UNIVERSITA' DEGLI STUDI DI NAPOLI

FEDERICO II

TESI DI DOTTORATO IN SCIENZE CHIMICHE

XXVIII CICLO

Proteomics and Cultural Heritage

Roberto Vinciguerra

Tutor

Prof. Leila Birolo

Coordinatore del XXVIII ciclo

Prof. Luigi Paduano

"Art is the act of triggering deep memories
of what it means to be fully human."

David Whyte

Aknowledgments

Department of Chemical Sciences

Prof. Leila Birolo
Prof. Angela Amoresano
Prof. Piero Pucci
Prof. Gennaro Marino

Prof. Alessandro Vergara

University of Naples
'Federico II'

DATABENC
Distretto ad Alta Tecnologia dei Beni Culturali

Department of Humanistic Studies
Prof. Giovanna Greco

Prof. Paola Villa
School of Geography,
Archaeology and Environmental Studies,
University of the Witwatersrand,
South Africa

MUSEUM OF NATURAL HISTORY
UNIVERSITY OF COLORADO

University of Colorado
Museum

Department of Chemistry
and Industrial Chemistry

Prof. Maria Perla Colombini
Prof. Ilaria Degano
Dr. Ilaria Bonaduce
Dr. Anna Lluveras Tenorio

University of Pisa

My colleagues:

Dr. Addolorata De Chiaro
Dr. Anna Illiano
Dr. Eugenio Galano
Dr. Andrea Carpentieri
Dr. Francesca Pane
Dr. Carolina Fontanarosa
Dr. Matteo Imbelloni
Dr. Grazia Lettieri

Index

Summary

Proteins represent an important constituent in many art and archaeological objects. Notable proteinaceous materials, such as animal glue, egg (both yolk and albumen) and milk have long been used in artwork as painting binders, as adhesives for gildings, or they were included in mortars and paint grounds. The identification of protein-based material is an important information to understand the manufacturing process; it provides insights in the technique used by an artist, essential information for art historians. Moreover, it may also support the choice of the most appropriate conservation or restoration procedures.

Proteomics is typically considered to be associated with the study of living organisms. However, its inherently multidisciplinary nature has recently led to its application to oddly assorted areas ranging from forensics, food analysis, and clinical medicine and even for studying the origins of life [1].

In recent years, proteomics procedures have become increasingly popular for the characterization of proteinaceous materials in ancient samples of several cultural heritage objects.

Proteomics for cultural heritage is still in its infancy, with the first paper dating back to the early 2000 [2]. In particular, protocols routinely applied for typical modern samples still need to be fully adapted to take into account the low amount of proteinaceous material, the heterogeneity and the unusual physical state of the samples, as well as the high levels of damage found in ancient samples.

Modern mass spectrometry instrumentations are perfectly adequate to afford the minimal quantities of ancient samples. A micro-invasive protocol, in fact, proved to be successful when applied to fragments of paintings from the collapsed vault of the Basilica di S. Francesco in Assisi [3], and samples collected from the Camposanto Monumentale in Pisa [4]. Generally, by the use of these methods, we can easily identify complex protein mixtures in very complex and heterogeneous matrix.

Nevertheless, samples coming from artwork have an intrinsic contamination problem that originates from environmental exposure. In addition, the physical state of the samples, enormously different from the natural environment of proteins, and the degradation processes undergone during aging provide unusual problems that require to be counteracted.

Therefore, all the steps of the proteomic procedure need to be thoughtfully adapted, from the optimization of specific protocols for sample preparation to the development of data analysis tools that can cope with ancient or damaged samples. In this perspective, my PhD project was

devoted to this challenge as well as to the specific application of the developed strategies to diverse samples from works of art.

After a brief introduction, a description of the meaning of Cultural Heritage is reported in Chapter 2 of this Ph.D. thesis, since this is the basis of the scientific interest in the search of new investigation procedures.

Since many difficult problems arise from the complex composition and alteration pathways of the materials considered, this Chapter also provides the information of organic materials used in both art and archaeology, focusing the attention on protein-based materials.

To understand the potential of proteomics techniques applied on protein-based materials in Cultural Heritage objects, a description of the common proteomics methods is given in Chapter 3. Moreover, to help the understanding of the proteomic strategies developed and the various mass spectrometric methods applied, in this Chapter the main concepts in mass spectrometric instrumentation are also discussed, whit a special emphasis on the selective chemical modification of proteins of interest in order to understand their behaviour under natural ageing.

The aim of this thesis project is focused in the Chapter 4, devoted to the development of a unique method, efficient and applicable to different kinds of samples containing proteinaceous materials [5-6].

In the Chapter 5, a description of different study cases demonstrate the success of the experimental approach developed [7].

As a further section of my project, I worked to the development of a dedicated peptide database and targeted MS method (Multiple Reaction Monitoring, MRM), as described in Chapter 6. The need to develop this method comes from the intrinsic analytical disadvantages due to the complexity of these unusual samples, in terms of chemical composition, presence of contaminating protein-based materials, as well as the high levels of molecular damage found in ancient samples.

This method can be considered the first example of the application of targeted proteomics to samples in the field of cultural heritage that will allow an increase in sensitivity of protein detection in highly complex and rather contaminated samples.

The improvement in selectivity and sensitivity achieved with the targeted approach in respect to classical profiling experiments could pave the way to the possible application of this most advanced analytical strategy in the field of archaeology and artworks.

References

1. en Have, S. T., Ahmad, Y. & Lamond, A. I. Proteomics– Current Novelties and Future Directions. *J. Anal. Bioanal. Tech.* **01,** 1–6 (2011).
2. Ostrom, P. H. *et al.* New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. *Geochim. Cosmochim. Acta* **64,** 1043–1050 (2000).
3. Leo, G. *et al.* Proteomic strategies for the identification of proteinaceous binders in paintings. *Anal. Bioanal. Chem.* **395,** 2269–2280 (2009).
4. Leo, G. *et al.* Deamidation at asparagine and glutamine As a major modification upon deterioration/aging of proteinaceous binders in mural paintings. *Anal. Chem.* **83,** 2056–2064 (2011).
5. Vinciguerra, R., De Chiaro, A., Pucci, P., Marino, G. & Birolo, L. Proteomic strategies for cultural heritage: form bones to paintings. *Microchem. J.* **126,** 341–348 (2015).
6. Vinciguerra, R. *et al.* Deglycosylation Step to Improve the Identification of Egg Proteins in Art Samples. *Anal. Chem.* (2015). doi:10.1021/acs.analchem.5b02423
7. Villa, P. *et al.* A Milk and Ochre Paint Mixture Used 49,000 Years Ago at Sibudu, South Africa. *PLoS One* **10,** e0131273 (2015).

I

Introduction

The continuous research of new survey methods for Cultural Heritage materials is a constantly expanding field. An analytical chemistry approach to art and archaeological objects started around 1780 because of the progressive practical application of the ideas of art historians like Johann Winck-Lemann (1717–1768) on art and technical history. According to these historians, the knowledge of objects should be based on the examination of the artwork itself rather than on a text-based methodology[1]. Subsequently, analytical techniques and methods applied to the study of artworks have constantly grown, particularly in the mid-late 20th century and, thus, nowadays there is a wide variety of scientific methodologies essential for the recovery of useful information about the history and conservation of cultural heritage.

It becomes obvious to all that the knowledge and progress of our society arise from deeper knowledge of our past, represented by artworks as frescoes or dry paintings, archaeological finds of previous civilizations as old manufactures or also by human and animal remains.

Every artwork or manufactured, particularly but not only of artistic interest, contains in itself the history of the building, including information on its production, on its use and its state of preservation. Similarly, ancient human and animal remains and in particular bones may be considered a time capsule of ancient biomolecules owing to their natural resistance to *post mortem* decay[2,3], that provide an avenue of delineating evolutionary divergence from extant species and an indications of varying diet and provenance in the life of the individual.

The study, conservation and restoration of materials of historico-cultural value, require a well-defined need of analytical methods (Figure I.1), which are able to provide detailed information on:

- the chemical composition of selected parts of cultural heritage materials in order to elucidate their provenance;
- the state of alteration (on the surface and/or internally) of objects as a result of aging or short-, medium- and long-term exposure to particular environmental conditions;
- the effect of conservation/restoration strategies during and after application.

Chemistry, in particular, can offer sophisticated techniques to study the properties more 'intimate' of the materials constituting a work of art or an archaeological find and their state of degradation, helping the development of innovative processes and products involved in the

early stages of restoration. It is an essential role as fascinating to know and preserve the historical and artistic cultural heritage.

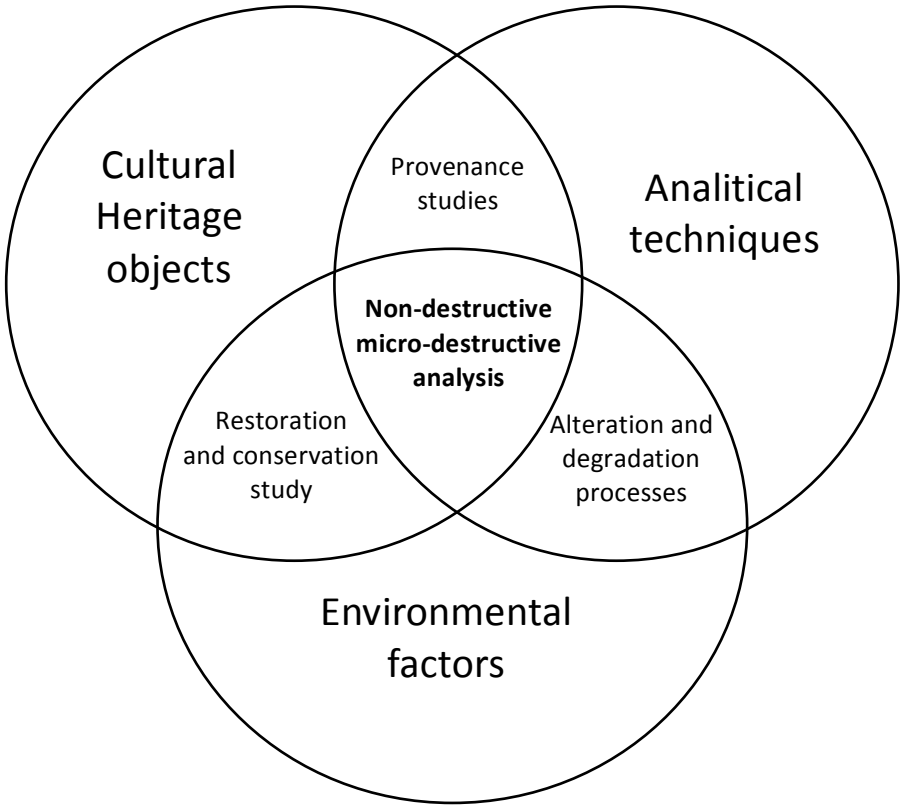


Figure I.1. Graphical representation of the interactions between cultural heritage materials, and the use of analytical techniques and environmental factors.

Unaware of their chemical characteristics, ancient civilizations and great exponents of the art world have adopted pictorial techniques that exploit the combination of natural pigments with organic compounds. A wide class of organic compounds (oils, resins, gums, waxes, and proteinaceous substances) were historically used in paintings given their ability to form uniform and flexible thin films (imprimatur or paint layers) when mixed with pigments and colouring materials on the appropriate supports [4]. Such naturally occurring substances used as binding media or protective coatings have attracted the attention of the conservation profession because of their considerable ability of undergoing physicochemical changes upon ageing.

The complexity of these materials makes the approach to their scientific investigation quite challenging, particularly when considering that analytical studies in the science of cultural heritage often involve invaluable works of art, for which destructive sampling techniques

should be avoided or severely limited, wherever possible. Consequently, scientists have made great efforts to put forward either non-destructive or micro-destructive analytical approaches to characterize painting materials and their alteration products.

An ideal method to analyse Cultural Heritage objects should be:

1. non-destructive or micro-destructive, i.e., it should respect the physical integrity of the object. Often valuable objects can only be investigated when the analysis does not result in any visible damage to the object. Usually, this completely eliminates sampling or limits it to very small amounts;
2. fast, so that large numbers of similar objects may be analysed or a single object investigated at various positions on its surface; this property is very valuable since this is the only way of being able to obtain the homogeneity of the data;
3. universal, so that by means of a single pre-treatment and analysis, many materials and objects of various shapes and dimensions may be investigated;
4. versatile, allowing with the same technique to obtain average compositional information as well as local information of small areas from materials with heterogeneous matrixes;
5. sensitive, so that analysis can be done by means of not only principal elements but also trace-element fingerprints;
6. multi-elemental, so that in a single measurement, information on many elements is obtained simultaneously and, more importantly, information is also obtained on elements which were not initially thought to be relevant to the investigation.

Among the naturally occurring substances, proteinaceous materials such as animal glue, egg (both yolk and albumen), and casein from milk have long been used for preparation of binders, coatings, and adhesives, used alone, mixed together, or in combination with oils and other organic materials.

Several methods have been proposed in the literature to characterize protein materials in the paintings, but relatively limited in sensitivity and specificity. So far, the identification of specific proteins occurring in binders has been successfully performed using mass spectrometry, providing promising results[5,6].

However, application of mass spectrometry to conservation science still deserves specific considerations arising from the intrinsic nature of the samples themselves that include:

- a minute quantity of available sample;
- a minimally invasive analysis;

- the complex and quite variable chemical composition of the paints themselves, because of the possible simultaneous presence of several binders and contaminants;
- degradation of the original materials affected by aging and pollution.

Therefore, it is essential to develop analytical protocols endowed with minimal invasiveness and extremely high sensitivity.

This doctoral thesis does not aim to highlight how the analytical method developed is more efficient in comparison to the disadvantages of different analytical methods used in the years, but aims specifically to illustrate the logic that led to the choice of sample treatment, the development of efficient methods and data analysis tools to obtain the maximum result and useful information from the experiment.

The remodelling of classical proteomic methods and the application of advanced analytical techniques of mass spectrometry is induced mainly by the choice of steps to be performed in the analysis of different ancient samples each characterized by different problems related to their nature in order to meet different purposes of investigation.

II

Cultural Heritage object and Proteinaceous materials

This inceptive chapter is briefly focuses on the term of Cultural Heritage and the social significance that outcomes from his definition. In order to understand how artworks and archaeological materials can be chemically analysed, it is essential to understand something about their context: sources, technics, technologies and use. Organic materials are more subject to degradation than inorganic ones, so if we can understand their composition, then we are able to ensure that ancient artefacts will remain part of our cultural heritage. In this chapter, a classification of organic materials that can be found as binders is reported, focusing particular attention on proteinaceous constituents and the category of the protein contained in binders.

II.1 What does Cultural Heritage mean?

The word ‘culture’ is sometimes used to refer to the highest intellectual endeavors and the pursuit of perfection and beauty. As the poet and critic Matthew Arnold claimed, culture is 'the best that has been thought and known in the world'[7]. Cultural Heritage is a wide concept defined as an expression of the ways of living developed by a community and passed on from generation to generation, including customs, practices, places, objects, artistic expressions and values. From a general point of view, cultural heritage can be divided into tangible culture (such as artifacts, works of art, buildings, monuments, books and pictures) and intangible culture (such as rural landscapes, coasts and shorelines, language, oral history and knowledge)[8].

As part of human activity, Cultural Heritage produces tangible representations of the value systems, beliefs, traditions and lifestyles. As an essential part of culture as a whole, instead, it contains these visible and tangible from antiquity to the recent past.

Therefore, it is accurate to define cultural heritage as the set of physical artifacts and intangible attributes of a community that is inherited from past generations, maintained in the present and bestowed for the benefit of future generations.

Societies must continue to protect and preserve their cultural heritage, for reasons ranging from education to historical research to the desire to reinforce a sense of identity that in times of war and conflict become all the more important.

Cultural heritage, moreover, contribute to create our history and need to be safeguarded and protected in order to assure that they will be passed out from one generation to another.

II.2 Paintings and organic materials

Paintings are complex materials characterized by intrinsic heterogeneity on a microscale. They generally consists of several layers of 10–200 μm characterized by:

- the support (plaster, canvas, wood, paper, parchment);
- the ground layer or priming (for instance gypsum and animal glue);
- one or more paint layers where the pigment is dispersed into an organic binder to obtain the colored;
- film-forming paint capable of adhering to the ground layer;
- eventually an outer layer of varnish[9].

Ancient paintings were made with a variety of organic materials used as paint, and the empirical attempts to develop the best recipes have led to the acknowledged use of a few natural materials for this purpose. The following table shows the main elements used by artists from the past classified according to the type of constituent molecule.

Table II.1 Category of principal organic materials used in ancient paintings

CATEGORY	ORGANIC MATERIALS
Proteins	Egg, milk, animal glue, silk, wool, vegetable proteins (e.g. garlic), human and animal tissues.
Glycerolipids	Animal fats, vegetable oils (e.g.olive oil) including drying oils (e.g. linseed, walnut, poppy seed).
Waxes	Beeswax, spermaceti, Chinese wax, lanolin (animal waxes); carnauba, candelilla, Japan wax, esparto (vegetable waxes); paraffin, ozokerite (fossil waxes).
Natural resins	Pine resins, sandarac, copals, mastic, dammar, amber, frankincense, benzoe, styrax, myrrh, (plant resins); shellac (animal resin); tar and pitch (from thermal treatment of plant resins or wood).
Polysaccharide	Starch, cellulose, plant gums (arabic gum, tragacanth, karaya, ghatti, guar, locust bean, fruit tree gum).
Bituminous materials	Bitumen, asphalt.
Organic dyes	Cochineal, madder, kermes, saffron, purple, indigo, synthethic dyes.

Synthetic polymers	Polyacrylates, cellulose nitrate, phenolic resins, polyethylene, poly(vinyl acetate), polystyrene.

In Europe, organic materials were constituted essentially by proteinaceous binders, polysaccharides gums, and siccative oils [10]. Other natural organic materials widely used in paintings as plant resins and waxes, mostly serving as varnishes, surface coatings and organic dyestuffs, used as coloring matter [11].

The choice of the binding medium is often specific to the artist's technique but also adapted to the artwork support and to the materials availability. Consequently, the identification of binders provides fundamental information for art techniques and historians but also for restorers and conservators. For example, the use of casein or oil as a binding medium in paintings results in very different painting practices: casein paints are brittle after dying, and thus, they are less suitable for heavy impastos than oil paints especially when used on flexible supports. Furthermore, the material examination can provide information on the trading customs of art supplies in certain regions and has the potential to facilitate the authentication of an object.

II.3 Protein-based materials

Of the proteins containing substances animal glue, egg (both yolk and albumen), milk and casein glue have long been used by artists in the technique known as *tempera*, for preparation of binders, coatings, and adhesives, used alone, mixed together, or in combination with oils and other organic materials [10].

Proteins play an essential role in every living organism and as such, many natural products made from animal or plant sources contain to some extent proteins. They, moreover, can be expected as a major or minor component in many other natural products used in paint, in conservation treatment products or they can be introduced by contamination. Most of the publications on protein binders are restricted to animal glues, eggs and milk based binders, and so these will be the main focus in this chapter.

II.3.1 Milk-based binders

Milk is an aqueous emulsion of proteins and lipids (dried whole cow milk contains around 26% protein, 26% lipids, and 39% sugars [4]). Most milk-based binders are based on the curd or

calcium caseinate. This is the solid fraction of phosphoprotein complex involved in the emulsification of the lipid fraction that precipitates by the acidification of skimmed milk. It was mainly applied as glue and eventually as a tempera medium. Alternatively, a colloidal suspention is obtained by enzymatic or heating treatment of milk, soluble in basic aqueous solutions. Calcium caseinate is a strong viscous glue widely used in restoration work, while ammonium caseinate was used as a binder, mainly in mural paintings [9]. The curd proteins s1-casein, αs2-casein, β-casein and κ-casein are phosphoproteins, and account for 80% of the total proteins in milk.

II.3.2 Egg-based binders

Eggs were a valid alternative to milk tempera for artists in the past. Chicken whole egg and yolk have been the fairly widespread media in tempera paintings, and have also been used as fixatives and unifying in restoration. Paints that are based on yolk and water are very often denominated egg temperas, but its composition varies, sometimes including egg white. The most abundant protein in the white is ovalbumin (ca. 54%), a 45 kDa phosphorylated glycoprotein. Other major proteins are ovotransferrin (formerly conalbumin, ca. 12%, 76 kDa), ovomucoid (ca. 11%, 28 kDa), ovoglobulins G2 and G3 (each ca. 4%, 49 kDa), ovomucin (ca. 3.5%, 5.5-8.3 kDa) and hen egg white lysozyme (ca. 3.4%, 14.4 kDa) [12].

Since egg yolk contains a consistent fraction of lipids (dried whole hen egg contains about 45% protein, 41% lipids, and 2% cholesterol [4]), as it dries it forms a particularly resistant and elastic film, whose permeability and solubility decreases in the course of curing and ageing due to denaturation of globular hydrophilic proteins.

The proteins in the form of complex lipoproteins are responsible for emulsifying the complete lipid fraction, that is either present in triglycerides and phospholipids. The protein fraction of these lipoproteins are in this context called apoproteins. On the basis of its dry matter, yolk has five major constituents: 68% low-density lipoproteins (LDL, lipovitellenins), 16 % high-density lipoproteins (HDL, lipovitellins), 10% globular proteins (livetins), 4% phosphoprotein (phosvitin), and 2% minor proteins[12].

There are six major LDL apoproteins (apovitellenins), all of which are glycosylated. There is a lack of knowledge concerning the exact identification of the apoproteins of LDL. The HDL or lipovitellins are dimeric complexes and have a more molecular form and much lower emulsifying properties: 75-80% proteins versus only 20-25% lipid fraction. Each monomer of HDL is composed of about five glycosylated apoproteins (vitellins), which have their origins in the precursor vitellogenin chains.

II.3.3 Animal glue-based binders

Animal glue was widely used not only as binder for pigments but also mixed with chalk for the preparation of the ground layer in canvas paintings, as adhesive, or for gilding on manuscripts, wood or building embellishment[13,14]. It also found application in e.g. woodworking (typically hide glue), book binding, as a varnish coating, for consolidation of flaking paint. Traditionally rabbit skin glue was used, due to its higher elasticity compared to other hide glue. Because of the hygroscopic nature of collagen, paints based upon animal glue have a relatively high risk of cracking [10].

Animal glue was obtained by boiling tissues such as bones, skins, and cartilaginous parts of animals or fishes. It is made up of collagen, a protein characterized by the presence of a high content of glycine, proline and hydroxyproline [4]. Its characteristics will be described in the following paragraph.

