Spectroscopy and Multiphotonic Processes

“Ogni cosa è illuminata’: Un ‘turista per caso’ nella catalisi fotoredox” organized by Società Chimica Italiana, held online in date 27/11/2020, by Prof. Pier Giorgio Cozzi from the University of Bologna, Italy

Attended congresses /summer schools

MaxQuant summer school on 8-15 July 2018 in Barcelona

IMSC2018 International Mass Spectrometry Conference on 26th 31st August 2018, in Florence. Oral presentation (MOLECULAR SIGNATURES OF PROTEIN AGING IN ANCIENT HUMAN BONES FROM THE ERUPTION AREA OF VESUVIUS. Georgia Ntasi, Piero Pucci, Gennaro Marino, Leila Birolo).

European Night of researchers on 28 November 2018, Museum of Cappella Sansevero, in Naples

ASMS2019 – American Society for Mass Spectrometry on 31th - 4st May 2019, in Atlanta,USA. Oral presentation “A minimally invasive and portable tool for MS identification of proteins in ancient paintings” Ntasi G, Cicatiello P, Marino G, Giardina P, Birolo

Summer school “Advanced Mass Spectrometry applied to cultural heritage” 17th – 21st June, 2019, Bordeaux, France

Periods abroad

Secondment of 5 months in the laboratory of Prof. Enrico Cappellini in the University of Copenhagen, Section for Evolutionary Genomics (March-July, 2019), Denmark

Support for teaching activity

I have experience teaching by mentoring after my Prof Leila Birolo students at both undergraduate and graduate level. Encouraging free expression, creativity, and directly participation in the experiments, students were introduced in the field of paleoproteomics, completing their thesis with worthy results. Also teaching activities has already taken place on inside lab projects in an attempt to collaborate and expand the paleoproteomics application. In total during these three years I was involved in the supervision of 7 undergraduate projects in Chemistry, 1 undergraduate project in Molecular and industrial Biotechnology and 4 postgraduate projects in Chemistry although one postgraduate program is still in progress.

Publications

- **Published**

[1] R. Linn, I. Bonaduce, G. Ntasi, L. Birolo, A. Yasur-Landau, E.H. Cline, A. Nevin, A. Lluveras-Tenorio, Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings, *Angew. Chemie Int. Ed.* 57 (2018) 13257–13260. <https://doi.org/10.1002/anie.201806520>.

[2] P. Cennamo, M. Rosaria, B. Lumaga, G. Fatigati, A. Amoresano, A. Carpentieri, CONSERVATION SCIENCE A MULTIDISCIPLINARY ASSESSMENT TO INVESTIGATE A XXII, 11 (2020) 25–38. ISSN: 2067-533X

[3] P. Cicatiello, G. Ntasi, M. Rossi, G. Marino, P. Giardina, L. Birolo, Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings, *Anal. Chem.* 90 (2018) 10128–10133. <https://doi.org/10.1021/acs.analchem.8b01718>.

[4] G. Ntasi, D.P. Kirby, I. Stanzione, A. Carpentieri, P. Somma, P. Cicatiello, G. Marino, P. Giardina, L. Birolo, A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects, *J. Proteomics.* 231 (2021). <https://doi.org/10.1016/j.jprot.2020.104039>.

- **Under Submission**

[1] Written in bone proteins: molecular signatures in the skeletons from Herculaneum and Pompeii; Ntasi G; Cappellini E, Marino G, Birolo L .manuscript in preparation

[2] Selection of peptide biomarkers in human bone collagen for the evaluation of deamidation (N, Q) and oxidation (M) by targeted proteomics; Ntasi G, Palomo. I.P, De Chiaro A, Marino G, Petrone P.P and Birolo L. manuscript in preparation

[3] Molecular Characterization of animal glues by a proteomic approach. Ntasi G; Sbriglia S, Marino G, Birolo L manuscript in preparation

[4] Palaeoproteomic analysis of bone and teeth, Ntasi G, Welker F,chapter of the upcoming book: Proteomic analysis in cultural heritage , chemical review under submission, Springer Nature

[5] Organic remains in amphorae from the temple of Hera in Paestum shed some lights on solemn ceremonies Ntasi G, Birolo L ,Vergara A, manuscript in preparation

Innovations

Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings ;Ntasi G, Cicatiello P, Rossi M, Marino G, Giardina P, Birolo L ,Anal Chem. (2018) (<https://www.innoradar.eu/innovation/36157>)