II.3.4 Vegetable proteins-based binders

Plant sources, was occasionally used in paints and contain about 10% (w/w) protein matter. Gliadin and glutenin comprise about 80% of the protein contained in wheat seed. Garlic (Allium sativum), a member of the Liliaceae family and one of the most important vegetable containing proteins, was as an adhesive in gildings [15]. It contains 0.1–0.4% volatile oil, carbohydrates (making up 75% of the dry matter), and proteins (15–17% of the dry matter) [16]. Also plant gums such as arabic gum, which is mainly composed of polysaccharides, contain a minor proteinaceous fraction. High amounts up to 29% of hydroxyproline are present in Arabic gum [17], the most frequently and widespread gum made from Acacia spp. and was used as an adhesive or as a paint or ink medium.

II.3.5 Archeological materials

Archeological heritage means any mould, tumulus, any tombs, place of interment, any cave, any sculptures, inscriptions or an immovable object which is of historical or artistic interest and which has been in existence for not less than hundred years.

Glass, pigments and dyes, mortars and cement, shall include only some of the kinds of raw materials and finished products that can be found on an archaeological site and a variety of archaeological materials may contain trace of organic compounds, including ceramics, stone tools, grinding stones, cooking slabs, plaster, soil, and sediments. Some of the important

categories of materials that can be preserved along with their archaeological context are listed in Table II.2 [18].

All of the kinds of things that people originally used and made, may be present either physically or chemically.

Organic remains such as plant materials, food, skin, and leather items are often the first things found in sites with excellent preservation.

Pigments and dyes were used for painting, decoration, and clothing; glass was originally made into beads, jewelry; mollusk shells from various species were deposited as waste from food consumption, used in construction.

In order to understand something about their presence in archaeological context: sources, technology, use, and disposal; it is essential to chemically analyze the materials found at archaeological sites. A brief parenthesis on archaeological chemistry is provided in order to give a sense of the self-development of the field compared to the art and the restoration chemistry. Archaeological chemistry began almost simultaneously with archaeology itself and has grown dramatically in the last 50 years as the instruments and procedures of chemical analysis become more readily available to scientists outside of chemistry [19]. While archaeologists commonly describe tools functionally in terms of their use (e.g., flake, projectile point, borer, ground stone, etc.), archaeological chemists focus upon the compositional characterization of these materials in order to obtain information from historico-cultural point of view, but also to improve preservation techniques.

The best preservation seems to be in artifacts such as pottery that have absorbed trace of organic compounds into their structural matrix, reducing the accumulation of "contaminants" from *diagenesis*, a post depositional alteration extremely relevant from archaeological excavations and handling, as well as the oxygen-induced degradation that can interfere with the identification of the original parent material. These organic molecules appear to bond to clay and to cling tightly to that bond over long periods of time[18].

Tab. II.2. The type of molecule and archaeological contexts where these residues are found.

Category	Contexts
Proteins	Vegetable (e.g. Garlic) in vessel residues, human and animal tissues, tools;
Glycerolipids	Vessel residues, soils;
DNA	Animal and plants;
Polysaccharides, starches	Vessel residues, soils;
Resins	Gums, resins binders;
Bituminous materials	Bitumen, asphalt;

Moreover ancient human and animal remains are included in this category of cultural heritage too [20].

Similarly important, is the preservation of the bones, the most commonly recovered tissue in archaeological sites. Bones represent a time capsule of ancient biomolecules, owing to their natural resistance to post mortem decay arising from a unique combination of mechanical, structural, and chemical properties.

Bones and other skeletal tissue from animals (teeth, antler, shell, horn) was often used to make certain kinds of tools and equipment. Moreover, they can answer questions about whether animals were scavenged, hunted, or herded, their age and sex, about how animals were butchered, about how important meat was to the diet, when animals died, and the process of domestication. Similar kinds of information are present in human remains. Bones, a combination of mineral and organic matter, is more susceptible to decomposition and contamination. The distinction between cortical and trabecular bone tissue is an important one in understanding the survival of skeletal parts.

An enormous amount of information about both deceased individual is stored in human bones. The length and thickness of long bones provide an indication of an individual's size and strength. Evidence of disease or illness is often embedded in bone. It is often possible to determine age at death, cause of death, sex, history of disease or accident, occupation, movement, and nutritional status from the analysis of prehistoric human bone. Because bone is organic, it sometimes holds ancient DNA that, if it is not subjected to an advanced degradation state, can tell us many things about past activities and genealogical relationships[21].

Bone is in fact a remarkable "space-age" material, a network of mineralized fibers, composed of a matrix of organic collagen filled with inorganic calcium phosphate (apatite) crystals. Bone has three major components: a mineral fraction (apatite), an organic matrix (collagen), and water. By dry weight, the organic materials constitute about 30% and minerals about 70% of bone.

Collagen is a major structural protein, forming molecular cables that strengthen the tendons and vast, resilient sheets that support the skin and internal organs. It represents the most abundant protein in mammals, making up from 25% to 35% of the whole-body protein content. It is an extracellular protein characterized by three parallel polypeptide chains organized into insoluble fibers of great tensile strength. This suits collagen to its role as the major stress-bearing component of bone, teeth, cartilage, tendon, ligament, and the fibrous matrices of skin and blood vessels. Vertebrates have 46 genetically distinct polypeptide chains comprising 28 distinct collagen types that occur in different tissues of the same individual. The most prominent of these is Collagen Type I. A single molecule of Type I collagen is composed of three polypeptide chains with an aggregate molecular mass of 285 kDa. The amino acid sequence of bovine collagen 1(I), which is similar to that of other collagens, consists of monotonously repeating triplets of sequence Gly-X-Y over a continuous 1011-residue stretch of its 1042-residue polypeptide chain. About repeating triplets, X is often Proline and Y is often 4-hydroxyprolyl (Hyp), which confers stability on collagen, possibly through intramolecular hydrogen bonds that involve bridging water molecules[10].

From the analysis of human and animal bones, it is possible to understand the phylogenetic relationship between different species; furthermore, it is possible to carry out anthropological and environmental studies by determining post-translational modifications on specific amino acids in collagen proteins. Besides, the ability to discriminate between bone fragments from different species is not important only in archeology and paleontology, in fact it has many applications in the realms of modern life too, in areas such as food safety, fishing industry, forensics and trading standards.

II.4 Contamination as other sources of proteins

Other sources of proteins in paints are introduced by contamination, such as skin, or hair fragments, saliva and bacterial or fungal activity. Keratins are major proteins in both the epidermis and hair; as such, avoiding contamination of paints with keratin is nearly impossible. Saliva was often used as a "ready-to-use" cleaning agent or used to wet the paint-

brush. It contains a small amount of digestive and antimicrobial enzymes. Finally, in non-ideal storage circumstances, a high humidity for example, microbe or fungal activity might introduce proteins into the paint layers.

III

Proteomics and Mass Spectrometry in Cultural Heritage

For a suitable knowledge of the protein-based binders in artworks and archaeological materials, it is essential to identify the individual proteinaceous additives and to distinguish them even in materials where they are present in very small amounts, in insoluble forms and often in matrices unsuitable from an analytical point of view. To meet this aim, methodologies characterized by high selectivity, specificity and sensitivity are required.

A proteomics approach, typically considered to be associated with the study of living organisms, has been proven to be an effective tool also for the scientific analysis of Cultural Heritage materials and and mass spectrometry forms a fundamental platform for this new methodology.

III.1 Proteomics approach

Proteomics is the science dedicated to the study of proteins, including their identification, quantification, and the study of their modification (chemical and post-translational)[22,23]. The word "proteome" was proposed for the first time in 1995 to designate a protein set expressed by the genome of a cell, a tissue, and an organism at a precise moment of its development and in a precise environment[24].

Whereas the genome of an organism is essentially invariant and static in all its cells,

proteomes vary from cell to cell, with time and as a function under developmental, physiological, pathological, pharmacological and aging conditions. Proteomics allows a dynamic description of gene expression and it has deep roots in the early 1970s, as a result of important technical advances in biochemistry and genetics.

Nowadays proteomics can rely on powerful analytical protein-separation technologies (chromatography and electrophoresis), that serve to simplify complex protein mixtures [25,26]. However the most powerful analytical tools for proteomic analysis is mass spectrometry (MS), whose instrumentation has undergone huge changes over the past years, culminating in the development of highly sensitive, robust instruments that can reliably analyse biomolecules, particularly proteins and peptides. Moreover, the application of bioinformatics

tools, that integrates protein databases, allowing the identification of proteins and their modification using MS data.

III.2 Mass Spectrometry

MS is one of the most powerful methodologies for identifying, structurally characterizing, and quantitating wide classes of molecules, ranging from small to very big species. It plays a key role in proteomics studies because of its capability of identifying proteins and characterizing post-translational modifications.

Most important features of mass spectrometry are reproducibility, sensitivity, resolution and accuracy. In particular, mass spectrometry is a reproducible methodology based on the determination of the molecular mass, which is not dependent on the experimental conditions.

III.2.1 Principles and instrumentations

All approaches that use MS are based on the study of ions in the gas phase [27]. These ions can interact with an electric field and can be resolved following their electro-dynamic attitude, which is dependent on their mass-to-charge ratio

In every mass spectrometer, independently from its features and performance, ions are generated, accelerated, driven, analyzed and detected. In principle, all molecular and atomic ions are accessible to mass spectrometric analysis; for this reason, mass spectrometry is a very powerful tool for chemical and biochemical analysis in general.

Today a wide variety of mass spectrometers are available, all sharing the capability to assign mass-to-charge values to ions, although the principles of operation and the types of experiments that can be done on these instruments differ greatly.

In every MS experiment the first step is the introduction of the sample into the mass spectrometer. It follows that the first part of every instrument is the inlet system that allows the introduction of the sample, generally molecules, into the mass spectrometer. There are different ways to introduce the sample, depending on its purity and properties. Once molecules have been introduced inside the mass spectrometer, they must be ionized, i.e. transformed into ions. The ionization process occurs in the ion source by using an ionization technique. There are different ionization techniques depending on the physico-chemical properties of the molecules. The ions thus produced are accelerated and driven from the source to the analyzer that separates them according to their *m/z* ratios. Finally, the detector reveals ions. All mass spectrometers have different stages of pumping in order to maintain the analyzer and detector regions under high vacuum, i.e. 10^{-7}–10^{-8} torr or higher, in order to

prevent collisions of ions with residual gas molecules in the analyser during the flight from the ion source to the detector. Depending upon the ionization technique, the inlet system and the ion source must be or not be under vacuum. Mass spectrometers have four essential parts: a system for sample introduction; a source that produces gas phase ions from the sample; one or more mass analyser to separate ions; an ion detector (Figure III.1.).

Initially limited to the analysis of volatile and low molecular-weight components, the developments in MS were based on the application of new ion sources able to transfer organic macromolecules (such as peptides, proteins, nucleic acids, lipids and so on) from liquid or solid state to gaseous phase. These problems were solved with the development of new ionization methods and in particular with the introduction of Matrix Assisted Laser Desorption Ionization (MALDI) [28] and Electrospray Ionization (ESI) [29] techniques, that are the most common ionization techniques used for biochemical applications. These ionization methods are considered "soft", because the generated ions undergo poor fragmentation.

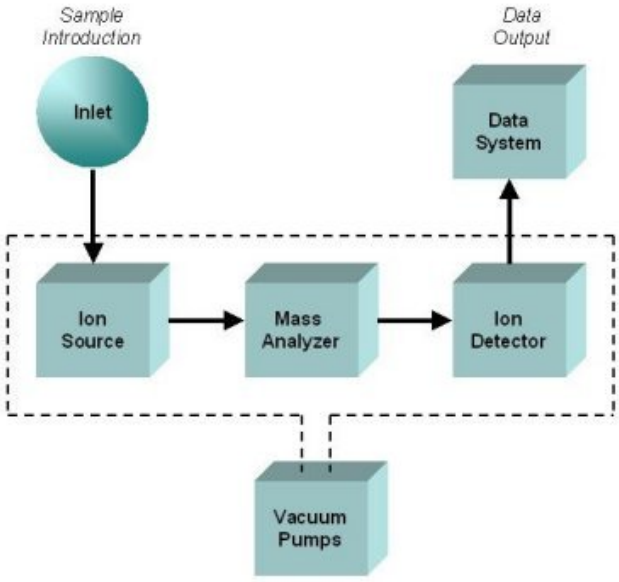


Figure III.1. A schematic diagram of a mass spectrometer.

Matrix Assisted Laser Desorption Ionization. During the development of MS, a lot of studies have been devoted to the use of laser light as an energy source for ionizing molecules. As a result, MALDI, developed by Karas & Hillenkamp in the mid-1980s, was introduced and soon applied to the study of large molecules [28].

To generate gas phase, protonated molecules, a large excess of matrix material is co-crystallized with analyte molecules. The resulting solid is then irradiated by nanosecond laser pulses, usually from small nitrogen lasers with a wavelength of 337 nm. The matrix is typically a small organic molecule with absorbance at the wavelength of the laser employed and the matrices differ from each other in the amount of energy they impart to the biomolecules during desorption and ionization and hence the degree of fragmentation (unimolecular decay) that they cause. Generation of ions is believed to arise through ion/molecule reactions in the gas phase. Generally, $[M+H]^+$ ions are preferentially formed in the positive ion mode, and $[M-H]^-$ ions in the negative ion mode. The presence of singly charged signals in MALDI spectra allows to correlate each peak one-to-one with analytes, thus allowing the analysis of complex mixtures.

Electrospray ionization. Although the electrospray phenomenon was known since the beginning of the twentieth century, the breakthrough for ESI came in the mid-1980s by Fenn et al. [29] The term electrospray reveals important features of the process: first of all, and differently from the ionization methods described so far, the ionization occurs on a spray. This is produced by forcing a liquid containing the analyte to pass through a capillary, at low microliter-per-minute flow rates through a fine needle at high voltage. The first part of the ionization's term contains further important information: an electric field (3–5 kV) is applied to the capillary and it causes ionization. The overall effect is the formation of small liquid-charged droplets from which gas phase ions are formed. ESI is governed by a large number of chemical and physical parameters that together determine the quality of the process.

The charged droplets decrease in volume by solvent evaporation, assisted by a warm flow of nitrogen, known as desolvatation gas, which flows perpendicularly to the front of the ionization source. Once the droplets have reached the Rayleigh limit, ions are desorbed from the droplet generating gas-phase in a "continuous beam" fashion.

This technique gives rise to multiply-charged molecular-related ions such as $(M+nH)^{n+}$ in positive ionization mode and $(M-nH)^{n-}$ in negative ionization mode; the number of charges depends on the chemical-physical characteristics of analytes; thus each analyte could assume more than one charge complicating the form assumed by the signal. For this reason this source is not useful to analyse complex mixtures except when combined to the use of using pre-fractionation systems. In fact most uses of ESI-MS are made in combination with liquid chromatographic separation, the so called LC-MS technique.

A great enhancement of ESI ion sources has come from the reduction of the flow rate of the liquid used to create the spray to a nano-scale level. This device leads to a higher efficiency in creating ions [30] because the charge density at the Rayleigh limit increases significantly with decreasing droplet size. Others advantages related to the use on μLC and nLC are the low consumption of sample and the higher sensitivity of the methods because of the increase in the concentration of the analyte as it elutes off the column.

Once the ions are formed in the ion source, they are accelerated towards the mass analyzer where separation according to their *m/z* ratio occurs. It is, literally and figuratively, central to the technology.

Before to specify the main analyzers, it is better to consider some feature common to all of them and most important in the context of proteomics, its key parameters are sensitivity, resolution, mass accuracy and the ability to generate information-rich ion mass spectra from peptide fragments (tandem mass or MS/MS spectra) [31,32].

There are four basic types of mass analyser currently used in proteomics research (Fig.III.2). These are the quadrupole (Q), ion trap (IT), time-of-flight (TOF), and Fourier transform ion cyclotron resonance (FT-ICR) analysers [33]. They are very different in design and performance, but also for their resolution, sensitivity and accuracy and for the types of experiments that are able to perform. These analysers can be stand alone or, in some cases, put together in tandem to take advantage of the strengths of each.

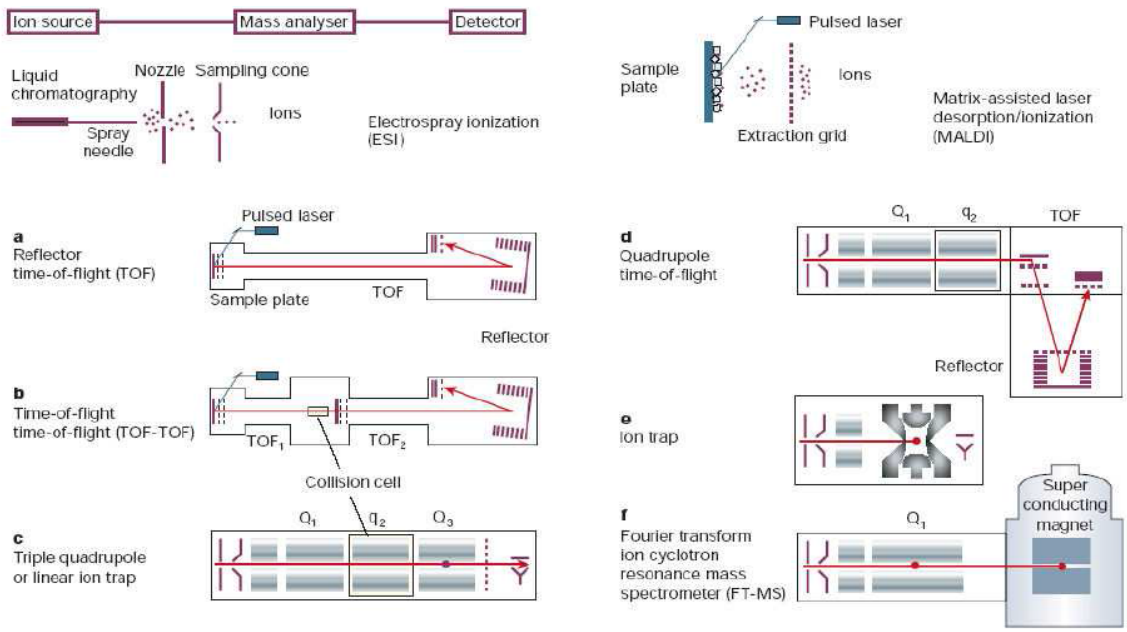


Figure III.2. Typical mass spectrometers used for proteomics analysis.

TOF analysers are typically used in combination with MALDI (MALDI-TOF MS instruments); however, both MALDI sources and TOF analysers can be used in different configurations. TOF analyser working under high vacuum (10^{-8}-10^{-9} atm) and ions can be separated during their flight in a field-free drift region basing upon the time they use to pass through the tube. Their different velocities depend on their *m/z* ratios. All ions entering the TOF tube have a fixed kinetic energy, which is proportional to the applied accelerating voltage and the charge. This implicates that the higher is the mass of the ion the lower is its velocity and the longer it takes before the ion reaches the detector. A detector at the end of the tube records the time of flight of the ions.

MALDI ionization process causes a spread of kinetic energy of ions resulting in different points, in time and space, of ions formation within the ion source. Thereby ions with the same mass obtain different kinetic energies and velocities during their extraction out of the ion source. This results in peak broadening, causing a loss in resolution. This peak broadening has been reduced by the introduction of the 'delayed extraction' and of a 'reflectron' at the end of the linear flight tube. The reflectron allows ions to slow down and reverse their flight path to a second detector. Ions with lower kinetic energy do not penetrate the reflectron as deep and thus turn around faster, catching up with ions of slightly greater kinetic energy that penetrate the reflectron deeper. Thereby the flight times of ions with identical *m/z* values, but different kinetic energy values, will be corrected when the ions arrive to the detector. Commercial TOF based instruments have high resolution of about 10,000 or greater and high accuracy characteristics.

The quadrupole mass filter (Q) is composed of four parallel cylindrical, or ideally hyperbolic, rods to which a direct current (DC) and a radiofrequency (RF) are applied. The separation of ions according to their *m/z* ratio occurs by scanning simultaneously DC and RF, but maintaining their ratio to be constant. For a given value of DC and RF only ions with a given *m/z* value will be able to pass through the entire quadrupole and they will be transmitted to the detector. By changing DC and RF another family of resonant ions will be transmitted and so on. A complete scan allows all the *m/z* values in a given range to be analysed in different instants.

Commercially available instruments usually have mass/charge limits ranging from 0 to 4000 *m/z* and at best are normally set to resolve the various ^{13}C isotope peaks for a singly charged ion, although the resolution may be intentionally degraded to improve sensitivity. In ESI, multiple charging enables quadrupole mass measurement of molecules >100,000 Da, if the molecule can be charged sufficiently.

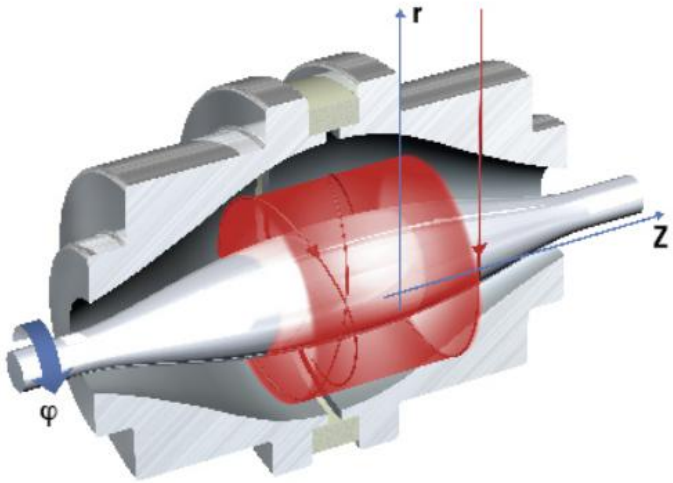
The quadrupole ion trap (QIT) is based on the same principle as the quadrupole mass filter, except that the quadrupole field is generated within a three-dimensional trap. The trap consist of three electrodes, a rotationally symmetrical ring electrode of hyperbolic shape and two hemispherical end cap electrodes of the same cross-section. One end cap electrode usually has a hole in it to allow ions to enter the trap. Ions are injected axially and to overcome the exit of the ions, their high kinetic energy is dissipated by cooling them collisionally, filling the trap with gas molecules [34]. Ions are dynamically stored in a three-dimensional quadrupole ion storage device by an electric quadrupole field generated by the applied RF voltage. The RF voltage can be scanned to eject successive mass-to-charge ratios from the trap to the detector. The advantages of the ion-trap mass spectrometer include compact size and the ability to trap and accumulate ions (typically the ion trap can hold up about 105–106 ions before coulombic repulsions significantly affect their trajectories and reduce the mass resolution) to increase the signal-to-noise ratio of a measurement. Mass spectrometer equipped with QIT technology has good sensitivity but low resolution and mass accuracy. Due to its small trapping volume, QIT has a limited trap capacity; when the number of ions in the trap is too high, we assist to deterioration in the mass spectrum and loss of dynamic range due to space-charging interactions. To avoid these effects, the number of ions introduced into the trap has to be strictly controlled [34]. Trapping of ions can also be performed in linear ion trap devices (LIT) [35]. LITs have two major advantages over QIT: a larger ion storage capacity and a higher trapping efficiency. In LIT devices analysis is performed by ejecting the ions radially, through the slits of the quadrupole rods, using the mass instability mode as a QIT. Detection is performed by two detectors placed axially along to the rods [36].

In FT-ICR and the Orbitrap analyzers the *m/z* values of the ions are not directly measured, but they are obtained by Fourier transform treatment of the signal.

The Orbitrap analyzer,[37] invented by Alexander Makarov, has been defined by the company that commercially produces it as 'the first totally new mass analyzer to be introduced to the market in more than 20 years'. Its name recalls the concept of trapping ions. Indeed, ions are trapped in an electrostatic field produced by two electrodes: a central spindle-shaped and an outer barrel-like electrode. Ions are moving in harmonic, complex spiral-like movements around the central electrode while shuttling back and forth over its long axis in harmonic motion with frequencies

dependent only on their *m/z* values (Figure III.3). The latter can be determined by Fourier transform treatment of the signal. Note that, in contrast to other trapping analyzers, neither RF

nor a magnet field are used. Moreover, the Orbitrap allows very high resolution to be achieved (the resolving power in commercial instruments is 100 000, rivalling that of FT-ICR instruments) and routine mass measurement accuracies less than 2 ppm.



$$U(r,z) = \frac{k}{2} \cdot \left\{ z^2 - r^2/2 + R_m^2 \cdot \ln(r/R_m) \right\}$$

Figure III.3. Orbitrap mass analyser.

The FT-ICR analyzer is based on the interaction of a charged species with an external magnetic field produced by superconducting magnets. A charged species subject to a magnetic field covers a circular trajectory.

The rotational frequency of a certain ion on this orbit is called the 'natural cyclotronic frequency' and it is dependent on the magnetic field and on the *m/z* ratio of the ion.

An additional RF is applied and the net result is a spiral motion. To obtain the mass spectrum the RF is scanned while the magnetic field is maintained constant. In FT-ICR mass spectrometers, ion isolation and detection occur in the same region. In fact all ions coming from the source are simultaneously excited applying a RF pulse of a large wide band. The resulting signal is then converted by Fourier transform to obtain the mass spectrum. The main feature of this analyzer is its extremely high resolution [38]. However, owing to the difficulty of operation and its very high cost, the FT-ICR analyzer cannot be used for routine purposes.

Once ions have been produced and analysed they must be detected. Indeed, the detector is the final part of a mass spectrometer. At the very sunrise of MS, detectors were composed of a fluorescent screen or a photographic plate; the modern instruments are equipped with detectors able to transform the signal produced by the ion beam into an electric current that is

transmitted to the data system. At the moment, there is a variety of different types of detectors whose choice is determined by the instrument design and applications[39].

III.2.2 Tandem mass spectrometry

Tandem MS [40] involves two or more analyzers which are synchronized. This is particularly important to obtain structural information and to produce quantitative analysis by fragmenting specific analytes inside the mass spectrometer and analysing the fragment ions. Tandem mass spectrometry also enables specific compounds to be detected in complex mixtures on account of their specific and characteristic 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, using two different mass analysers; the first mass analyser is used to selectively pass an ion into another reaction region while the second mass analyzer is used to record the *m/z* values of the dissociation products.

The collision cell is the region into which an inert gas, e.g. nitrogen, argon or helium is admitted to collide with the selected ions causing the fragmentation. This process is known as Collision Induced Dissociation (CID) or Collision Activated Dissociation (CAD) [40,41]. The collisions between the precursor ion with a high translational energy and a neutral target gas cause the conversion of part of the translational energy into internal energy of the ion, leading to subsequent fragmentation. The CID process is highly dependent on the relative masses of the two species. The overall CID process is expected to occur by a two-step mechanism, where the excitation of the precursors and their fragmentations are separated in time. Generally, fragmentation of a precursor ion can occur if the collision energy is high enough to allow the ion to be excited beyond its threshold of dissociation. 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 QqLIT and QqTOF, trapping devices, such as ion traps (IT), and Fourier-transform ion cyclotron resonance (FTICR) instruments.

The observed fragmentation pattern depends on various parameters including the amino acid composition and size of the peptide, excitation method, time scale of the instrument, the charge state of the ion, etc [33]. Peptide precursor ions, dissociated by the most usual low-energy collision conditions, fragment along the backbone at the amide bonds [42,43] forming structurally informative sequence ions and less useful non-sequence ions by losing small neutrals like water, ammonia, etc.

The amino acid backbone has three different types of bonds (NH-CαH, CαH-CO and CO-NH) and each of them can be fragmented originating different fragment ions. Each bond breakage gives rise to two species, one neutral and the other one charged, and only the charged species is monitored by the mass spectrometer. Hence there are six possible fragment ions for each amino acid residue and these are labelled as in the Figure III.3, with the *a*, *b*, and *c* ions having the charge retained on the N-terminal fragment, and the *x*, *y* and *z* ions having the charge on the C-terminal fragment. The most common cleavage sites are at the CONH bonds, which give rise to the *b* and/or the *y* ions [44]. The most comprehensive model currently available to describe how protonated peptides fragment is the so called *mobile proton* model [52]. According to this model peptides activated under low-energy CID, fragment mainly by charge directed reaction. In fact, protonation of the amide nitrogen along the amino acid backbone leads to considerable weakening of the amide bond and leads to cleavage, generating fragment ions. Considering, for example, a doubly protonated tryptic peptide, one proton will be localized on C terminal Arg or Lys side chain and the second may be localized at one of the amide bonds or the N-terminal. In this case, the *heterogeneous population model* assumes that, when the second proton is mobilized along the amino acid backbone, it will exist different protonated forms of the precursor ion that can easily fragment giving complementary series.

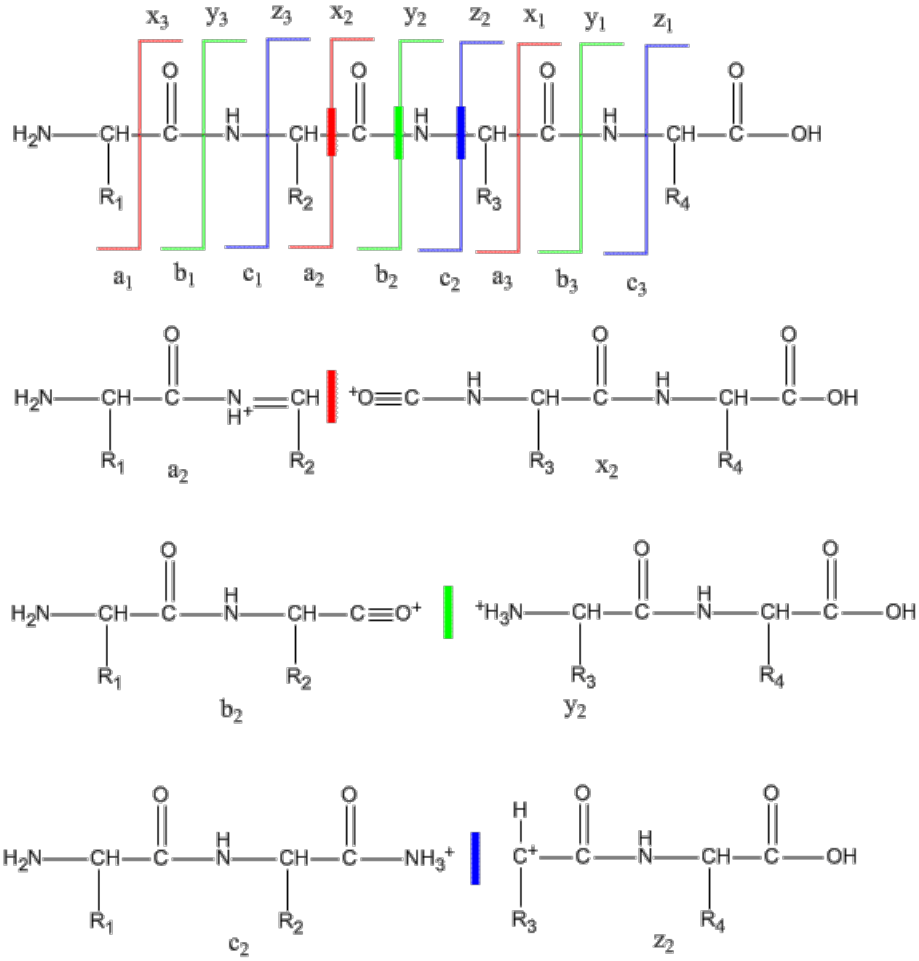


Figure.III.3. Peptide fragmentation scheme.

III.2.3 Multiple Reaction Monitoring (MRM)

The most used method that require a specific experimental set-up for is the multiple reaction monitoring (MRM). By considering that it is characterize by highly sensitivity, it is possible to carry out also for detection and quantitative determination of specific compounds at trace level[4546].

In MRM the first and the third quadrupoles act as filters to specifically select predefined *m/z* values corresponding to the peptide ion and different specific fragments ion of the peptide, whereas the second quadrupole serves as collision cell (Figure III.4)[47].

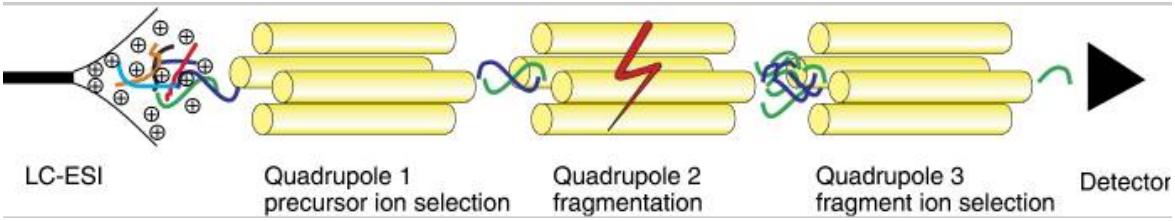


Figure III.4: MRM analysis on QQQ MS. Several analytes are coeluting from the chromatographic system. The specific *m/z* selection in the first quadrupole filters out most coeluting ions. However, owing to identical mass, one interfering ion (blue) remains. In quadrupole 2, the analytes are fragmented. The *m/z* selection in the third quadrupole filters out all the fragments of the blue analyte and leaves only a set of fragments of the green analyte for specific detection.

Despite the high specificity of MRM analysis, obtained by the two filters of mass of the QQQ, a particular transition (precursor/product) may not be specific to a molecule. In fact, in a complex mixture, non-specific signals can be derived from other molecules with transitions of similar masses, thus affecting the validity of the analysis. The acquisition of two or more parallel transitions to the same molecule (precursor → product1, product2) gives the analysis further selectivity. In fact, the typical elution time of the analyte in question, these transitions will produce a set of peaks "co-eluted" only if derived from the same molecule. Therefore, the acquisition of multiple transitions in parallel is most important in the MRM method set up. The total number of transitions per stroke LC-MS/MS is limited; commonly are selected for quantitative analyzes, only the best 2-4 transitions for each molecule.

III.2.4 Quadrupole-time of flight (QqTof) mass spectrometer

QqTOF tandem mass spectrometer can be described as a triple quadrupole with the last quadrupole section replaced by a TOF analyser. It have rapidly been embraced as powerful and robust instruments with unique capabilities. QqTOF combines the high performance of TOF devices, in both the MS and MS/MS modes, with the widely used techniques of electrospray ionization. These mass spectrometers are characterized by high sensitivity, high mass accuracy and high mass resolution for both precursor and product ions [48].

The popularity of the QqTOF has been significantly advanced by the rapid growth of semi-automated instrument control and data processing and by continuing improvements in the core performance characteristics of mass resolution and sensitivity. The instrument (Fig 2.4) consists of three quadrupoles, Q0, Q1 and q2, followed by a reflecting TOF mass analyzer with orthogonal injection of ions.

The resulting spectra benefit from the high resolution and mass accuracy of the TOF instruments, and also from their ability to record all ions in parallel, without scanning. Both

Q0 and q2 are operated in the RF only mode: the RF field creates a potential that provides radial confinement of the precursor and/or fragment ions. Since the RF quadrupoles are normally operated at a pressure of several millitorr, they provide both radial and axial collisional damping of ion motion. Ions are thermalized in collisions with neutral gas molecules, reducing both the energy spread and the beam diameter and resulting in better transmission into and through both the Q [49] and TOF [50] analysers. For MS/MS, Q1 is operated in the mass filter mode to transmit only the parent ion of interest. The ion is then accelerated to an energy of between 20 and 200 eV before it enters the collision cell q2, where it undergoes collision induced dissociation (CID) with neutral gas molecules. The resulting fragment ions are focused by RF fields and re-accelerated to the required energy (usually several tens of eV per unit charge), and focused by ion optics into a parallel beam that continuously enters the ion modulator of the TOF analyzer. Initially the modulator region is field-free, so ions continue to move in their original direction in the gap. A pulsed electric field is applied at a frequency of several kHz across the modulator gap, pushing ions in a direction orthogonal to their original trajectory into the accelerating column, where they acquire their final energy of several keV per charge. From the accelerating column, ions arrive in the field-free drift space, where TOF mass separation occurs. The ratio of velocities (or energies) in the two orthogonal directions is selected such that ions reach the ion mirror and then the TOF detector naturally, without requiring an additional deflection in the drift region, which could affect the mass resolution [51].

The ESI Q-TOF mass spectrometer generally can also be equipped with nano-HPLC-CHIP, a different and innovative technology compared to the classic LC-MS systems. In classical LC/MS instruments a precolumn (o enrichment column) where the samples is washed from salts and detergents and concentrated composes the liquid chromatography. After the reverse-phase separation, another post-column capillary transport the analyte to the electrospray. The passage through all the column and the volume of elution in the capillary volume involve a broadening of chromatographic peaks and the decrease of resolution that affect the sensitivity of more dilute species.

The HPLC-Chip is made of sandwiched polyimide films. In particular, the laser ablation is used to generate grooves of specific dimensions which can be filled with packing material, and used as pre-column, reverse phase separation nanocolumn and nanospray tip. In this way diffusion in dead volumes cannot affect the broadening of peaks and the sensitivity is strongly enhanced. On end of the chip tapes off into a polyimide nanospray emitter (tip). A stator and a

rotor fix the chip when it is placed into the HPLC-Chip Cube and the rotation of the latter allows the switch between loading and analysis (Figure III.5).

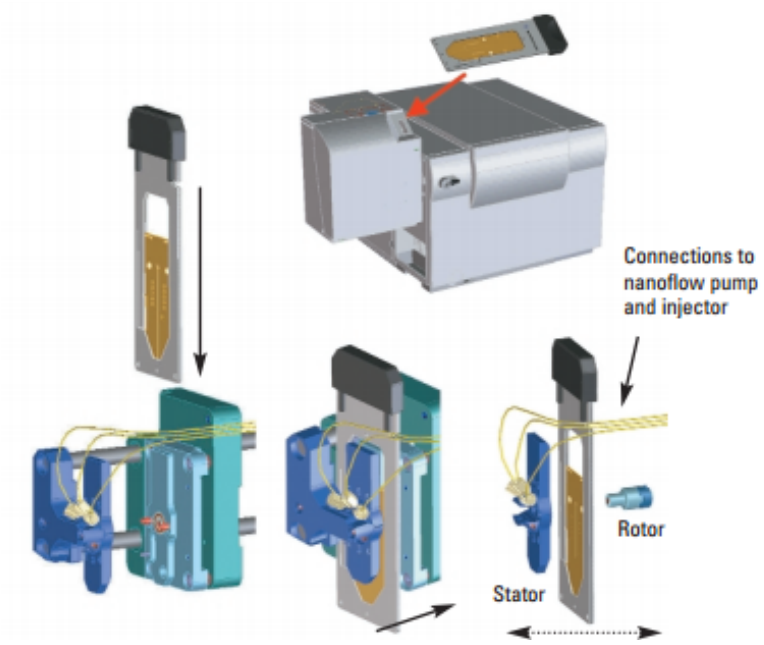


Figure III.5. The HPLC-Chip Cube automatically loads the chip, establishes leak tight connections and positions the chip orthogonal to the MS inlet simply by clicking on the operate command in the ChemStation menu.
(http://www.chem.agilent.com/cag/EMEA/LCMS_Literature/Brochures/5989-5492EN.pdf)

During loading, the sample is transferred at a flow rate of 4 µl/min onto the enrichment column. Peptides are retained at the front end of the column while the loading solvent is directed to waste.

During analysis the outlet end of the enrichment column is connected to the analytical column; as the nanoflow gradient starts (400 nl/min), the retained peptides are eluted onto the analytical column where they are further separated by reversed phase chromatography.

The peptides emerging from the tip enter the spray chamber as an aerosol thanks to a nebulizing gas. Droplets are led by an electric field as described before for an ESI source.

III.3 Proteomics and Mass Spectrometry – based approach for Cultural Heritage

Proteomics approach applied to the identification of protein-based materials in Cultural Heritage samples was mentioned for the first time in the 2000s[285, 492–495]. On the bases of mass spectrometric analysis, and database searching, a classical proteomic strategy based on bottom-up approach was generally applied to protein identification in these kind of objects[52].

Bottom-up is the current proteomics mainstream that is based on the analysis of the peptide mixture resulting from protein hydrolysis (commonly with trypsin, an endoprotease that catalysed the cleavage of proteins at the carboxyl side of Lys or Arg amino acids, except when they are followed by Pro) either by Peptide Mass Fingerprint (PMF) or by peptide sequencing using Tandem Mass Spectrometry (MS/MS).

In the PMF, protein identification depends on the accurate measurement of peptides molecular mass, whereas in MS/MS, peptides are fragmented in the source or whitin the mass spectrometer and amino acid sequences are obtained by the masses of fragments.

The proteomic approach is composed of several steps including the hydrolysis of the proteins, the MS-based analysis of the resulting peptides, and data processing using bioinformatics tools.

Studies of artwork materials, where proteomic analysis is complementary to other techniques such as optical microscopy, SEM with energy dispersive X-ray spectroscopy, micro-Raman, GC−MS, pyrolysis GC-MS, ELISA, SYPRO Ruby staining [53–56].

Focusing on proteins, tandem mass spectrometry (MS/MS) allows a deep structural approach with the accurate identification of the peptide backbone sequence and its modifications. Thus, protein identification can be accurately performed on the basis of a few peptidic sequences.

MS/MS allows the characterization of the amino acid sequences of peptides, providing thus an accurate identification of proteins. Additional information can also be obtained such as the biological origin of the proteins or information on sample degradation state or aging.

The common strategy to interpreter tandem mass spectrometry data is matching the observed spectra with predicted mass values rather than the actual determination of peptide sequence. These peptides are predicted by theoretical tryptic cleavage of known protein sequences, as found in protein databases such as UniProt (http://www.uniprot.org/) or National Centre for Biotechnology Information (NCBInr: http://www.ncbi.nlm.nih.gov/protein/) consulted by dedicated search programs such as Mascot MS/MS Ions Search (Matrix Science), SEQUEST (Thermo Fisher) and X!Tandem (The Global Proteome Machine Organization, GPM).

Moreover, besides the accurate identification of the protein material, tandem mass spectrometry and bioinformatics tools allow also the identification of protein modifications that can be extremely useful in the search of the molecular effects of aging [57–60].

In this Ph.D. project, an updated Mascot version was used. Mascot format requires the insertion of experimental parameters that are necessary to obtain reliable results. Such parameters are: which protease has been used; the maximum number of missed cleavages and which fixed and/or variable post-translational modifications have to be taken into account. An

indication of the instrument type that has been used, since this influences the type of fragmentation and therefore the software should take into account when calculating the match score. A sequence database should be chosen, with eventual taxonomic restrictions. Moreover, mass tolerances, both for peptides and for fragments, and peptide charges to be considered.

A specific algorithm used by Mascot software calculates all possible peptides for all protein sequences available in the selected sequence database. The virtual masses are compared with the masses observed in single MS mode and with fragments in MS/MS mode. For each of the observed components in the measurement, a list of possible peptide annotations is thus composed, based solely on the precursor mass, charge, and ion fragments.

A similarity score MOWSE score (MOlecular Weight SEarch score) [61] is attributed to each of the possible peptide annotations based on the tandem MS spectrum. It uses a statistical model that calculates the probability of a number of matching peaks by random chance by assigning a probability value to each matched *m/z* peak using a training set of sequences and multiplying all probability values to compute the composite probability P expressed as -10 log P. When this MOWSE score is above a limit value, the corresponding peptide annotation is considered as a "hit". Note that MOWSE and Mascot's derived ion scores are not real probabilities, since there is no requirement that a higher score for one sequence reduces the score for other sequences (http://www.proteomics2.com/?p=65). Furthermore, Mascot software regroups all hits per protein and on the base of individual MOWSE scores of the individual peptides, a protein score is calculated (http://www.matrixscience.com/help.html).

All methods reported above belong to the so-called discovery proteomics approaches, where all the peptides brought to the mass spectrometer are fragmented and potentially identified. Recently, targeted proteomic approaches, based on MRM and experimentally observable peptides uniquely identifying a specific protein [62], have gained popularity for their peculiar feature to provide a reliable quantitative evaluation of low-abundance proteins and PTMs [63]. MRM experiments can follow multiple transitions characteristic for the investigated component, thus concentrating only on the analysis of the species of interest within a complex peptide mixture. The high specificity and sensitivity of MRM have made this MS technique highly suitable for the analysis of protein in complex matrixes as in cultural heritage objects and for the detection of PTMs [64]. By adding a stable isotope-labeled version of the peptide of interest, absolute quantization can also be achieved [65].

In contrast to conventional shotgun proteomic studies, the MRM measurements are strictly targeted to predetermined set of peptides and depend on specific MRM transitions for each

targeted peptide. Previous information is required to define these transitions. Specifically, three types of information are of critical importance:

- First, the proteins that constitute the targeted protein set have to be selected;
- Second, for each targeted protein, those peptides that present good MS responses and uniquely identify the targeted protein, or a specific isoform thereof, have to be identified. Such peptides have been termed as proteotypic peptides (PTPs)[66];
- Third, for each PTP, those fragment ions that provide optimal signal intensity and discriminate the targeted peptide from other species present in the sample have to be identified.

These optimized transitions are the essence of this kind of assay. The time and effort required to establish these conditions is the price to pay for the excellent performance experiments.

Unlike in other MS-based proteomic techniques, no full mass spectra are recorded in QQQ-based MRM analysis. The non-scanning nature of this mode of operation translates into an increased sensitivity by five orders of magnitude compared with conventional 'full scan' techniques. This enables the detection of low-abundance proteins in highly complex mixtures, which is crucial for systematic quantitative studies[67].

III.3.1 Protein modifications in ancient samples

Post-translational modifications (PTM) of proteins is part of what makes proteomics so much challenging. They are chemical or enzymatic modifications of amino acids after protein synthesis that change their molecular weights, the fundamental physical property measured by mass spectrometry.

The information that can be obtained in the study of chemical modifications of proteins are very useful to investigate several questions such as the conservation state of a material, the effect of aging, or the effect of environment (such as pollution, humidity, effect of restoration procedures, and so on…).

The most frequently observed PTM in this study is 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. Phosphorylation of serine and threonine (+80 Da) are commonly observed in milk proteins. In case of animal glue, a high number of hydroxyproline and hydroxylysine modifications were observed in fibrillar collagens, and most importantly, more confident identifications are commonly obtained for these proteins by taking into consideration these extensive post-translational modifications when collagen is present.

More 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 post-synthetic modification that plays an important role in protein degradation[68]. It occurs in Asn and Gln amino acids resulting, respectively, in aspartic/isoaspartic acid and glutamic/γ-glutamic acid, and it represents a potential aging marker[69]. Conversions of Asn and Gln to aspartic and glutamic acids can be detected by MS techniques and MS/MS methods, although the modification results in a mass shift of 0.984 Da.

Deamidations were observed in ancient materials, [21,70,71] but it is only recently that studies were focused on the relationship between the deamidation rate and mechanisms such as degradation or aging of a Cultural Heritage sample.

Deamidation was deeply investigated in the study of a deteriorated XIVth century wall painting containing casein and animal glue binders from Camposanto Monumentale of Pisa, Italy[72]. The procedure showed that most of the Gln and Asn containing peptides were deamidated.

III.4 Aim of the PhD thesis

Proteomics multidisciplinary nature has recently led to the application of analytical methods to oddly assorted areas ranging from forensics, food analysis, clinical medicine and; it has been proven to be an effective tool also for the scientific analysis of cultural heritage materials [1].

Suitably adapted, these strategies can be applied to art, archeological, and paleontological samples, exploiting all the advantages and the power offered by proteomics.

In this perspective, my PhD project was devoted to the development of proteomic approaches, which provided the optimization of specific protocols for sample preparation, the application of mass spectrometry methodologies and data analysis tools that can cope with ancient or damaged samples.

In addition, the specific application of the developed strategies to different samples from works of art and archaeological finds was an essential part of the thesis project that aims to validate and demonstrate the efficiency and versatility of the developed methodological approach on real case studies.

C**hapter IV** describes the development and the adaptation of classical proteomic strategies to different kinds of samples containing proteinaceous materials. The aim is to show how the different analytical needs can be satisfied in order to obtain desirable results in the cultural heritage field. The use of specific samples pre-treatments and data analysis tools are important since it improves the quality of the experimental data that allow to highlight specific aspects of both identification and characterization, and consequently it helps extensive studies such as the searching for unpredicted modifications induced by aging processes. In this respect, classical protocols might be adapted to the specific requirements of these particular samples, to obtain a high number of peptides and consequently higher sequence coverage. This is particularly important in the search for proteotypic peptides, which are useful to acquire specie-specific information that are extremely useful for delineating domestication timelines and paleo-dietary habits, for instance

In **chapter V,** the developed strategies were successful employed to examine specific case studies. Diverse samples, different for origin, age and typology, were treated. The rescued fragments of the destroyed Giant Buddhas of Bāmiyān were analyzed by this proteomics approach and advanced mass spectrometry methods. Despite their age and degradation, caseins were identified in the painting layers, and study of sequence alignment allowed us to demonstrate the application of cow and goat milk as binder in ancient painting techniques.

The characterization of wall paintings from a Roman Archeological site in Cuma was also carried out. Differences between a private indoor (*domus*) and public outdoor (*temple*) painting decorations were observed. Moreover, a first example of art medium known as tempera was observed by the analysis of a little stone flake from a 49,000 years-old level of Sibudu (South Africa). In particular, the amazing discovery was to identify a mixture milk and collagen from *Bos taurus* in presence of ochre as pigment, in a sample that dates back to an age where domestication was not discovered yet.

Chapter VI deals with the first example of the application of targeted MS method (Multiple Reaction Monitoring, MRM) to samples in the field of cultural heritage of this dissertation. Despite the tremendous success of standard based proteomics approaches, the demand of improvement in sensitivity of protein detection in highly complex and rather contaminated samples drove the development of a targeted assay. To make this approach rational and efficient, a database was created to identify peptide markers of commonly used protein-based binders (milk, eggs and animal glue), by collecting and rationalizing the set of analyses that had been carried out on 107 pictorial samples during years with standard discovery protocols in our laboratory.

References

1. Nadolny, J. The first century of published scientific analyses of the materials of historical painting and polychromy,circa 1780-1880. *Stud. Conserv.* **48,** 39–51 (2003).
2. Hofreiter, M., Collins, M. & Stewart, J. R. Ancient biomolecules in Quaternary palaeoecology. *Quat. Sci. Rev.* **33,** 1–13 (2012).
3. Schwarz, C. *et al.* New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res.* **37,** 3215–3229 (2009).
4. Rees-Jones, S. G. *The Organic Chemistry of Museum Objects*. (Elsevier Science, 1987).
5. Hynekl, R., Kuckova, S., Hradilova, J. & Kodicek, M. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry as a tool for fast identification of protein binders in color layers of paintings. *Rapid Commun. Mass Spectrom.* **18,** 1896–1900 (2004).
6. Kuckova, S., Hynek, R. & Kodicek, M. Identification of proteinaceous binders used in artworks by MALDI-TOF mass spectrometry. *Anal. Bioanal. Chem.* **388,** 201–206 (2007).
7. Fallis, A. . No Title No Title. *J. Chem. Inf. Model.* **53,** 1689–1699 (2013).
8. International Council on Monuments and Sites. ICOMOS International Cultural Tourism Charter : Principles And Guidelines For Managing Tourism At Places Of Cultural And Heritage Significance. *Tourism* 1–9 (2008).
9. Colombini, M. P. & Modugno, F. Characterisation of proteinaceous binders in artistic paintings by chromatographic techniques. *J. Sep. Sci.* **27,** 147–160 (2004).
10. Lewandowski, C. M., Co-investigator, N. & Lewandowski, C. M. *Organic Mass Spectrometry in Art and Archaeology*. *The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference* **1,** (2015).
11. Aakacha, M. A., Pawlak, K. & Jarosz, M. Identification of anthraquinone coloring matters in natural red dyestuffs by high performance liquid chromatography with ultraviolet and electrospray mass spectrometric detection. *J. Sep. Sci.* **26,** 1028–1034 (2003).
12. Eunice, C. Y., Li-Chan & Kim, H.-O. Structure and chemical composition of eggs. *Egg Biosci. Biotechnol.* 1–96 (2008). doi:10.1002/9780470181249.ch1
13. Harrison, S. M., Kaml, I., Prokoratova, V., Mazanek, M. & Kenndler, E. Animal glues in mixtures of natural binding media used in artistic and historic objects: Identification by capillary zone electrophoresis. *Anal. Bioanal. Chem.* **382,** 1520–1526 (2005).
14. Dallongeville, S., Garnier, N., Rolando, C. & Tokarski, C. Proteins in Art, Archaeology, and Paleontology: From Detection to Identification. *Chem. Rev.* (2015). doi:10.1021/acs.chemrev.5b00037
15. Bonaduce, I., Colombini, M. P. & Diring, S. Identification of garlic in old gildings by gas chromatography-mass spectrometry. *J. Chromatogr. A* **1107,** 226–232 (2006).
16. Ahuja, J. K. C. *et al.* Composition of Foods Raw , Processed , Prepared USDA National Nutrient Database for Standard Reference , Release 27 Documentation and User Guide. *U.S. Dep. Agric. Agric. Res. Serv. Beltsv. Hum. Nutr. Res. Cent. Nutr. Data Lab.* **2,** 1–136 (2013).
17. Schilling, M. R. *et al.* Gas Chromatographic Analysis of Amino Acids as Ethyl Chloroformate Derivatives **35,** (2014).
18. Bott, R. *An Introduction to Archaeological Chemistry*. *Igarss 2014* (2014). doi:10.1007/s13398-014-0173-7.2
19. Jakes, K. A. *Archaeological Chemistry*. (American Chemical Society, 2002). doi:10.1021/bk-2002-0831
20. Smith, L. Archaeological Theory and the Politics of Cultural Heritage. 200–260 (2004).
21. Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **23,** 3843–54 (2009).

22. Gowda, H. Advances in Genomics and Proteomics : Their Role in Future of Molecular Medicine.
23. Blackstock, W. P. & Weir, M. P. Proteomics: Quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **17,** 121–127 (1999).
24. Graves, P. R. & Haystead, T. a J. Molecular Biologist ' s Guide to Proteomics. *Microbiol. Mol. Biol. Rev. Microbiol. Mol. Biol. Rev.* **66,** 39–63 (2002).
25. Fournier, M. L., Gilmore, J. M., Martin-Brown, S. A. & Washburn, M. P. Multidimensional separations-based shotgun proteomics. *Chem. Rev.* **107,** 3654–3686 (2007).
26. Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10,** 1785–1793 (2011).
27. Cooks, R. G., Howe, I. & Williams, D. H. Structure and fragmentation mechanisms of organic ions in the mass spectrometer. *Org. Mass Spectrom.* **2,** 137–156 (1969).
28. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60,** 2299–2301 (1988).
29. Whitehouse, C. M., Dreyer, R. N., Yamashita, M. & Fenn, J. B. Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.* **57,** 675–679 (1985).
30. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68,** 1–8 (1996).
31. Pandey, a & Mann, M. Proteomics to study genes and genomes. *Nature* **405,** 837–846 (2000).
32. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).
33. Wysocki, V. H., Resing, K. A., Zhang, Q. & Cheng, G. Mass spectrometry of peptides and proteins. *Methods* **35,** 211–222 (2005).
34. Payne, A. H. & Glish, G. L. Tandem mass spectrometry in quadrupole ion trap and ion cyclotron resonance mass spectrometers. *Methods Enzymol.* **402,** 109–148 (2005).
35. Hager, J. W. Recent trends in mass spectrometer development. *Anal. Bioanal. Chem.* **378,** 845–850 (2004).
36. Schwartz, J. C., Senko, M. W. & Syka, J. E. P. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **13,** 659–669 (2002).
37. Wadsworth, C. & Buckley, M. Proteome degradation in fossils: Investigating the longevity of protein survival in ancient bone. *Rapid Commun. Mass Spectrom.* **28,** 605–615 (2014).
38. Makarov, A. *et al.* Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **78,** 2113–2120 (2006).
39. Schilling, G. D. *et al.* Detection of Positive and Negative Ions from a Flowing Atmospheric Pressure Afterglow Using a Mattauch-Herzog Mass Spectrograph Equipped with a Faraday-Strip Array Detector. *J. Am. Soc. Mass Spectrom.* **21,** 97–103 (2010).
40. McLuckey, S. a, Goeringer, D. E. & Glish, G. L. Collisional activation with random noise in ion trap mass spectrometry. *Anal. Chem.* **64,** 1455–60 (1992).
41. McLafferty, F. W. Tandem Mass Spectrometry. **8,** 85–107 (1985).
42. Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S. & Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 6233–7 (1986).
43. Papayannopoulos, I. A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* **14,** 49–73 (1995).
44. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5,** 699–711 (2004).
45. Applications, Q. *et al.* of Mass Spectrometry Quantitative Applications of Mass Spectrometry. (2006).
46. Boyd, R. K. ., Basic, C. & Bethem, R. A. *Trace Quantitative Analysis by Mass Spectrometry*.

(2008). doi:10.1002/9780470727140
47. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4,** 222 (2008).
48. Loboda, a V, Krutchinsky, a N., Bromirski, M., Ens, W. & Standing, K. G. A tandem quadrupole / time-of-flight mass spectrometer (QqTOF) with a MALDI source : design and performance. *Rapid Commun. Mass Spectrom.* (2000).
49. Douglas, D. J. & French, J. B. Collisional focusing effects in radio frequency quadrupoles. *J. Am. Soc. Mass Spectrom.* **3,** 398–408 (1992).
50. Krutchinsky, an, Chernushevich, I., Spicer, V., Ens, W. & Standing, K. Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. *J. Am. Soc. Mass Spectrom.* **9,** 569 (1998).
51. Guilhaus, M., Selby, D. & Mlynski, V. Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrom. Rev.* **19,** 65–107 (2000).
52. Kelleher, N. L. *et al.* Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **121,** 806–812 (1999).
53. Sandu, I. C. A. *et al.* An innovative, interdisciplinary, and multi-technique study of gilding and painting techniques in the decoration of the main altarpiece of Miranda do Douro Cathedral (XVII-XVIIIth centuries, Portugal). *Microsc. Res. Tech.* **76,** 733–743 (2013).
54. Van Der Werf, I. D., Calvano, C. D., Laviano, R., Simonetti, A. & Sabbatini, L. Multi-technique chemical characterisation of a 12-13th-century painted Crucifix. *Microchem. J.* **106,** 87–94 (2013).
55. Kirby, D. P., Khandekar, N., Arslanoglu, J. & Sutherland, K. Protein identification in artworks by peptide mass fingerprinting. *ICOM-CC 16th Trienn. Conf. Lisbon 19-23 Sept. 2011 Prepr.* 1–10 (2011). doi:10.1385/1-59745-275-0:61
56. Kuckova, S. *et al.* Complementary cross-section based protocol of investigation of polychrome samples of a 16th century Moravian Sculpture by optical, vibrational and mass spectrometric techniques. *Microchem. J.* **110,** 538–544 (2013).
57. Corso, G. *et al.* Characterization of pigments and ligands in a wall painting fragment from Liternum archaeological park (Italy). *J. Sep. Sci.* **35,** 2986–2993 (2012).
58. Dallongeville, S. *et al.* Proteomics applied to the authentication of fish glue: application to a 17th century artwork sample. *Analyst* **138,** 5357–5364 (2013).
59. Chambery, A. *et al.* Improved procedure for protein binder analysis in mural painting by LC-ESI/Q-q-TOF mass spectrometry: Detection of different milk species by casein proteotypic peptides. *Anal. Bioanal. Chem.* **395,** 2281–2291 (2009).
60. Fremout, W. *et al.* Tryptic peptide analysis of protein binders in works of art by liquid chromatography-tandem mass spectrometry. *Anal. Chim. Acta* **658,** 156–162 (2010).
61. Damodaran, S., Wood, T. D., Nagarajan, P. & Rabin, R. A. Evaluating Peptide Mass Fingerprinting-based Protein Identification. *Genomics, Proteomics Bioinforma.* **5,** 152–157 (2007).
62. Kuster, B., Schirle, M., Mallick, P. & Aebersold, R. Scoring proteomes with proteotypic peptide probes. *Nat. Rev. Mol. Cell Biol.* **6,** 577–83 (2005).
63. Cox, J. & Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80,** 273–299 (2011).
64. Held, I. M. *et al.* Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Clim.* **23,** 2418–2427 (2010).
65. Kettenbach, A. N., Rush, J. & Gerber, S. A. Absolute quantification of protein and post-translational modification abundance with stable isotope–labeled synthetic peptides. *Nat. Protoc.* **6,** 175–186 (2011).

66. Brönstrup, M. Absolute quantification strategies in proteomics based on mass spectrometry. *Expert Rev. Proteomics* **1,** 503–512 (2004).
67. Brun, V. *et al.* Isotope-labeled Protein Standards: Toward Absolute Quantitative Proteomics. *Mol. Cell. Proteomics* **6,** 2139–2149 (2007).
68. van Doorn, N. L., Wilson, J., Hollund, H., Soressi, M. & Collins, M. J. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun. Mass Spectrom.* **26,** 2319–2327 (2012).
69. Robinson, N. E. & Robinson, a B. Molecular clocks. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 944–9 (2001).
70. Schweitzer, M. H. *et al.* and Protein Sequences of the. 626–631 (2009). doi:10.1126/science.1165069
71. Ostrom, P. H. *et al.* New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. *Geochim. Cosmochim. Acta* **64,** 1043–1050 (2000).
72. Leo, G. *et al.* Deamidation at asparagine and glutamine As a major modification upon deterioration/aging of proteinaceous binders in mural paintings. *Anal. Chem.* **83,** 2056–2064 (2011).

IV

Development of proteomic strategies for Cultural Heritage

Nowadays the knowledge of the materials used in a work of art is of greatest significance, not only to give an insight in the historical context of objects and artists, but also to analyze degradation processes taking place in aged objects and to develop appropriate conservation and/or restoration treatments[1].

Proteomic procedures, routinely applied for typical modern samples, still need to be fully adapted to heterogeneous materials in ancient samples of several cultural heritage objects. For this purpose is important to take into account the low amount of proteinaceous material, the heterogeneity and the unusual physical state of the samples, as well as the high levels of damage found in ancient samples.

This chapter deals with some examples of the development and adaptation of classical proteomic strategies in the analysis of ancient samples to meet the different aims in the cultural heritage field. Overall, it is intended to illustrate the logic that might address the choice of sample treatment and data analysis to get the most of the experiment.

IV.1. Introduction

Following the seminal papers by Hynek et al.[2] and Tokarsky et al. [3] several proteomic strategies, aimed at the unequivocal identification of proteins present in art objects or in archaeological remains, have been designed by extensively exploiting modern mass spectrometry.

In particular, although analyses in proteomics are per se invasive, modern mass spectrometry instrumentations enable the characterization of proteins with extremely high sensitivity even in crude mixtures in which the dynamic range of components abundance exceeds 1000-fold, and on very limited amount of sample, typically less than 10 μg. Modern instruments are therefore perfectly adequate to afford the minimal quantities of ancient samples.

However, all the steps of the proteomic procedure need to be thoughtfully adapted, from the optimization of specific protocols for sample preparation to the development of data analysis tools in order to cope with ancient, damaged samples.

IV.1.1 The proteomics for Cultural Heritage so far

The very critical step in the application of proteomic approach is related to the efficiency of either protein extraction from solid matrices or proteolytic digestion of substrates incorporated within the matrix itself. The commonly used "bottom-up" approach to identify proteins is based on the enzymatic digestion of the proteins and can be directly performed in heterogeneous phase on sample fragments followed by mass spectrometric analyses of the released peptides (Figure IV.1). Even when the protein is embedded in a complex mixture such as that of a painting layer, few peptides released by the protease without any pre-treatment of the sample are enough to identify the protein [4], without significantly affecting the sample itself. This avoid the extraction of the whole protein from the sample using harsh methods, since digestion can rather be carried out by depositing on the surface of a small sample an aqueous neutral solution containing the enzyme that can directly trims protruding peptides. The solution is then gently removed and, once released, peptides can be analysed by mass spectrometric methods such as MALDI-TOF and LC-MS/MS, and database searches by bioinformatics tools such as Mascot (www.matrixscience.com) allow protein identification. This micro-invasive protocol (i.e. the intervention is intrinsically invasive but requires only a minimal quantity of material to work with) proved to be successful when applied to fragments of paintings from the collapsed vault of the Basilica di S. Francesco in Assisi [1], and samples collected from the Camposanto Monumentale in Pisa [3], and it was not significantly affected by the pigment that was present in the sample, i.e. the different metals do not actually affect the quality of the results[4]. It is worth mentioning that this protocol is closely similar to the procedure adopted when bio-cleaning of works of art is carried out [5], and therefore the meaning of "destructive or invasive" approach should be resized and this can be safely considered as a minimally invasive or rather micro-invasive procedure.

Protein identification can also fail because of overwhelming proteins from unavoidable contaminants. Samples coming from artworks have, indeed, an intrinsic contamination problem that originates from i.e. environmental exposure, restoration interventions, and so on, that cannot be overcome by just operating all the chemical manipulations in controlled conditions as in ordinary proteomic analysis. To circumvent unavoidable problems arising from "historical" contaminations, an exclusion list of the peptides that, in a first LC-MS/MS run allowed for identification of keratins or other protein contaminants and that in a further subsequent run have to be ignored by the mass spectrometer for fragmentation, can be adopted. It is a sort of instrumental trick to avoid "waste" of the mass spectrometer's time in fragmenting "useless" peptides derived from trypsin autodigestion or common protein

contaminants [4]. The list is created *ad hoc* from the raw data of the LC-MS/MS analysis and a second analysis is carried out with the same LC-MS/MS method, but for the addition of the exclusion list.

Moreover, although for merely identifying purposes, the detection of as few as two peptides is sufficient to properly pinpoint the protein, the characterization of the modifications induced by aging and deterioration processes requires a deeper examination of most of the protein primary structure.

IV.1.2 Analytical strategies for more specific analyses

While two or very few peptides can be enough to merely identify materials, more extended sequence coverage could be needed for more detailed analyses, such as species discrimination, and/or conservation state evaluation. In fact, if the goal is to discriminate the organism of origin of proteins that are highly conserved throughout the evolution and among species, it is important to identify proteotypic peptides, i.e. peptides that are unique to the protein sequence specific for an individual organism and not in common to other species.

This can be easily explained with the example of milk proteins: among the peptides that have been identified in the analysis of an ancient food residue by Hong et al. [6], some are shared by goat/sheep and cattle sequences while others can be used to discern the origin of milk in the ancient residue. Similarly, detection of β-lactoglobulin in dental calculus is per se a direct evidence of milk consumption and can constitute a signature for adoption of dairying habits. Most interestingly, identification of specie-specific peptides allowed discrimination of the origin of dairy product whether they are cattle, sheep or goat dairy product [7]. Specie-specific information that are peculiar of proteomics are therefore extremely useful for delineating domestication timelines and paleodietary habits. To meet aims like these, however, the mild protocols described above might be not always adequate and alternative procedures might be required in order to obtain a higher number of peptides and consequently higher sequence coverage, thus increasing the probability of detecting proteotypic peptides.

This is definitively the case of collagen from bones, where an in depth analysis might be needed to gain extremely important information that can be used for a wide range of purposes, from taxonomic analyses [8–11], to the characterization of degradation processes [12]. Proteins and collagen in particular, are indeed gaining momentum and are now supporting DNA in evolutionary studies, since they are more stable than nucleic acids, can persist much longer, and, moreover, the survived molecules bear the signature of time.

As commented by the authors of a tremendous paper in Nature this year [10,13] "ancient proteins could now prove as revolutionary as DNA for studying the tree of life", since we could be able to find proteins that are orders of magnitude older than the oldest DNA discovered so far. Intrinsic collagen protein stability and its entrapment within the hydroxylapatite protective cage make the collagen the longest surviving protein in ancient bone [14]. Consequently and adversely, collagen is highly resistant to extraction and it can be quite challenging to obtain a good sequence coverage that would allow highly reliable sequence comparisons for taphonomic studies (i.e. studies of decaying organisms over time).

Deamidation can occur also as a by-product of sample preparation. To consider deamidation as a signature of aging, its genuine pre-extraction origin must be verified. $H_2{}^{18}O$ labelling can be used to assess any deamidation occurring during digestion process [15–17], and Figure IV.2 illustrates the different mass shifts that would be observed upon hydrolysis in $H_2{}^{18}O$ in a glutamine containing peptide as a function of deamidation occurring before or during the treatment. However, simple rules can also be used in evaluating the significance of deamidation: conditions that could favour deamidation reaction such as extreme of pH or high temperature should be avoided, for instance, in collagen extraction, or, since glutamine deamidation is much slower than asparagines conversion to aspartic, glutamine containing peptides should be preferred over asparagines containing ones in the seek for markers of deamidation.

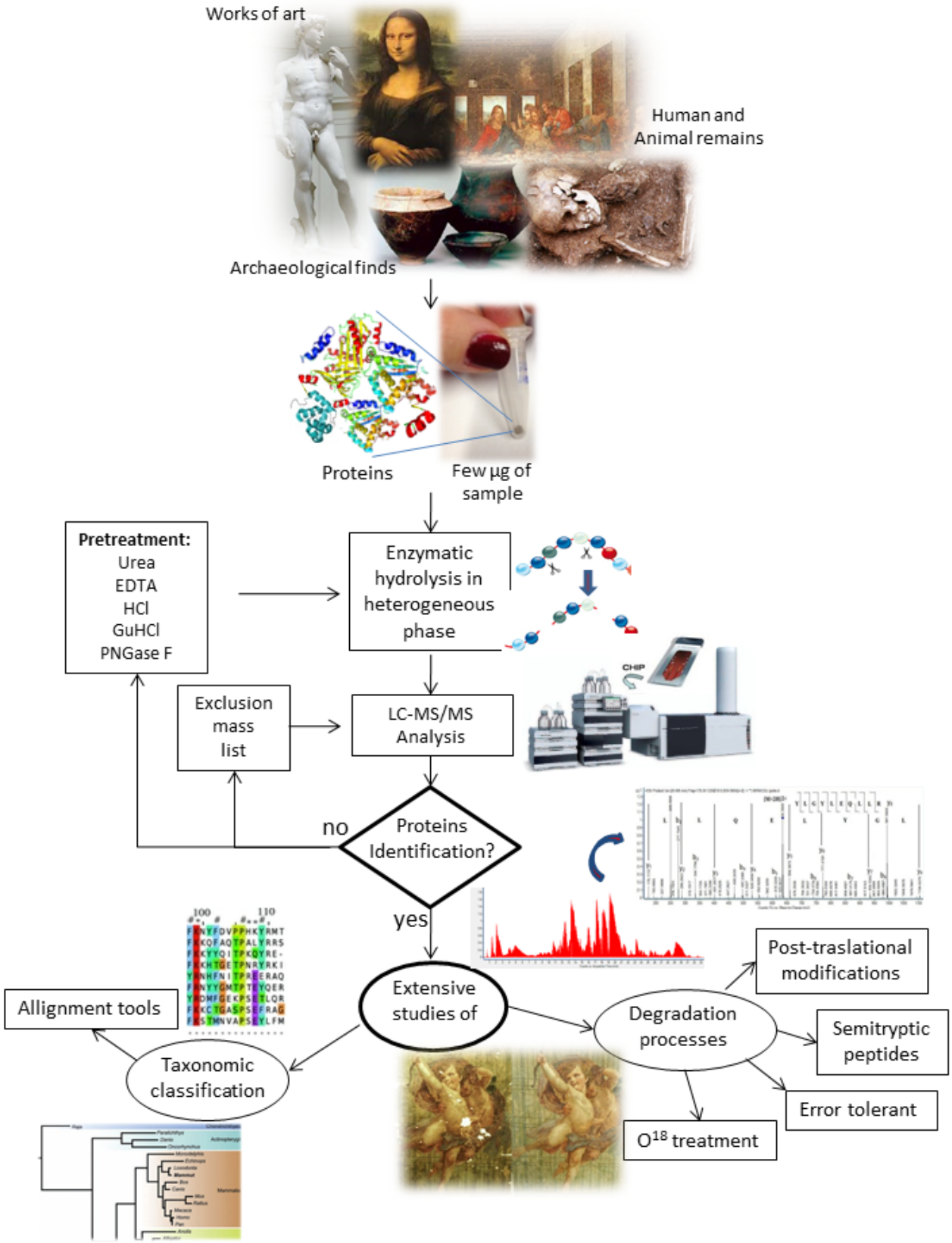


Figure IV.1. General flow chart of proteomic investigations.

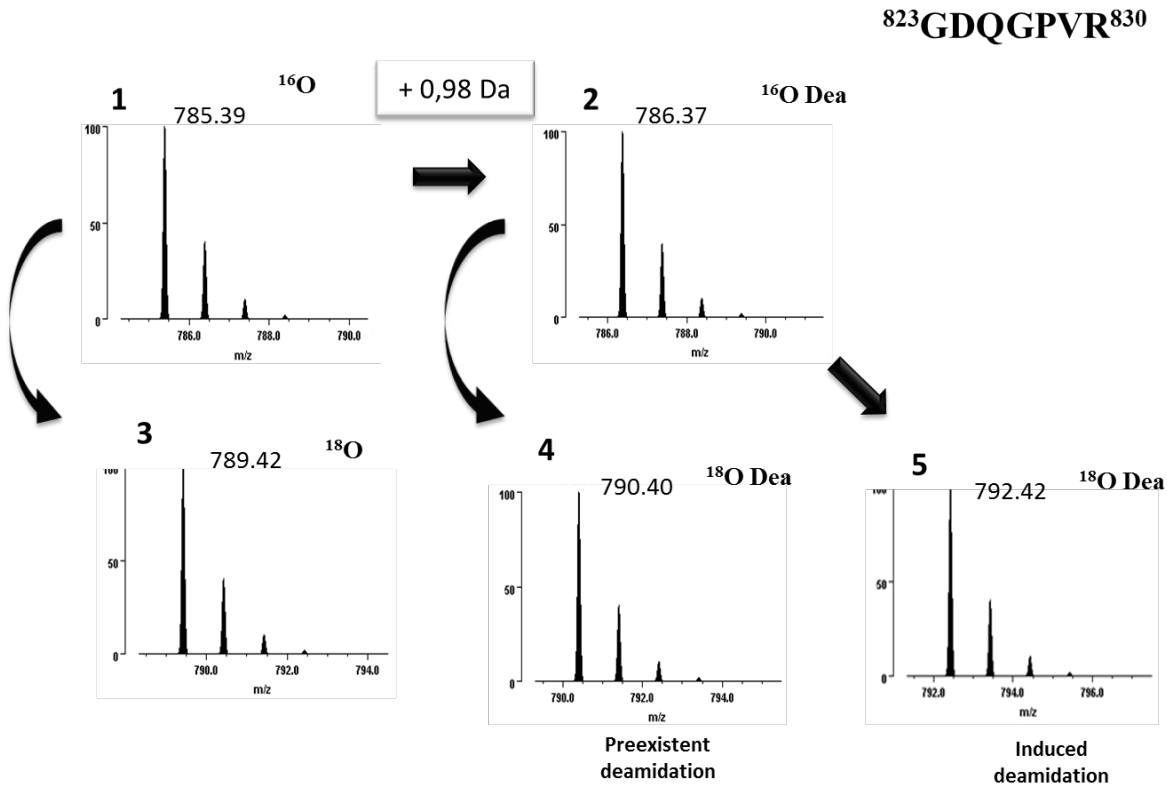


Figure IV.2. Scheme of expected mass shift upon enzymatic digestion in $H_2{}^{18}O$ of the peptide GDQGPVGR of collagen alpha-1(I). Upon digestion with trypsin in $H_2{}^{18}O$ up to 2 ^{18}O are incorporated at the C-terminus of the generated peptide with a mass shift of about 4 Da (3) in respect to digestion in $H_2{}^{16}O$ (1). Different mass shifts are observed when deamidation occurs during the trypsin digestion (5) or is a preexistent modification of the peptide (4) in respect to enzymatic hydrolysis in $H_2{}^{16}O$ (2).

The use of extraction protocol as mild as possible is also important when searching for unpredicted modifications induced by aging processes, since our knowledge of alterations occurring to proteins because of diagenetic events as well as environmental factors is not yet complete, and we need to be sure that the detected modifications are not induced during sample processing as it could for instance occur in arsh sample treatments.

The experimental part is not the only part that has to be thoughtfully considered and adopted to the specific requirements of ancient samples. The bioinformatics tools, that can address the specific issues, such as the identification of aging signatures in proteins of ancient samples or to handle extinct species when genomic data are available for the extant ones, are another challenge in the field of proteomics for cultural heritage.

A main difficulty in identifying ancient proteins is indeed the paucity of ancient genomic data. In shotgun proteomics, proteins are routinely identified by matching experimental MS/MS spectra of enzymatic digests of protein samples to simulated spectra from protein databases,

usually derived from genomic sequences. Identification in database search procedures rely on how well experimental spectra fit to theoretical spectra obtained from sequences that are present in databases. While this procedure works for most of the identifying purposes such as whether egg, milk or animal glue has been used as binder in a painting [4,3] or organic component in food residues [6,18] the issue might be more complex when trying to assess the species of origin and/or in the case of extinct organisms. Standard database search algorithms fail to identify peptides that are not exactly contained in a protein database, such as those arising from unreported mutations occurred throughout evolution. Good quality fragmentation spectra are in principle discarded because of even point mutation that makes them non-ascribable to known sequences. Identification of the specific species will then depend on the detection of two or few more peptides that are conserved between the ancient protein and its modern counterpart, despite possible differences in other peptides, which will be ignored in a standard database searches. This is an intrinsic limit of the otherwise powerful proteomic approach that will mask novel sequences or points of divergence with reported sequences.

The tremendous attractive possibility of proteomics of reaching much further back in time to gain information on more ancient samples than genomics, thanks to the higher stability of proteins in respect to nucleic acids, and to characterize in depth diagenetic alterations directly on surviving molecules, can be therefore limited in principle by the lacks of corresponding DNA information.

However, many software tools have been developed for the automated identification of peptides by de novo sequencing directly from the MS/MS spectrum of peptides. Representative de novo sequencing software packages include PEAKS [19], PepNovo [20], NovoHMM [21], and Lutefisk [22], and very recently, Welker et al. [10] extensively and successfully used PEAKS to overcome the absence of corresponding genomic data to resolve the evolutionary history of Darwin's south American ungulates.

Moreover, the lacks of corresponding DNA information can also be partially overcome with bioinformatic tools such as the error tolerant search utility in Mascot, which allows for single substitution in peptides in respect to sequences which are present in databases, taking for granted that any obtained match has to be manually confirmed afterwards. These approaches allow taxonomic attribution of extinct species, in specimens that no longer can yield DNA [9] and can be combined to a wide panel of other bioinformatics tools such as generation *ad hoc* of peptide databases as experimented by Waters and collaborators [23] that used a collagen sequences database to analyse the mass spectrometric data obtained from a mastodon bone, and classical homology search tools like BLAST [10].

IV.2. Material and methods

IV.2.1 Reagents

Ammonium hydrogen carbonate (AMBIC), Ethylenediaminetetraacetic acid (EDTA); Tri(hydroxymethyl)aminomethane (TRIS), Urea, GuHCl and TPCK-treated trypsin were from Sigma; recombinant Peptide N-Glycosidase F (PNGaseF) was from Roche Formic acid and Acetonitrile (ACN) were purchased from Baker. Deionized water was obtained from Millipore cartridge equipment. Hydrochloric acid was purchased from Carlo Erba.

IV.2.2 Panting samples

Models of paint layers were prepared with egg white, milk and animal glue as binders and azurite ($Cu_3(CO_3)_2(OH)_2$), red ochre (Fe_2O_3), minium (Pb_3O_4), calcite ($CaCO_3$), and vermilion (HgS) as pigments on glass slides and skimmed milk as control without pigments. Paint replicas were left to dry at RT on the bench for one month. Paint replicas were pre-pared in 2010, left to dry at room temperature for one month, then artificially aged for 720 h at 25 °C, 50% rela-tive humidity in indoor light conditions, and stored since then in the darkness at room temperature. Details are re-ported elsewhere. [24,25]. Sample DSFL6 was collected from the gilded aureole of the angel in "Holy Conversation", mural painting by Amico Aspertini, 1506-1510, San Frediano Church, Lucca, Italy. Sample Purple 12M080 was collected from a mural painting of the urban district of Cuma archeological site (first century AD). Bone sample, was a fragment from human bone of the first century A.D.

IV.2.3 Sample treatment

Urea pre-treatment: 10 µL of a solution of 6M Urea were added to micro-samples (ca 300-800 µg) and incubated for 10 min at RT, followed by sonication for 20 min. Urea was then 6-fold diluted with water. EDTA pre-treatment: about 100 µL of a solution of 0.5M EDTA were added to the bone fragment 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 Urea protocol described above was applied.

HCl pre-treatment: 50 µL of 0.6M HCl were added to the bone fragment and incubated at 4°C for 4 hours. After centrifugation for 2 min at 10,000 rpm in a benchtop microfuge the supernatant was removed and washed with 20 µL of 10mM Ambic. Washes were repeated for four times. 100 µL of 50mM Ambic were added and sample was left at 65°C for 3 hours. After centrifugation at 10,000 rpm for 15 minutes the supernatant was removed.

GuHCl pre-treatment: 200 µL of 0.6M HCl were added and incubated at 4°C for 18 hours. After centrifugation for 1 min at 14,000 rpm in a benchtop microfuge, and the acid-insoluble pellet washed three times with 200 µL of distilled water. The pellet was incubated at 4 °C for 72 h in a buffer containing 100mM Tris and 6M GuHCl at pH 7.4. The sample was then centrifuged for 1 min at 14,000 rpm in a benchtop microfuge. The supernatant was buffer-exchanged into 10mM Ambic using 3 K molecular weight cut-off Amicon Ultra, centrifugal filter unit.

PNGaseF pre-treatment: 50 µL of AMBIC 50 mM containing 60m U/µL of PNGaseF solution were added to microsamples (ca 300-800 µg), and incubating at 37°C for two hours. The reaction was stopped by incubation of the sample in boiling water for 2 min.

IV.2.4 Protein digestion and LC-MS/MS analysis

After any pre-treatment of the sample, enzymatic digestion was carried out as in the minimally invasive proteomic analytical procedure described by Leo *et al.* [4]. Briefly, trypsin was added to a final concentration of 10ng/µL to micro-samples (ca 300-800 µg) as directly suspended in 50 µL of 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 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 400nL/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 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.

IV.2.5 Data handling

The acquired MS/MS spectra were transformed in Mascot Generic files (.mgf) format and used to query the SwissProt database 2015_04 (548208 sequences; 195282524 residues), with *Chordata* as taxonomy restriction for protein identification in paint reconstruction samples and with *Homo sapiens* for bone samples.

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, formation of pyroglutammic acid from glutamine residues at the N-terminal position of peptides, and deamidation at asparagines and glutamines were considered as variable modifications [11]. 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 [11]. 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 generally 31 for paintings and 43 for human samples. Spectra with Mascot score <10 were rejected.

IV.3 Results and discussion.

IV.3.1 Denaturation pre-treatments

It is evident from what outlined in the introduction that accessibility of proteases to the protein in the sample is likely the most important aspect for a successful proteomic experiment in cultural heritage application (obviously beside the actual presence in the sample of proteinaceous material). Whatever reduces the proteolysis yield and, therefore, determines an inefficient and poor production of peptides, would sensibly affect protein identification. For instance, the persistence of stabilizing interactions such as those occurring in structured proteins can greatly impair the efficacy of the enzymatic digestion. By reasoning as in classical biochemical experiments, where denatured, unfolded proteins are digested much more efficiently than structured, folded proteins, we introduced a denaturing step before trypsin digestion to “open” residual structural elements in proteins in the perspective that flexibility of the polypeptide substrate is an absolute requirement for the protease to properly hydrolyze peptide bonds. In the hypothesis that treatment with traditional protein denaturing

agents such as urea or guanidinium chloride will make digestion sites more amenable to protease attack even in an unusual "dehydrated, non-soluble" physical state experienced by proteins in works of art or archaeological remains, we tested the denaturing pre-treatment in some cases where the minimally invasive approach described above failed. In our laboratory, this pre-treatment had already been successfully applied to some gilding samples [26] and to ink samples from the Qumran archaeological site [27], but never rationalized. I therefore set up a set of experiments aimed to systematically test and prove the efficiency of the denaturing treatment. As an example, Table IV.1 reports a comparison of the results obtained by the minimally invasive approach, and the analysis carried out on the same pictorial sample after pre-treatment with 6M urea, followed by dilution and tryptic digestion. The results clearly show that the pre-treatment with urea improves the quality of identification. I apply the protocol to a small stone flake from Sibudu Cave dated 49000 years ago as described in the next chapter [11].

This pre-treatment is slightly more aggressive and therefore more invasive than that described above, and it should be used as a second attempt, when the above protocol with only protease in bicarbonate buffer fails or when more peptides or higher quality spectra are required. It is worth noting that it can be also directly applied on the same sample after the first attempt with the protease in the simple ammonium bicarbonate buffer has failed, thus not reducing the sample size. In principle, after the removal of the bicarbonate buffer containing the protease from the solid sample, the urea pretreatment can be carried out without any further step in between.

Table IV.1. Proteins identified in the paint replica containing minium and milk by LC-MS/MS. Aliquots were treated in heterogeneous phase with trypsin with the minimally invasive protocol either without any sample pre-treatment or with Urea pre-treatment and the resulting peptide mixtures were analyzed by LC-MS/MS. Proteins were identified in the Uniprot database with Mascot MS/MS Ion Search software, with Chordata as taxonomic restriction, with methionine oxidation, formation of pyroGlu at the N-terminus of Gln, and deamidation (N, Q) of peptides as variable modifications. Individual ion scores >31 indicate identity or extensive homology.
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits (http://www.matrixscience.com/help/interpretation_help.html).

Sample	Protein (UniProt Accession number)	Minimally invasive protocol			Urea pre-treatment		
		Sequence coverage (%)	Protein score	n° of peptides	Sequence coverage (%)	Protein score	n° of peptides
Milk and minium	alpha-S1 casein (P02662)	68	548	11	75	828	22
	alpha-S2 casein (P02663)	33	259	8	54	753	13
	beta-casein (P02666)	49	298	6	79	536	11
	kappa-casein (P02668)	22	166	5	74	319	8
	beta-lactoglobulin (P02754)	38	308	6	36	309	7

IV.3.2 Decalcification pre-treatments

To understand how it is possible to identify a greater number of peptides from a complex matrix such as bones, the sequence coverage of human COL1α1 (collagen alpha-1(I)) and COL1α2 (collagen alpha-2(I)), the two chains constituting type I collagen, are reported in Table IV.2. These results were obtained in the analysis of a human bone dated the first century A.D., using selected different and opportunely optimized extraction procedures, that were proposed for protein extraction from ancient bones [13,14,28]. It can be noted that even with the minimally invasive protocol, where the bone fragment is simply deepened in a trypsin solution, the number of peptides detected can be high enough to obtain good sequence coverage (24% and 23% for COL1α1 and COL1α2, respectively. However, when decalcification is carried out with EDTA 0.5M, nearly the whole sequences can be verified, reaching 75% and 72% of sequence coverage for mature COL1α1 and COL1α2, respectively.
It has to be noted, however, that while decalcification with EDTA seems to be the ultimate choice at this stage for collagen protein sequence coverage, it might not be the best choice when non-collagenic proteins (NCPs) are searched[14,29]. NCPs have potentially higher phylogenetic value than collagen I because of their greater sequence variation, but they are less resistant than collagen I, constitute a relatively small fraction of the total protein content

of bones, and are more soluble than collagen and might then be lost in demineralization steps. The choice of the protocol to be used for bone treatment will ultimately depend on the protein/s of interest but also on the analytical application. In fact, the extraction protocol used might also influence the results in the following characterization of the proteins recovered from bones.

Table IV.2. Sequence coverage (%) of Collagen alpha-1(I) and Collagen alpha-2(I) from the digestion of an ancient human bone with different pretreatment protocols followed by LC-MS/MS analysis. Sequence coverage was calculated on the sequence of the mature form of the protein. Proteins were identified in the UniProt database with Mascot MS/MS Ion Search software, with *Homo sapiens* as taxonomic restriction, with methionine oxidation, formation of pyroGlu at the N-terminus of Gln, deamidation (N, Q), and hydroxylation (K, P) of peptides as variable modifications.

Protein (UniProt Accession number)	**Minimally invasive protocol** (%)	**Urea** (%)	**EDTA** (%)	**HCl** (%)	**GuHCl** (%)
Collagen alpha-1(I) (P02452)	24.0	60.0	75.0	37.5	63.4
Collagen alpha-2(I) (P08123)	22.8	59.4	72.0	27.5	52.2

IV.3.3 Detection of glutamine and asparagine deamidation

Obtaining as many peptides as possible and the choice of the appropriate protocol can be also extremely important to study the molecular modifications occurring post-mortem. It is still an open question whether deamidation of glutamine (Q) and asparagine (N) could be used as a dating technique in ancient sample, but it is widely accepted that deamidation can be considered as a biomolecular marker of deterioration and natural aging of proteins in artistic and archaeological materials [12,16,17,30]. Detection of deamidation is constantly considered in our data interpretation. It can offer interesting prospects in the evaluation of the conservation state of work of arts and archaeological remains. However, deamidation is a delicate modification, since it is strongly influenced by several parameters such as pH and temperature. Deamidation is also a delicate modification from a purely technical point of view, since it induces a mass shift of only 0.98 Da. Fragmentation spectra, however, not only clearly rule out any doubt about the modification but also localize the deamidation site within the peptide sequence (Fig. IV.3) since only fragments ions containing the deamidation site will differ between deamidated and non-deamidated peptide.

Several of the peptides identified in the analyses reported from here on throughout the thesis were deamidated. I systematically introduced deamidation as variable modification in the database searches and manually checked the fragmentation spectra.

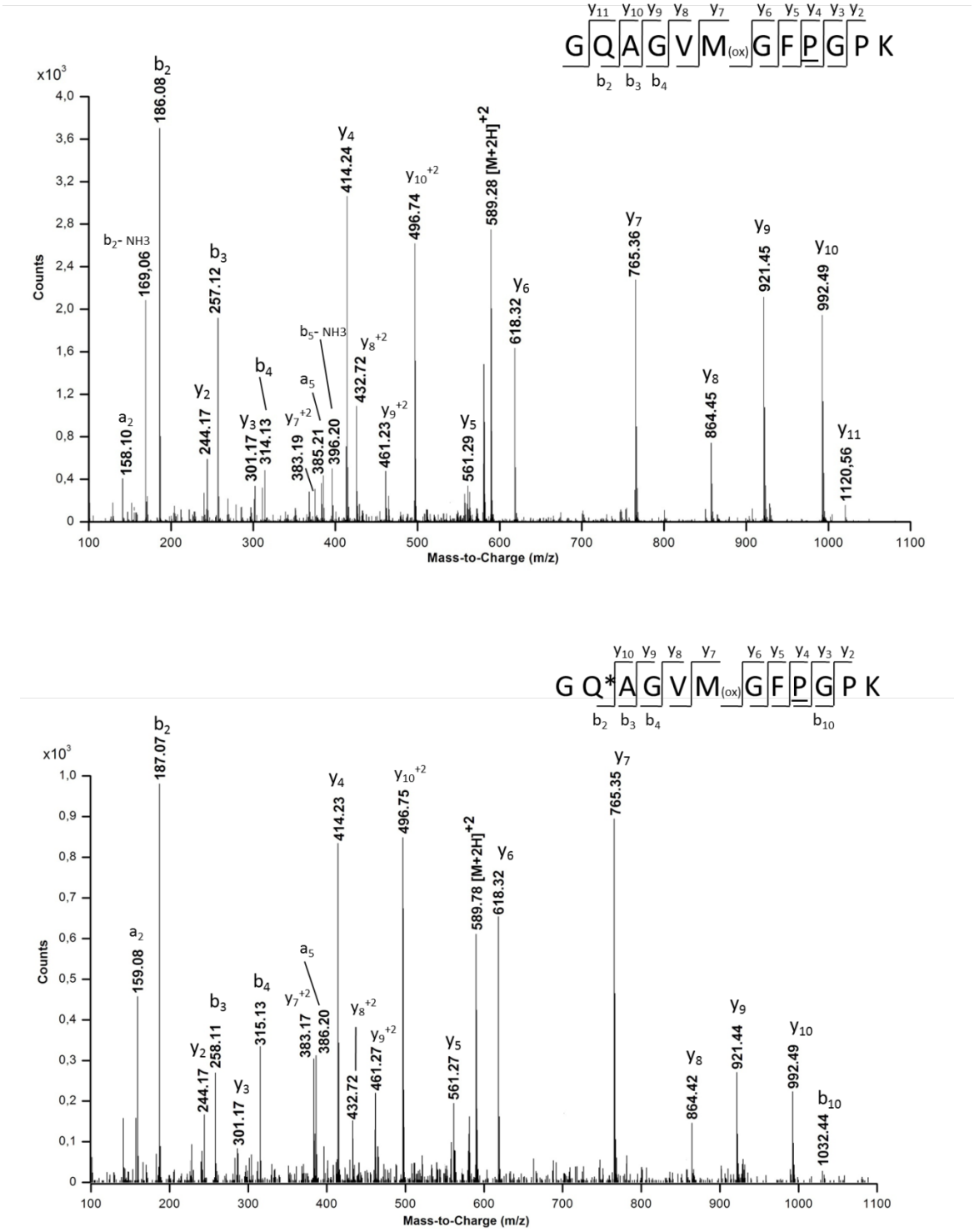


Figure IV.3. MS/MS spectra of the doubly charged ions at *m/z* 589.28 (A) of the peptide GQAGVMGFPGPK of human collagen alpha-1(I) (P02452) and its deamidated form at *m/z* 589.78 (B) identified in the analysis of the sample of human bone dated 1th A.D. The product ions are indicated with the observed mass. * Indicates deamidation site, and hydroxylation site is underlined.

V.3.4 A deglycosylation step to improve the identification of egg proteins

The very critical step in proteomic approach to ancient samples is the efficiency of either protein extraction from solid matrices or proteolytic digestion of substrates incorporated within the matrix itself. In some cases the difficulties related to the identification of the protein-based binders used in dry paints resides also in the molecular complexity of the protein in their physiological conditions. This is the case of the egg protein-based binders. In our experience [4], the detection of egg proteins remains often rather unsatisfactory and is certainly less confident than other most common binders, i.e. milk and animal glue.

It is well known that the most abundant proteins in hen egg, either in the albumen (i.e.: ovalbumin, ovotransferrin, ovomucoid, ovoglobulins, and beta-ovomucin) and in the yolk (i.e.: vitellogenin), are heavily glycosylated [31,32]. We reasoned that the extensive glycosylation of egg proteins might create a significant molecular hindrance, which hampers proteases to efficiently interact with the proteinaceous substrates, thus greatly decreasing proteolytic efficiency. This, in turn, would result in inefficient and poor production of peptides, impairing a confident protein identification. A possible approach to avoid this molecular hindrance would be to trim out the glycosidic decoration by preceding the proteolytic digestion with a deglycosylating step[33]. This working hypothesis was investigated and results are herein reported. The study demonstrated that the ingenuity of introducing a N-Glycosidase F digestion before the protease(s) treatment greatly increases the release of digested peptides, substantially improving the quality and reliability of egg identification in samples from artworks.

Therefore, the experimental observations that egg proteins are seldom identified in paint samples and, even when analyzing model samples, their identification is based on a small number of peptides, thus impairing a highly confident identification, prompted us to develop a novel analytical protocol to investigate and to circumvent this problem. Several enzymes have been success-fully applied to the release of N-linked glycans, such as peptide-N-glycosidase F (PNGaseF), endoglycosidase F and H. [34,35]. Among these, PNGaseF has emerged as a widely used glycoamidase [36,37]. As already shown in other circumstances[38], this enzyme is able to remove the glycosidic moieties leaving the protein substrate more amenable to tryptic digestion. A simple protocol was thus set up with a deglycosylating step before the tryptic digestion as de-scribed in the experimental section.

Paint reconstructions containing red ochre as pigment and alternatively albumen, or yolk or whole egg were used. Aliquots of each sample were analyzed using both the classical proteomic approach, based on simple treatment with trypsin (herein indicated as “Trypsin

alone") and the protocol with PNGaseF pretreatment (herein indicated as "PNGaseF and Trypsin"). It is worth noting that both steps of PNGaseF and trypsin digestions were carried out in heterogeneous phase, that means direct enzymatic digestion on solid samples without protein extraction [4]. This can be extremely useful when dealing with paint samples in which proteins are aged and bonded with pigments and fillers [24,25], and are thus difficult to solubilize. After digestion with trypsin, samples were analyzed by LC-MS/MS and proteins were identified by database search with the MS/MS Ion search mode within a licensed version of Mascot, with Chordata as taxonomic restriction in SwissProt protein database.

Table IV.3 reports a comparison of the results obtained on a replica made of albumen and red ochre. It is evident that pre-treating the sample with PNGaseF results in a much higher number of peptides obtained, greatly improving the reliability of the identification. In the aliquot treated with Trypsin alone, only ovalbumin and ovotransferrin were identified, with few detected peptides.

Table IV.3. Proteins identified in the paint replica containing red ochre and albumen by LC-MS/MS. Aliquots of paint replica were treated in heterogeneous phase with trypsin with or without treatment with PNGaseF (before and after tryptic digestion) and analysed by LC-MSMS. Details of the identification are given in supplementary Table IV.4.

Identified Protein. (Accession number)	Protocol					
	TRYPSIN		PNGaseF + TRYPSIN		TRYPSIN + PNGaseF	
	score	n° of peptides	score[a]	n° of peptides	score[a]	n° of peptides
Ovalbumin (P01012)	159	4	960	23	299	1123
Ovotransferrin (P02789)	90	2	2147	63	196	6
Ovostatin (P20740)			137	5	85	3
Ovalbumin releted protein Y (P01014)			452	11		
Mucin 5B (Q98UI9)			161	2		
Lysozyme C (P00698)			340	7		
Ovalbumin related protein X (P01013)			82	3		
Protein Tenp (O42273)			82	2		

[a] 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*Log(P), where P is the probability that the observed peptide match is a random event.

PNGaseF pretreatment led to a definitively more confident and reliable identification of 8 proteins from chicken egg white, showing also a much larger number of detected peptides for ovalbumin and ovotransferrin. The results are shown in Table IV.4, where the N-

glycosylation signature of deamidation in the Asn-X-Ser consensus sequence can be easily spotted in some of the observed peptides demonstrating the efficacy of the deglycosylation procedure. As control, an aliquot of sample was digested with trypsin and then with PNGaseF, in order to check whether the complexity of the sample mixture rather than the steric hindrance created by glycosylated peptides was responsible for the lack of results in PNGaseF untreated samples. The results clearly demonstrate that only when PNGase F treatment preceeds the enzymatic digestion, an improvement in the identification is observed.

Very similar results were obtained on paint replicas containing either whole egg or egg yolk as binder (Table IV.5).

In the absence of PNGaseF pretreatment, 7 proteins were identified in the whole egg samples while following PNGaseF incubation a total of 14 proteins could be confidently identified, mainly ovalbumin and related protein and vitellogenins ones. Similar results were obtained in the analysis of the egg yolk paint replica, where mainly vitellogenins were identified, demonstrating that the usefulness of PNGaseF treatment is not limited to sample containing proteins from albumen only, but it is extended to any sample where egg proteins are present (Table IV.5). Moreover, in all cases a higher number of peptides was detected compared to the trypsin alone protocol (from 2 to 19 times more).

In order to investigate if the PNGaseF treatment affects the identification of milk and animal glue, the other two common proteinaceous paint binders, we analyzed with the newly developed protocol paint replicas containing the same pigment, red ochre, mixed with milk and animal glue, as well as different combinations of binders. Table IV.6 shows that PNGaseF treatment had no substantial effects on the results when the sample does not contain egg. This result is not surprising, since milk proteins and collagen, the most abundant protein in animal glue, are much less glycosylated than egg proteins, and are thus good substrates for tryptic hydrolysis even in the absence of the deglycosylation step. On the contrary, whenever egg is present in mixture with other binders, PNGaseF pretreatment improved identifications.

As in paint samples inorganic materials are simultaneously present, mostly as pigments and dryers, it is necessary to investigate whether they can inhibit the activity of PNGaseF. To meet this aim, paint replicas containing a range of metals commonly occurring in paintings were selected. In particular the replicas contained albumen and cinnabar (HgS), minium (2 $PbO{\bullet}PbO_2$), and azurite ($Cu_3(CO_3)_2(OH)_2$), alongside the previously tested red ochre (Fe_2O_3). Experiments were carried out as above, on aliquots of each sample using the Trypsin alone and the PNGaseF + Trypsin protocols in parallel.

Data clearly indicate that PNGaseF treatment always improved the egg identifications, regardless the pigment used (Table IV.7), by increasing the number of peptides that are released from the proteins after tryptic digestion. These results indicate that PNGaseF is effective also in the presence of some widespread metal containing pigments, including copper. Moreover, the data clearly show that the newly developed protocol is fundamental in order to maximize the possibility of identifying egg in an aged paint sample. It is well known, in fact, that pigments can affect the protein identification by amino acid analysis if suitable purification steps are not adopted [39,40].

To fully understand the influence of pigments in the protein identification through proteomics procedures, the effect of pigments still needs be systematically investigated. Despite this, the data presented here clearly indicate that pigments do have an influence in the number of proteins that can be identified by proteomics analysis in a paint sample, as a consequence of the strong interactions taking place between pigments and proteins [25,41], and the developed protocol is a suitable analytical tool to help us to improve our success rate.

Table IV.4. Details of the identification of proteins in the paint replicas containing red ochre and albumen or whole egg by LC-MS/MS. Aliquots of paint replicas were treated in heterogeneous phase with trypsin with or without PNGaseF pretreatment and analysed by LC-MS/MS. Proteins were identified searching UniprotSprot database with MS/MS Ion search Mascot software (Matrix Science) with Chordata as taxonomy restriction, with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications. Consensus sequence for N-glycosilation is highlighted in grey in the matched sequence. Only proteins identified with at least two peptides were considered as significative.

ALBUMEN + RED OCHRE TRYPSIN				
Identified protein (Accession number)	**Total score**[a]	**Sequence coverage (%)**	**Individual ion scores**[b]	**Matched sequence**
Ovalbumin (P01012)	159	18	35 33 30 39	R.GGLEPINFQTAADQAR.E K.HIATNAVLFFGRCVSP.- K.ISQAVHAAHAEINEAGR.E R.ELINSWVESQTNGIIR.N
Ovotransferrin (P02789)	90	6	29 57	R.SAGWNIPIGTLLHR.G R.GAIEWEGIESGSVEQAVAK.F
ALBUMEN + RED OCHRE PNGaseF + TRYPSIN				
Identified protein (Accession number)	**Total score**[a]	**Sequence coverage (%)**	**Individual ion scores**[b]	**Matched sequence**
Ovalbumin (P01012)	960	44	30 29 19 52 21 16 46 (21) 39 48 (37) 17 (69) 77 80 (67) 67 (44) (88) 119 (82) (18) 29 22 51 56 (55)	K.VYLPR.M R.LYAEER.Y R.MKMEEK.Y R.VASMASEK.M K.IKVYLPR.M K.VYLPRMK.M R.TQINKVVR.F R.TQINKVVR.F + Dea (NQ) R.KIKVYLPR.M R.VASMASEKMK.I R.VASMASEKMK.I + Oxi (M) K.AFKDEDTQAMPFR.V K.LTEWTSSNVMEER.K K.LTEWTSSNVMEER.E + Oxi (M) R.GGLEPINFQTAADQAR.E R.GGLEPINFQTAADQAR.E + Dea (NQ) K.ISQAVHAAHAEINEAGR.E K.ISQAVHAAHAEINEAGR.E + Dea (NQ) R.ELINSWVESQTNGIIR.N R.ELINSWVESQTNGIIR.N + Dea (NQ) R.ELINSWVESQTNGIIR.N + 2 Dea (NQ) R.ELINSWVESQTNGIIR.N + 3 Dea (NQ) K.LTEWTSSNVMEERKIK.V R.LYAEERYPILPEYLQCVK.E K.ELYRGGLEPINFQTAADQAR.E R.DILNQITKPNDVYSFSLASR.L R.DILNQITKPNDVYSFSLASR.L + Dea (NQ)

			(37)	R.DILNQITKPNDVYSFSLASR.L + 2 Dea (NQ)
			(38)	R.VTEQESKPVQMMYQIGLFR.M
			49	R.VTEQESKPVQMMYQIGLFR.V + Dea (NQ)
			40	R.LYAEERYPILPEYLQCVKELYR.G
			20	R.VTEQESKPVQMMYQIGLFRVASMASEKMK.I
			37	R.GGLEPINFQTAADQARELINSWVESQTNGIIR.N
			14	R.LYAEERYPILPEYLQCVKELYRGGLEPINFQTAA DQAR.E
Ovotransferrin (P02789)	2147	66	21	K.DLLFK.D
			16	R.DLLER.Q
			29	K.GDVAFVK.H
			30	K.TCNWAR.V
			34	K.NKADWAK.N
			(18)	R.FGVNGSEK.S
			33	R.FGVNGSEK.S + Dea (NQ)
			53	R.ANVMDYR.E
			(39)	R.ANVMDYR.E + Oxi (M)
			25	R.DLTQQER.I
			22	K.DQLTPSPR.E
			35	K.IRDLLER.Q
			31	K.ATYLDCIK.A
			34	K.TSCHTGLGR.S
			38	K.RFGVNGSEK.S
			(38)	K.RFGVNGSEK.S + Dea (NQ)
			25	R.DLLERQEK.R
			61	R.FGVNGSEKSK.F
			(54)	R.FGVNGSEKSK.F + Dea (NQ)
			35	K.AQSDFGVDTK.S
			15	K.FMMFESQNK.D
			29	R.WCTISSPEEK.K
			43	K.DSNVNWNNLK.G
			(41)	K.RFGVNGSEKSK.F
			45	K.RFGVNGSEKSK.F + Dea (NQ)
			32	R.IQWCAVGKDEK.S
			20	K.IRDLLERQEK.R
			75	K.GTEFTVNDLQGK.T
			15	K.DQLTPSPRENR.I
			59	R.KDSNVNWNNLK.G
			17	K.SKFMMFESQNK.D
			(10)	K.SKFMMFESQNK.D + Oxi (M)
			24	K.DSNVNWNNLKGK.K
			43	K.DLLFKDSAIMLK.R
			(18)	K.DLLFKDSAIMLK.R + Oxi (M)
			40	K.KGTEFTVNDLQGK.T
			35	R.KDSNVNWNNLKGK.K
			(32)	R.KDSNVNWNNLKGK.K + Dea (NQ)
			17	K.DSNVNWNNLKGKK.S
			51	R.SAGWNIPIGTLLHR.G
			(36)	R.TAGWVIPMGLIHNR.T
			(41)	R.TAGWVIPMGLIHNR.T + Dea (NQ)
			46	R.TAGWVIPMGLIHNR.T + Dea (NQ); Oxi (M)
			52	K.FFSASCVPGATIEQK.L
			38	R.NAPYSGYSGAFHCLK.D
			36	R.KDSNVNWNNLKGKK.S
			20	K.TDERPASYFAVAVAR.K
			18	R.ENRIQWCAVGKDEK.S
			22	R.DDNKVEDIWSFLSK.A
			21	K.EFLGDKFYTVISSLK.T
			61	K.NLQMDDFELLCTDGR.R
			21	K.FMMFESQNKDLLFK.D
			22	K.TDERPASYFAVAVARK.D
			27	K.SDFHLFGPPGKKDPVLK.D
			26	K.GDVAFIQHSTVEENTGGK.N
			15	R.ENRIQWCAVGKDEKSK.C
			20	R.NAPYSGYSGAFHCLKDGK.G
			40	K.NLQMDDFELLCTDGRR.A

			17	R.ECNLAEVPTHAVVVRPEK.A
			78	R.GAIEWEGIESGSVEQAVAK.F
			26	K.GTEFTVNDLQGKTSCHTGLGR.S
			14	R.EGTTYKEFLGDKFYTVISSLK.T
			31	K.ADWAKNLQMDDFELLCTDGRR.A
			23	R.VAAHAVVARDDNKVEDIWSFLSK.A
			25	R.NAPYSGYSGAFHCLKDGKGDVAFVK.H
			26	R.DDNKVEDIWSFLSKAQSDFGVDTK.S
			16	R.ANVMDYRECNLAEVPTHAVVVRPEK.A+Oxi (M)
			21	R.ANVMDYRECNLAEVPTHAVVVRPEKANK.I
			60	R.SAGWNIPIGTLLHRGAIEWEGIESGSVEQAVAK.F
			20	R.VAAHAVVARDDNKVEDIWSFLSKAQSDFGVDT K.S
			56	K.RVPSLMDSQLYLGFEYYSAIQSMRKDQLTPSPR. E
			20	K.LKPIAAEVYEHTEGSTTSYYAVAVVKKGTEFTV NDLQGK.T
			16	K.HTTVNENAPDQKDEYELLCLDGSRQPVDNYKTC NWAR.V
			43	R.SAGWNIPIGTLLHRGAIEWEGIESGSVEQAVAKF FSASCVPGATIEQK.L
			46	R.SAGWNIPIGTLLHRGAIEWEGIESGSVEQAVAKF FSASCVPGATIEQK.L + Dea (NQ)
Ovostatin (P20740)	137	4		
			21	K.LSAEVAR.E
			15	R.GIPYFGQIK.L
			35	K.IFDPELSLK.A
			30	K.TIGYLVSGYQK.Q
			18	K.EVDGSKYWEQNQR.S
Ovalbumin related protein Y (P01014)	452	24	30	K.VYLPR.M
			16	K.VYLPRMK.I
			17	R.EMPFSMTK.E
			30	K.SMKVYLPR.M
			37	K.TINFDKLR.E
			35	R.QLINSWVEK.E
			17	K.KSMKVYLPR.M
			50	R.EWTSTNAMAK.K
			24	R.EWTSTNAMAKK.S
			51	K.FYTGGVEEVNFK.T
			29	R.KFYTGGVEEVNFK.T
			17	K.TINFDKLREWTSTNAMAK.S
			65	R.KFYTGGVEEVNFKTAAEEAR.Q
			(22)	R.KFYTGGVEEVNFKTAAEEAR.Q + Dea (NQ)
Mucin 5B (Q98UI9)	161	1		
			102	R.TATGAVEDSAAAFGNSWK.T
			(72)	R.TATGAVEDSAAAFGNSWK.T + Dea (NQ)
			(54)	R.TATGAVEDSAAAFGNSWKTR.A
			56	R.TATGAVEDSAAAFGNSWKTR.A + Dea (NQ)
Lysozyme C (P00698)	340	49	10	R.CELAAAMK.R
			35	K.RHGLDNYR.G
			27	K.GTDVQAWIR.G
			65	K.FESNFNTQATNR.N
			(42)	K.FESNFNTQATNR.N + Dea (NQ)
			(61)	K.FESNFNTQATNR.N + Dea (NQ)
			22	K.GTDVQAWIRGCRL.-
			58	K.IVSDGNGMNAWVAWR.N
			(58)	K.IVSDGNGMNAWVAWR.N + Dea (NQ)
			90	R.NTDGSTDYGILQINSR.W
			33	K.KIVSDGNGMNAWVAWR.N
Ovalbumin related protein X (P01013)	82	11	20	K.TINFEKLTEWTNPNTMEK.R
			30	K.TINFEKLTEWTNPNTMEKR.R
			22	K.TINFEKLTEWTNPNTMEKRR.V

Protein Tenp (O42273)	82	7	46 32	K.ITQVGSLYHEDLPITLSAALR.S K.ITQVGSLYHEDLPITLSAALRSSPR.V

[a] 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*Log(P), where P is the for deriving total probability that the observed peptide match is a random event.

[b]in brackets ion scores that Mascot did not consider protein score, but that we reported for the sake of completeness to show that a modified form of the same peptide was also observed.

Table IV.5. Proteins identified in the paint replicas containing red ochre and egg by LC-MSMS. Aliquots of paint replicas were treated in heterogeneous phase with trypsin with or without PNGaseF pretreatment and analysed by LC-MSMS. Proteins were identified searching UniprotSprot database with MSMS Ion search Mascot software (Matrix Science) with *Chordata* as taxonomy restriction, with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications. Only proteins identified with at least two peptides were considered as significative.

Sample	**Identified Protein (Accession number)**	**Protocol**			
		TRYPSIN		**PNGaseF + TRYPSIN**	
		score[a]	**n° of peptides**	**score[a]**	**n° of peptides**
WHOLE EGG and RED OCHRE	Ovalbumin (P01012)	37	2	496	10
	Ovotransferrin (P02789)	191	8	639	18
	Vitellogenin-2 (P02845)	1298	39	1769	52
	Vitellogenin-1 (P87498)	339	12	730	19
	Apolipoprotein B (P11682)	107	4	463	10
	Apolipoprotein A-I (P08250)	89	3	257	8
	Apovitellenin-1(P02659)	50	2	226	5
	Ovalbumin releted protein Y (P01014)			141	5
	Lysozyme (P00698) C			84	2
	Ovalbumin related protein X (P01013)			130	3
	Serum albumin (P19121)			203	5
	Ovocleidin-116 (F1NSM7)			91	3
	Cystatin (P01038)			75	2
	Dystrophin (P11533)			64	2
Sample	**Identified Protein (Accession number)**	**Protocol**			
		TRYPSIN		**PNGaseF + TRYPSIN**	
		score[a]	**n° of peptides**	**score[a]**	**n° of peptides**
YOLK and RED OCHRE	Vitellogenin-2 (P02845)	401	18	2671	62
	Vitellogenin-1 (P87498)	211	6	1617	35
	Apolipoprotein B (P11682)	52	2	402	11
	Ovalbumin (P01012)			323	6
	Apovitellenin-1 (P02659)			216	15
	Ovotransferrin (P02789)			151	4
	Lysozyme C (P00698)			139	2
	Serum albumin (P19121)			129	5
	Apolipoprotein A-1 (P08250)			95	2
	Vitellogenin-3 (Q91025)			51	2

[a] 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*Log(P), where P is the probability that the observed peptide match is a random event.

Table IV.6. Proteins identified in the paint replicas containing red ochre, casein, animal glue, and their combinations with egg by LC-MSMS. Aliquots of paint replicas were treated in heterogeneous phase with trypsin with or without PNGaseF pretreatment and analysed by LC-MSMS. Proteins were identified searching UniprotSprot database with MSMS Ion search Mascot software (Matrix Science) with Chordata as taxonomy restriction, with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications. Only proteins identified with at least two peptides were considered as significative.

Sample	Identified Protein (Accession number)	Protocol			
		TRYPSIN		PNGaseF + TRYPSIN	
		scorea	n° of peptides	scorea	n° of peptides
CASEIN and RED OCHRE	α -S1-casein (P02662)	678	15	666	16
	α -S2-casein (P02663)	502	14	659	15
	β -casein (P02666)	404	8	399	9
	β -lactoglobulin (P02754)	363	9	335	5
	k- casein (P02668)	260	5	307	8
	Xanthine dehydrogenase (P80457)	127	3	78	3
	Lactadherin (Q95114)	115	5		
	Fatty acid binding protein, heart (P10790)			72	2
	Complement C3 (Q2UVX4)	63	2	50	2
	β-1,4 galactosyltransferase (P08037)	59	2		
	Lipoprotein lipase (P11151)	45	2		
	Glycosylation-dependent cell adhesion molecule (P80195)			45	2
Sample	**Identified Protein (Accession number)**	**Protocol**			
		TRYPSIN		**PNGaseF + TRYPSIN**	
		scorea	**n° of peptides**	**scorea**	**n° of peptides**
COLLAGEN and RED OCHRE	Collagen alpha-1(I) chain (P02453)	101	6	602	13
	Collagen alpha-2(I) chain (P02465)	107	6	110	4
	Collagen alpha-1(III) chain (P04258)	78	3	88	3
	Collagen alpha-2(XI) chain (Q32S24)	45	2	335	5

Sample	Identified Protein (Accession number)	Protocol			
		TRYPSIN		PNGaseF + TRYPSIN	
		score[a]	n° of peptides	score[a]	n° of peptides
ALBUMEN, CASEIN and RED OCHRE	α -S1-casein (P02662)	190	7	465	11
	α -S2-casein (P02663)	220	7	342	11
	β -casein (P02666)	72	3	293	9
	k-casein (P02668)			207	3
	Ovalbumin (P01012)	38	2	358	8
	Ovotransferrin (P02789)			470	15
	Ovalbumin related protein Y (P01014)			92	4

Sample	Identified Protein (Accession number)	Protocol			
		TRYPSIN		PNGaseF + TRYPSIN	
		score[a]	n° of peptides	score[a]	n° of peptides
ALBUMEN, COLLAGEN and RED OCHRE	Collagen alpha-1(I) chain (P02453)	52	2	666	23
	Collagen alpha-2(I) chain (P02465)			366	14
	Ovalbumin (P01012)			193	7
	Ovotransferrin (P02789)			139	3

Sample	Identified Protein (Accession number)	Protocol			
		TRYPSIN		PNGaseF + TRYPSIN	
		score[a]	n° of peptides	score[a]	n° of peptides
CASEIN, COLLAGEN and RED OCHRE	α-S1-casein (P02662)	445	9	621	13
	α-S2-casein (P02663)	187	7	405	14
	β-casein (P02666)	117	4	228	7
	k-casein (P02668)	51	2	216	5
	β -lactoglobulin (P02754)	111	3	94	2
	Collagen alpha-1(I) chain (P02453)	1255	36	1247	35
	Collagen alpha-2(I) chain (P02465)	883	24	1142	32
	Collagen alpha-1(III) chain (P04258)			120	3

[a] 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*Log(P), where P is the probability that the observed peptide match is a random event.

Table IV.7. Proteins identified in the paint replicas containing different inorganic pigments and albumen by LC-MSMS. Aliquots of paint replicas were treated in heterogeneous phase with trypsin with or without PNGaseF pretreatment and analysed by LC-MSMS. Proteins were identified searching UniprotSprot database with MSMS Ion search Mascot software (Matrix Science) with Chordata as taxonomy restriction, with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications. Only proteins identified with at least two peptides were considered as significative.

Sample	Identified Protein (Accession number)	Protocol			
		TRYPSIN		PNGaseF + TRYPSIN	
ALBUMEN and CINNABAR		score[a]	n° of peptides	score[a]	n° of peptides
	Ovalbumin (P01012)	402	14	1148	40
	Ovotransferrin (P02789)	256	9	1420	41
	Ovalbumin releted protein Y (P01014)	83	4	697	22
	Ovalbumin releted protein X (P01013)	54	2	163	9
	Clusterin (Q9YGP0)	83	2	109	4
	Mucin 5B (Q98UI9)			201	4
	Lysozime C (P00698)			123	4
	Cystatin (P01038)			79	2
	Ovostatin (P20740)			128	4
Sample	**Identified Protein (Accession number)**	**Protocol**			
		TRYPSIN		**PNGaseF + TRYPSIN**	
ALBUMEN and MINIUM		**score[a]**	**n° of peptides**	**score[a]**	**n° of peptides**
	Ovalbumin (P01012)	232	8	388	10
	Ovotransferrin (P02789)	115	2	421	18
	Ovalbumin releted protein Y (P01014)			124	9
	Ovalbumin releted protein X (P01013)			57	3
	Clusterin (Q9YGP0)			42	2
Sample	**Identified Protein (Accession number)**	**Protocol**			
		TRYPSIN		**PNGaseF + TRYPSIN**	
ALBUMEN and AZURITE		**score[a]**	**n° of peptides**	**score[a]**	**n° of peptides**
	Ovalbumin (P01012)	210	8	445	10
	Ovotransferrin (P02789)	45	2	490	13
	Ovalbumin releted protein Y (P01014)	59	2	206	6
	Ovalbumin releted protein X (P01013)			73	3
	Lysozyme C (P00698)			162	2

[a] 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*Log(P), where P is the probability that the observed peptide match is a random event.

Historical samples. Finally, our improved protocol was tested on two samples collected from historical objects: sample Purple 12M080 from a mural painting from the urban district of

Cuma archeological site, and DSFL6 from a mural painting of San Frediano Church in Lucca, Italy (16th century). It is worth noting that sample DSFL6 was previously analyzed by GC-MS and the presence of egg was accordingly inferred [39]. Both samples provided no result when analyzed by the Trypsin alone protocol.

Aliquots of the two historical samples were analyzed following the newly developed procedure including PNGaseF pretreatment before trypsin digestion and LC-MSMS analysis of the resulting peptide mixtures. Table IV.8 reports the details of identifications of the proteins obtained in the two samples. A number of egg proteins were identified in both samples confirming the presence of an egg-containing binder and showing very clearly that deglycosylation of the samples before tryptic digestion led to the identification of egg proteins that could not be detected before. Moreover, it should be underlined that no proteins from albumen were identified in the sample from Cuma thus allowing us to confidently assess that only yolk was used as paint binder.

Table IV.8. Details of the identification of proteins in historical paint samples by LC-MSMS. An aliquot of the fresco fragment "Purple 12M080" from the urban district of Cuma archeological site, and an aliquot of the mural painting "Sacra Conversazione" (DSFL6) by Amico Aspertini 1506-1510 from the S. Frediano Basilica (Lucca), that had been treated in heterogeneous phase with the classical trypsin alone protocol, providing no protein identification, were subjected again to trypsin digestion after PNGaseF pretreatment and the resulting peptide mixtures analysed by LC-MSMS. Proteins were identified searching UniprotSprot database with MSMS Ion search Mascot software (Matrix Science) with Chordata as taxonomy restriction, with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications. Only proteins identified with at least two peptides were considered as significative.

Purple 12M080				
Identified protein (Accession number)	**Total score[a]**	**Sequence coverage (%)**	**Individual ion scores[b]**	**Matched sequence**
Vitellogenin-2 (P02845)	1163	24	13	K.ALQGWK.E
			29	R.IANQIR.N
			14	K.TFNEVK.F
			40	R.AAVSVEGK.M
			25	K.TVVEPADR.N
			24	R.YVIQEDR.K
			29	R.NIGELGVEK.R
			35	K.VSTELVTGR.F
			32	R.MVVALTSPR.T
			37	K.ELLQQVMK.T + Oxidation (M)
			52	K.TVQLAGVDSK.C
			64	R.QQLTLVEVR.S
			13	R.TMFPSAIISK.L + Oxidation (M)
			28	R.NIGELGVEKR.T
			22	K.MTPPLTGDFR.L
			(22)	K.MTPPLTGDFR.L + Oxidation (M)

			32	R.ADTYFDNYR.V
			18	R.VEGLADVIMKR.N + Oxidation (M)
			14	K.LEWPKVPSNVR.S
			16	R.NIPFAEYPTYK.Q
			67	R.IANADNLESIWR.Q
			(34)	R.LSQLLESTMQIR.S
			53	R.LSQLLESTMQIR.S + Oxidation (M)
			15	R.DASFIQNTYLHK.L
			33	R.VGATGEIFVVNSPR.T
			46	K.ILGQEVAFININK.E
			106	K.QSDSGTLITDVSSR.Q
			13	K.LEISGLPENAYLLK.V
			58	R.SPQVEEYNGVWPR.D
			26	K.ELPTETPLVSAYLK.I
			13	K.IVLMPVHTDADIDK.I
			15	R.KKPMDEEENDQVK.Q + Oxidation (M)
			35	R.WLLSAVSASGTTETLK.F
			32	K.LKQSDSGTLITDVSSR.Q
			35	K.EALQPIHDLADEAISR.G
			33	R.NSIAGQWTQPVWMGELR.Y
			(18)	R.NSIAGQWTQPVWMGELR.Y + Deamidated (NQ)
			36	R.QVYQISPFNEPTGVAVMEAR.Q
Vitellogenin-1 (P87498)	626	13	10	
			38	K.FIITTR.K
			14	K.ANLVDVTK.S
			29	K.QVAEVPPK.E
			36	K.LALIEVQK.Q
			24	K.ALLLSEIR.C
			18	R.NIEDLAASK.M
			21	K.KANLVDVTK.S
			40	K.NVNFDGEILK.V
			18	R.LTELLNSNVR.L
			34	K.ALGNVGHPASIK.H
			57	R.IAALATTGQMAR.K + Oxidation (M)
				K.SNIEEVLLALK.A
			19	
			57	K.ILDDTDNQATR.N
			77	K.VAGNVQAQITPSPR.S
			33	R.AANEENYESVWK.Q
			25	R.YLLDLLPAAASHR.S
			26	R.MSFKPVYSDVPIEK.I
			(26)	K.MTPVLLPEAVPDIMK.M
			(18)	K.MTPVLLPEAVPDIMK.M + Oxidation (M)
			12	K.MTPVLLPEAVPDIMK.M + 2 Oxidation (M)
			27	R.GYSPDKDWETNYDFR.E
				R.VADPIEVGIAAEGLQEMFVR.G + Deamidated (NQ); Oxidation (M)
Apolipoprotein B (Fragment) (P11682)	107	13	12	
			26	R.YWSAVAK.R
			52	R.VPASETILR.G
			9	R.QTFVTLQEADFAGK.L
				K.YTGEELYLMTTEK.A + Oxidation (M)
Apovitellenin-1 (P02659)	90	16	58	
			33	K.LAEQLMEK.I + Oxidation (M)
				R.NFLINETAR.L

DSFL6				
Identified protein (Accession number)	**Total score[a]**	**Sequence coverage (%)**	**Individual ion scores**	**Matched sequence**
Ovalbumin	20	5	20	R.GGLEPINFQTAADQAR.E
(P01012)			13	R.LYAEER.Y

[a] 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*Log(P), where P is the probability that the observed peptide match is a random event.
[b]in brackets ion scores that Mascot did not consider for deriving total protein score, but that we reported for the sake of completeness to show that.

IV.3.5 Bioinformatics tools for deterioration detection

While much effort has been already devoted to the development of identification tools, a lot of work has still to be done to understand and characterize the whole range of modifications occurring upon aging on deteriorating proteins in samples that have been exposed to a wide spectrum of different environmental conditions, thus contributing to what can be called the field of paleo proteomics.

The modifications that occur on proteins in natural environment over time are still under investigation and delineating both *in vivo* and diagenetically derived alterations will provide important information on the physiology and/or phylogenies of organisms, as well as on the aging mechanisms. These can be function of specific environmental factors, and their knowledge will lead to a more conscious preservation of ancient samples.

Moreover, the intrinsic damaged nature of the ancient proteins can intuitively be expected to be the first evidence of authenticity of a sample, that ruling out possible contaminations from modern materials [42].

A clever use of the "open mass" modification search utility in the Protein Prospector database search, as recently suggested by Hill *et al* [12], in the analysis of fossil bones, allowed to identify extensive surviving galactosylation and glucosyl-galactosylation of hydroxylysine residues in collagen. This bioinformatics strategy of analysis offers the potentiality to detect unexpected modifications and certainly will hold interesting prospects in the characterization of molecular details of degradation processes.

Error tolerant searches can also be used to detect diagenetic modifications [43], among which the observation of semi-tryptic peptides accounting for partial hydrolysis of the polypeptidic chain within the samples [7,12]. High occurrence of partial hydrolysis is indeed expected as degradation effect. Table IV.9 reports a comparison of the results that we obtained on several pictorial models and on a sample from the Camposanto Monumentale in Pisa when identification is carried out, on the same set of experimental data, with Mascot search program allowing for semi-tryptic cleavages and only for specific tryptic cleavages: the abundance of non-tryptic termini suggests hydrolysis as a likely and expected effect of protein degradation.

Table IV.9. Proteins identified in the paint replica containing different inorganic pigments mixed with milk as binder, and in the sample from the Camposanto Monumentale of Pisa, by trypsin digestion in heterogeneous phase and LC-MS/MS analysis. Proteins were identified in the UniProt database with Mascot MS/MS Ion Search software, with *Chordata* as taxonomic restriction, with methionine oxidation, formation of pyroGlu at the N-terminus of Gln, and deamidation (N, Q) of peptides as variable modifications. Individual ion scores >31 indicate identity or extensive homology. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits (http://www.matrixscience.com/help/interpretation_help.html).

Sample	Protein (UniProt Accession number)	Tryptic identification			Semitryptic identification		
		Sequence coverage (%)	Protein score	n° of peptides	Sequence Coverage (%)	Protein score	n° of peptides (specific tryptic peptides)
Casein and minium	alpha-S1 casein (P02662)	76	641	16	77	1496	27 (16)
	alpha-S2 casein (P02663)	56	512	17	56	603	21 (17)
	beta-casein (P02666)	53	426	9	69	1104	14 (9)
	kappa-casein (P02668)	24	242	5	50	520	8 (5)
Casein and cinnabar	alpha-S1 casein (P02662)	69	666	15	69	1718	31 (15)
	alpha-S2 casein (P02663)	52	530	12	53	792	14 (12)
	beta-casein (P02666)	79	465	12	79	1070	21 (12)
	kappa-casein (P02668)	46	345	6	53	728	8 (6)
Casein and $CaCO_3$	alpha-S1 casein (P02662)	64	483	12	64	1072	21 (12)
	alpha-S2 casein (P02663)	52	356	6	52	453	13 (6)
	beta-casein (P02666)	47	266	7	61	732	21 (7)
	kappa-casein (P02668)	26	214	3	47	363	6 (3)
Casein	alpha-S1 casein (P02662)	71	589	14	72	1339	23 (14)
	alpha-S2 casein (P02663)	53	512	14	53	560	18 (14)
	beta-casein (P02666)	79	445	12	79	1058	19 (12)
	kappa-casein (P02668)	44	272	7	53	567	9 (7)
Sample from Camposanto Monumentale	alpha-S1 casein (P02662)	50	565	11	56	1377	31 (11)
	alpha-S2 casein (P02663)	27	252	7	29	374	10 (7)
	beta-casein (P02666)	71	432	8	72	1601	32 (8)
	kappa-casein (P02668)	22	141	3	46	391	10 (3)
	Collagen alpha-1(I) chain (P02453)	36	854	25	42	879	28 (25)
	Collagen alpha-2(I) chain (P02465)	23	522	21	25	558	22 (21)

IV.4 Conclusions

The physical state of the samples, enormously different from the natural environment of proteins, and the degradation processes undergone during ageing, provide unusual challenges that require to be counteracted by specific adaptations of the classically adopted protocols used in the analysis of biological samples. Experiments to improve and develop new protocol have been described in this chapter. A denaturing step with 6M urea was used before trypsin digestion to "open" residual structural elements in proteins in the perspective that flexibility of the polypeptide substrate is an absolute requirement for the protease to properly hydrolyze peptide bonds. Results obtained in comparison with minimally invasive approach and the analysis carried out clearly indicate that the denaturating pre-treatment with urea improves the quality of the identifications. Moreover, as discussed in the next chapter, this protocol was extensively used for analytical investigation of historical samples of cultural heritage with excellent results.

Egg proteins can be easily identified when occurring in solution with glycosylation not dramatically hampering the access of proteases to proteolytic linkages. However, when egg is used as binder in paintings, in a solid state mixed with pigments, the oligosaccharide moieties seem to create a molecular hindrance that greatly reduces the accessibility of proteases, thus impairing the identification process. The implementation of a deglycosylation step in the analysis of paint samples prior to the tryptic digestion has proven to significantly improve the number of identified peptides from egg proteins in several different paint reconstructions as well as on two historical samples of completely different origin. Moreover, it was shown that the same glycosylation step does not affect the capability of correctly identifying other proteinaceous paint binders, such as milk and animal glue.

The protocol developed to improve the identification egg based binders, based on a sample pretreatment with PNGase F, has revealed to be useful, reliable, cost effective and sensitive enough to cope with the small amounts of degraded proteins that can be found in samples from artworks.

Moreover, it was demonstrated that a clever use of bioinformatics tools such as operating Mascot in an error tolerant search mode, can allow to identify degradation processes such as random proteolysis.

References

1. Vinciguerra, R., De Chiaro, A., Pucci, P., Marino, G. & Birolo, L. Proteomic strategies for cultural heritage: form bones to paintings. *Microchem. J.* **126,** 341–348 (2015).
2. Hynekl, R., Kuckova, S., Hradilova, J. & Kodicek, M. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry as a tool for fast identification of protein binders in color layers of paintings. *Rapid Commun. Mass Spectrom.* **18,** 1896–1900 (2004).
3. Tokarski, C., Martin, E., Rolando, C., Post-traductionnelles, M. & Cedex, A. Identification of Proteins in Renaissance Paintings by Proteomics John the Baptist , St . Sebastian (XVth century), and. *Analysis* **78,** 1494–1502 (2006).
4. Leo, G. *et al.* Proteomic strategies for the identification of proteinaceous binders in paintings.
5. Ranalli, G. *et al.* Biotechnology applied to cultural heritage: Biorestoration of frescoes using viable bacterial cells and enzymes. *J. Appl. Microbiol.* **98,** 73–83 (2005).
6. Hong, C. *et al.* Identification of milk component in ancient food residue by proteomics. *PLoS One* **7,** 1–7 (2012).
7. Warinner, C. *et al.* Direct evidence of milk consumption from ancient human dental calculus. *Sci. Rep.* **4,** 7104 (2014).
8. Rybczynski, N. *et al.* Mid-Pliocene warm-period deposits in the High Arctic yield insight into camel evolution. *Nat. Commun.* **4,** 1550 (2013).
9. Buckley, M. A Molecular Phylogeny of Plesiorycteropus Reassigns the Extinct Mammalian Order 'Bibymalagasia'. *PLoS One* **8,** (2013).
10. Welker, F. *et al.* Ancient proteins resolve the evolutionary history of Darwin/'s South American ungulates. *Nature* **522,** 81–84 (2015).
11. Villa, P. *et al.* A Milk and Ochre Paint Mixture Used 49,000 Years Ago at Sibudu, South Africa. *PLoS One* **10,** e0131273 (2015).
12. Hill, R. C. *et al.* Preserved Proteins from Extinct Bison Latifrons Identified by Tandem Mass Spectrometry; Hydroxylysine Glycosides are a Common Feature of Ancient Collagen. *MCP Pap. Press. Publ. May* **6,** (2015).
13. Marchant, J. Mystery of Darwin's 'strange animals' solved. *Nature* 1–3 (2015). doi:10.1038/nature.2015.17138
14. Wadsworth, C. & Buckley, M. Proteome degradation in fossils: Investigating the longevity of protein survival in ancient bone. *Rapid Commun. Mass Spectrom.* **28,** 605–615 (2014).
15. Van Doorn, N. L., Wilson, J., Hollund, H., Soressi, M. & Collins, M. J. Site-specific deamidation of glutamine: A new marker of bone collagen deterioration. *Rapid Commun. Mass Spectrom.* **26,** 2319–2327 (2012).
16. Leo, G. *et al.* Deamidation at asparagine and glutamine As a major modification upon deterioration/aging of proteinaceous binders in mural paintings. *Anal. Chem.* **83,** 2056–2064 (2011).
17. Wilson, J., Van Doorn, N. L. & Collins, M. J. Assessing the extent of bone degradation using glutamine deamidation in collagen. *Anal. Chem.* **84,** 9041–9048 (2012).
18. Buckley, M., Melton, N. D. & Montgomery, J. Proteomics analysis of ancient food vessel stitching reveals >4000-year-old milk protein. *Rapid Commun. Mass Spectrom.* **27,** 531–538 (2013).
19. Ma, B. *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **17,** 2337–2342 (2003).
20. Frank, A. & Pevzner, P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **77,** 964–73 (2005).

21. Fischer, B. *et al.* A Hidden Markov Model for de Novo Peptide Sequencing. *Anal. Chem.* **77,** 7265–7273 (2005).
22. Taylor, J. A. & Johnson, R. S. Sequence database searches via de Novo peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **11,** 1067–1075 (1997).
23. Waters, M. R. *et al.* Pre-Clovis Mastodon Hunting 13,800 Years Ago at the Manis Site, Washington. *Sci.* **334 ,** 351–353 (2011).
24. Amoresano, A. *et al.* Structural characterization of the oligosaccharide chains of human ??1-microglobulin from urine and amniotic fluid. *Eur. J. Biochem.* **267,** 2105–2112 (2000).
25. Duce, C. *et al.* Physico-chemical characterization of protein-pigment interactions in tempera paint reconstructions: Casein/cinnabar and albumin/cinnabar. *Anal. Bioanal. Chem.* **402,** 2183–2193 (2012).
26. Gambino, M. *et al.* A simple and reliable methodology to detect egg white in art samples. *J. Biosci.* **38,** 397–408 (2013).
27. Rasmussen, K. L. *et al.* Reply to Ira Rabin's Comment on our paper Rasmussen etal. (2012). *J. Archaeol. Sci.* **43,** 155–158 (2014).
28. Jiang, X. *et al.* Method development of efficient protein extraction in bone tissue for proteome analysis. *J. Proteome Res.* **6,** 2287–2294 (2007).
29. Cleland, T. P., Voegele, K. & Schweitzer, M. H. Empirical evaluation of bone extraction protocols. *PLoS One* **7,** 1–9 (2012).
30. Solazzo, C. *et al.* Modeling Deamidation in Sheep α Keratin Peptides and Application to Archeological Wool Textiles. (2014).
31. Catherine Guérin-Dubiard *, † *et al.* Proteomic Analysis of Hen Egg White. *J. Agric. Food Chem.* **54,** 3901–3910 (2006).
32. Jiang, K. *et al.* Comparison of Chicken and Pheasant Ovotransferrin N-Glycoforms via Electrospray Ionization Mass Spectrometry and Liquid Chromatography Coupled with Mass Spectrometry. *J. Agric. Food Chem.* **62,** 7245–7254 (2014).
33. Bailey, U.-M. & Schulz, B. L. Deglycosylation systematically improves N-glycoprotein identification in liquid chromatography–tandem mass spectrometry proteomics for analysis of cell wall stress responses in Saccharomyces cerevisiae lacking Alg3p. *J. Chromatogr. B* **923–924,** 16–21 (2013).
34. O'Neill, R. A. Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis. *J. Chromatogr. A* **720,** 201–215 (1996).
35. Zhang, W., Wang, H., Zhang, L., Yao, J. & Yang, P. Large-scale assignment of N-glycosylation sites using complementary enzymatic deglycosylation. *Talanta* **85,** 499–505 (2011).
36. Tarentino, a L., Gómez, C. M. & Plummer, T. H. Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24,** 4665–71 (1985).
37. Chu, F. K. Requirements of cleavage of high mannose oligosaccharides in glycoproteins by peptide N-glycosidase F. *J. Biol. Chem.* **261,** 172–177 (1986).
38. Longobardi, S. *et al.* A simple MALDI plate functionalization by Vmh2 hydrophobin for serial multi-enzymatic protein digestions. *Anal. Bioanal. Chem.* **407,** 487–496 (2015).
39. Bonaduce, I., Colombini, M. P. & Diring, S. Identification of garlic in old gildings by gas chromatography-mass spectrometry. *J. Chromatogr. A* **1107,** 226–232 (2006).
40. De La Cruz-Cañizares, J., Doménech-Carbó, M. T., Gimeno-Adelantado, J. V., Mateo-Castro, R. & Bosch-Reig, F. Suppression of pigment interference in the gas chromatographic analysis of proteinaceous binding media in paintings with EDTA. *J. Chromatogr. A* **1025,** 277–285 (2004).
41. Duce, C. *et al.* Interactions between inorganic pigments and proteinaceous binders in reference

paint reconstructions. *Dalt. Trans.* **42,** 5975–5984 (2013).
42. Buckley, M. *et al.* Comment on ‘Protein Sequences from Mastodon and Tyrannosaurus rex Revealed by Mass Spectrometry’. *Science (80-.).* **319,** 33 (2008).
43. Cappellini, E. *et al.* Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J. Proteome Res.* **11,** 917–926 (2012).

V

Application of develop strategy to case studies

Investigating the cultural heritage objects does not only allow us to get isolated data on those invaluable works of art or archaeological finds, but contributes at assembling a big "puzzle" to understand our past and the development of our culture, implementing and supporting written sources, stylistic and anthropologic studies with molecular data.

Moreover, the knowledge of protein-based materials is essential to design a long-term preservation of cultural heritage objects, to assess the best conservation and display conditions, to prevent and slow down the decay processes, and to plan the best restoration procedures.

In this chapter, the application of developed proteomics-based techniques and mass spectrometry analysis approaches to specific cases, described in the previous chapter, are shown. The aim is to show how these methods can be employed, either separately or in combination with other analytical techniques, to solve concrete problems. Among these case studies, both investigations that aim to gain information of the provenance of cultural artefacts as well as studies that seek to evaluate and monitor preservation/restoration treatments have been included.

The strategies described in the previous chapter and that are based on urea pre-treatment of the samples, was successful applied to the chemical investigation of the organic paint binders that have been used in the decoration of the Giant Buddhas of Bāmiyān. The analysis was carried out on a selection of rescued fragments in collaboration with ICOMOS (International Council on Monuments and Sites) and the University of Pisa, under aegis of UNESCO. In particular, the proteomics approach revealed to be able to identify the source of the milk proteins present in the restoration layers, despite their age and degradation. More in detail, cow and goat milk were both identified, in agreement with the documented presence of rich pastures in the Bāmiyān valley at the time when the historical restorations were carried out.

The chemical analysis of a residue on a stone flake from a 49,000 years-old level of Sibudu (South Africa) was performed showing a mixture of ochre, casein from milk and collagen likely obtained by killing a wild bovid lactating female. In antiquity, the mixture of pigment with milk as a binder was an art medium known as tempera. Ochre use is well documented in

Middle Stone Age South African sites but there was no evidence of a binder so far. The results can be interpreted as a precursor to rock art painting, documented in Southern Africa at 27, 500 years BP, or to the art of body painting and in rock art images. In either cases, it shows that toward the end of the Middle Stone Age, forms of symbolic communication were evolving that later become key elements of the San culture and social life.

As a part of multi-methodological characterization of wall paintings from a Roman Archaeological site in Cuma, a pre-treatment step with PNGaseF, as described in the previous chapter, was applied and allowed us to extract information about protein-based binders; while a GC-MS analysis, applied in collaboration with Prof. Amoresano from the Department of Chemical Sciences (University of Naples 'Federico II') revealed the presence of organic-based binders. The research was focused on the characterization of the differences and the similarities between a private indoor (*domus*) and a public outdoor fabrication (*temple*) in the archaeological excavation of Cuma. Both pigments, binders and mortars were studied via a combination of destructive/microdestructutive (mass spectrometry, ionic chromatography, ICP-based techniques) and non-destructive (Raman microscopy, X-ray diffraction and SANS) methodologies in a multidisciplinary approach at Department of Chemical Sciences and the Department of Humanities in collaboration with Prof. Giovanna Greco.

Particularly, the systematic presence of dolomite only in temple mortars suggests an intentional use of such limestone for the outdoor fabrication of public interest. Differences between *temple* and *domus* extend to binder composition is discussed: proteinaceous binders in *domus* (egg and animal glue), whereas in TCP drying oils were detected.

V.1 The Giant Buddhas of Bāmiyān valley – Afghanistan

V.1.1 Introduction

The Buddha statues of Bāmiyān, outstanding representation of the Gandharan School in the Central Asian region, became well known in the West when it was too late to save them. Disregarding international protests, the Taliban destroyed the statues during the war in Afghanistan [1].

After their destruction in March 2001, activities of an ICOMOS mission started in July 2002 to safeguard the remains and support conservation in the Bāmiyān valley. In 2003, Bāmiyān was included in the World Heritage List and contemporarily placed in the List of World Heritage in Danger. The ICOMOS mission focused in a first moment on rescuing and storing the fragments of the figures, combined with the demining of the area, and the stabilisation of the niches which were pervaded by cracks due to the force of the explosion [2]. These processes are now concluded, but there are still disputes among the international community on how to proceed with the conservation of the Giant Buddhas. Whatever will be the final decision on how to best tribute the destroyed Buddhas, the knowledge of the technology and the materials used for their realisation represents a unique opportunity to investigate the original appearance of the statues and their manufacturing techniques.

The giant standing Buddhas of the Bāmiyān valley, 35 and 55 m in height, were the two most prominent clay statues in the world (Figure VI.1.1 show one of the statues). On the basis of the AMS ^{14}C dating of the plant material the Eastern and Western Buddha were built in the period 544 to 595 AD and 591 and 644 AD, respectively [3]. The earliest description of one Buddha dates back to the travelogue *xi* you *ji* ("The journey to the West") by the Chinese monk Xuanzang who came to this region in about 630 AD. More accounts of the statues appear in text by Persian and Arab authors of the medieval times when Bāmiyān was already Islamic and the Buddhist origin of the "idols" was forgotten. The statues became well known only in the 19th century when a larger number of Europeans reached Bāmiyān as travellers, adventurers or members of military campaigns. The 20th century brought a scientific approach: under the conduct of the Délegation Archéologique Française en Afghanistan (DAFA) examinations were carried out in the 1920's and 1930's. From 1969 to 1978 the Archaeological Survey of India (ASI) in cooperation with the Afghans carried out a comprehensive restoration work on the statues.

The giant Buddhas of Bāmiyān were fundamental pieces of art history not only because of their colossal dimensions, but also because they were precursors, inspiration and models determining the style of clay sculptures in farther Asia [4]. Located on the Silk Road, and being geographically the westernmost of the Buddhist sites, Bāmiyān was in position to receive influences not only from India but also from Sassanian and Byzantine artistic tradition and transmit them on to East Asia [5]. This means that from the historic-artistic point of view Bāmiyān is a key knot to study the west-east influences in ancient Central Asia as well as the evolution of Buddhist art through the centuries.

A **B** **C**

Figure.V.1.1 A photo of one of the two statues (1978) (A) and a more recent photo (2001) with the part of the mountain in the Bāmiyān valley where the statue was previously located (B). In C, it is possible to appreciate in detail the shallow part of the intact mantle carved and decorated, from which are derived samples analyzed.

In this chapter we present our part in the research dedicated at the establishing the constituting materials of Giant Buddhas of Bāmiyān and the technique of their realisation. Although the most recent photographs of the Giant Buddhas depict them as stone-coloured, once the statues appeared in bright colours surrounded by niches full of colourful, richly detailed murals. In order to apply the pigments – a fine powder of inorganic or organic coloured material – on the surface of these sculptures, an organic binder must have been used [6]. The organic binder is a film-forming material able to disperse the pigment particles, to ensure their adhesion onto the substrate and their cohesion within the paint film. The identification of organic paint constituents in micro-samples from works of art is very challenging from an analytical point of view [7], as they are aged and present in relatively low amount. Moreover, several organic

and inorganic substances are often simultaneously present in a layered structure, and can give rise to strong interactions between each other [8–12].

Here we present only the results of the characterisation of protein-based binders, which was achieved by using a proteomic approach based on the use of liquid chromatography tandem mass spectrometry LC-MS/MS. A combination with gas chromatography mass spectrometry (GC-MS) analysis, carried out at the research group of Prof. M.P. Colombini, at the Department of Chemistry and Industrial Chemistry, University of Pisa, allowed us to acquire invaluable information about the materials and techniques used, supporting the preservation of the statues' memory, and improving our understanding of historical livestock, and the interchange of cultures and technical know-how of Central Asia between the 6th and 8th centuries AD.

V.1.2 Material and Methods

V.1.2.1 Samples

Between 2004 and 2008 about 10000 fragments were recovered from the rubble at the feet of the statues. The paint layers were investigated on 275 small pieces of painted clay of unknown origin, mostly between 1 and 4 cm in length. Fragments show the remnants of several superimposed thick and slightly powdery paint layers meaning that the statues have been repainted over the centuries. Both statues show the same layer sequence. Underneath the pigmented paint layers, often a preparation layer can be observed which seems to have penetrated into the clay surface. In some samples a white non coherent priming layer is found. As far as the pigmented paint layers are concerned, it has been possible to ascribe them to three different historical moments: the lowest is the most antique and probably original, and the following two belong to two different overpaintings performed in following eras. On top of the paint layers, almost all samples show a light ochre layer that can be attributed to the Indo-Afghan restoration. Organic paint materials were investigated on 19 of these fragments selected as representative of the different paint layer structures individuated [13].

V.1.2.2 Proteomic analyses

Reagents. Ammonium hydrogen carbonate (Ambic), urea, TPCK-treated trypsin, KOH ethanol, trifluoroacetic acid, N,O-Bis(trimethylsilyl)trifluoroacetamide and ammonia were from Sigma; Formic acid and Acetonitrile (ACN) were purchased from Baker; deionized water was obtained from Millipore cartridge equipment.

Sample treatment. Proteinaceous material was identified following a minimally invasive proteomic analytical procedure as described by Leo et al. (2009), except for a pre-treatment of the sample in strongly protein denaturing conditions (6M Urea), as described in the previous chapter, that was introduced in order to favour the exposure of the proteinaceous material to the action of proteases, as also reported in (Gambino et al., 2013). Briefly, microsamples (ca 100-500 µg) were incubated for 1 h in 20 µL in 6 M urea followed by sonication for 30 min at room temperature. The samples were then 6-fold diluted with ammonium bicarbonate 10 mM pH 7.5 and enzymatic digestion carried out by 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), 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, 0.1 % Formic acid in Milli-Q water and analysed by LC-MS/MS. The eluate was analyzed by LC-MSMS 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 400nL/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 50min.

Peptide analysis was performed using 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. Each LC-MSMS analysis was preceded and followed by blank runs to avoid carryover contamination.

Double and triple charged ions were preferably isolated and fragmented. The acquired MS/MS spectra were transformed in mzData (.XML) format and used for protein identification with a licensed version of MASCOT software (www.matrixscience.com) version 2.4. with 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, formation of pyroglutammic acid from glutamine residues at the N-terminal position of peptides, and deamidation at asparagines and glutamines were considered as variable modifications (Vinciguerra *et al*., 2016) to query SwissProt and NCBI databases, without any

taxonomy restriction. Only proteins presenting two or more peptides were considered as positively identified.

Selected Ion Monitoring (SIM) mode analysis was used to improve the quality of the fragmentation of selected ions. Protein sequences were aligned using the Align tool available online at UniProt [14], which uses Clustal-Omega program [15].

V.1.3 Results

To characterise the proteinaceous binders, a selection of 30 sub-samples (obtained from 11 different fragments) were analysed by proteomics procedures.

The samples were subject to enzymatic digestion with trypsin, and the obtained mixtures of peptides were purified and concentrated using reverse-phase C18 pipette tips and analysed by LC-MS/MS [4]. Proteins were confidently identified in 16 out of the 30 sub-samples analysed (Table V.1.1, and details of the identifications are given in Table V.1.2).

The presence of several milk proteins, i.e. Alpha S1 casein, Alpha S2 casein, Beta casein and Beta lactoglobulin, allowed us to confidently assess that milk was extensively used. Alpha S1 casein was identified in almost all the samples where a positive identification was observed, namely in 15 out of the 16 positive samples, while Alpha S2 casein was identified in 6 samples, Beta casein in 7, and Beta lactoglobulin in 4 samples (Table V.1.1).

Table V.1.1. Milk proteins identified in the samples from Buddhas of Bāmiyān by LC-MSMS. Proteins were identified searching UniprotSprot database with MSMS Ion search Mascot software (Matrix Science) with Chordata as taxonomy restriction, with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications. Only proteins identified with at least two peptides were considered as significative. Details of the identifications are reported in Table V.1.2. (Appendix A)

Sample / Protein	214-clay	206-4-3	206-6-7-5-4	188	235-bulk	235-grey-blue	235-grey-blue II	235-black	235-clay1	214-sup	214-blue	18-2	18-clay	22-2	16—b	16-a
Alpha-S1-casein	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Alpha-S2-casein	X		X		X		X			X	X					
Beta-casein	X				X		X			X	X	X				X
Beta-lactoglobulin	X				X					X	X					

Most interestingly, proteomic analyses allowed to establish the coexistence of milk from cow and from goat. In fact, we identified four possible sources for the milk (*Bos taurus*, *Bubalis bubalis*, *Capra hircus* and *Ovis aries*) (Table V.1.2 in Appendix A), provided the high

percentage of similarity between the homologous proteins from different sources that results in multiple possible assignments with very similar scores. Table V.1.3 (Appendix A) reports the comprehensive list of the peptides from milk proteins that have been identified in the whole set of samples. Some peptides are in common between proteins from different organisms, while some are unique for a single sequence, thus allowing species identification. Let's focus our attention on the peptides of alpha S1 and alpha S2 caseins (Table V.1.4), since these are the proteins that were identified in the largest number of cases with high confidence.

Table V.1.4. Peptides of alpha S1-casein and alpha S2-casein that have been identified. The comprehensive list of the matched peptides that have been observed in the whole set of samples. The position of the starting and the ending residues of each peptide are indicated as superscript. In bold the sequences that are unique to the species; in brackets the accession number in the UniProt database.

Protein	*Bos taurus* (P02662)	*Capra hircus* (P18626)
Alpha-S1-casein	^{99}EDVPSER^{105} ^{106}YLGYLEQLLR^{115} ^{107}LGYLEQLLR^{115} ^{19}**HQGLPQEVLNENLLR**^{37} ^{29}**EVLNENLLR**^{37} ^{38}**FFVAPFPEVFGK**^{49} ^{50}**EKVNELSK**^{57} ^{121}**VPQLEIVPNSAEER**^{134}	^{99}EDVPSER^{105} ^{106}YLGYLEQLLR^{115} ^{107}LGYLEQLLR^{115} ^{38}**FVVAPFPEVFR**^{48}
Alpha-S2-casein	***Bos taurus* (P02663)**	***Capra hircus* (P33049)**
	^{96}ALNEINQFYQK^{106} ^{141}EQLSTSEENSK^{151} ^{130}**NAVPITPTLNR**^{140} ^{153}**TVDMESTEVFTK**^{164} ^{189}**FALPQYLK**^{196} ^{215}**VIPYVRYL**^{222}	^{96}ALNEINQFYQK^{106} ^{141}EQLSTSEENSK^{151} ^{62}**NANEEEYSIR**^{72} ^{130}**NAGPFTPTVNR**^{140} ^{213}**TNAIPYVR**^{220}

The peptide FFVAPFPEVFGK of alpha S1 casein was detected in 15 samples and it is a unique peptide for bovine alpha S1 casein (either from *Bos taurus* (P02662) or from *Bubalus bubalis* (O62823)). Similarly, the peptide FVVAPFPEVFR is unique of caprine alpha S1 casein (either from *Capra hircus* (P18626) or from *Ovis aries* (P04653)). The concomitant identification of peptides that are unique to either *Bos taurus* and *Capra hircus* homologous proteins allowed to suggest that both cattle and goat milk were used. Most interestingly, in the alignment of the sequences of the four alpha S1 caseins considered above (Fig. V.1.2), these two sequences are aligned, and their compresence in the sample indicates the coexistence of the homologous proteins in mixture.

```
1    MKLLILTCLVAVALARPKHPIKHQGLPQEVLNENLLRFFVAPFPEVFGKEKVNELSKDIG   60  P02662   CASA1_BOVIN
1    MKLLILTCLVAVALARPKQPIKHQGLPQGVLNENLLRFFVAPFPEVFGKEKVNELSTDIG   60  O62823   CASA1_BUBBU
1    MKLLILTCLVAVALARPKHPINHRGLSPEVPNENLLRFVVAPFPEVFRKENINELSKDIG   60  P18626   CASA1_CAPHI
1    MKLLILTCLVAVALARPKHPIKHQGLSSEVLNENLLRFVVAPFPEVFRKENINELSKDIG   60  P04653   CASA1_SHEEP
     ******************:**:*:**   * *******.******** **::****.***

61   SESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYK  120  P02662   CASA1_BOVIN
61   SESTEDQAMEDIKQMEAESISSSEEIVPISVEQKHIQKEDVPSERYLGYLEQLLRLKKYN  120  O62823   CASA1_BUBBU
61   SESTEDQAMEDAKQMKAGSSSSSEEIVPNSAEQKYIQKEDVPSERYLGYLEQLLRLKKYN  120  P18626   CASA1_CAPHI
61   SESIEDQAMEDAKQMKAGSSSSSEEIVPNSAEQKYIQKEDVPSERYLGYLEQLLRLKKYN  120  P04653   CASA1_SHEEP
     *** ******* ***:* * ******** *.***:************************:

121  VPQLEIVPNSAEERLHSMKEGIHAQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWY  180  P02662   CASA1_BOVIN
121  VPQLEIVPNLAEEQLHSMKEGIHAQQKEPMIGVNQELAYFYPQLFRQFYQLDAYPSGAWY  180  O62823   CASA1_BUBBU
121  VPQLEIVPKSAEEQLHSMKEGNPAHQKQPMIAVNQELAYFYPQLFRQFYQLDAYPSGAWY  180  P18626   CASA1_CAPHI
121  VPQLEIVPKSAEEQLHSMKEGNPAHQKQPMIAVNQELAYFYPQLFRQFYQLDAYPSGAWY  180  P04653   CASA1_SHEEP
     ********: ***:*******  *:**:***.**********:*****************

181  YVPLGTQYTDAPSFSDIPNPIGSENSEKTTMPLW  214  P02662   CASA1_BOVIN
181  YVPLGTQYPDAPSFSDIPNPIGSENSGKTTMPLW  214  O62823   CASA1_BUBBU
181  YLPLGTQYTDAPSFSDIPNPIGSENSGKTTMPLW  214  P18626   CASA1_CAPHI
181  YLPLGTQYTDAPSFSDIPNPIGSENSGKITMPLW  214  P04653   CASA1_SHEEP
     *:****** ***************** * *****
```

Figure V.1.2. Alignment of the sequences of Alpha S1 caseins. The aminoacidic sequences from *Bos taurus* (P02662), *Bubalus bubalis* (O62823), *Capra hircus* (P18626) and *Ovis aries* (P04653). The experimentally identified peptides are highlighted in grey. Strikethrough sequence is the signal peptide that is absent in the mature form of the protein. Protein sequences were aligned using Align tool available online at UniProt [14]which uses Clustal-Omega program [15]. * (asterisk) indicates positions which have a single, fully conserved residue; (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix; . (period) indicates conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix.

As a confirm of the effective presence of the two unique peptides in the same peptide mixture, we selectively searched them by running selected samples in SIM mode, and manually interpreting the MS/MS spectra. As an example, Figure V.1.3 reports the fragmentation spectra of the ions that were interpreted as peptide ^{38}FFVAPFPEVFGK^{49} of alpha S1 casein from *Bos taurus* (P02662, UniProt entry) and peptide ^{38}FVVAPFPEVFR^{48} of alpha S1 casein from *Capra hircus* (P18626), respectively, as detected in the LC-MS/MS analysis in one of the samples. Both spectra can be manually inspected and unambiguously interpreted, thus confirming the assignments to the peptides inferred above.

Similar reasoning was carried out for the peptide ^{130}NAVPITPTLNR^{140} from alpha S2 casein from *Bos taurus* (P02663) and peptide ^{130}NAGPFTPTVNR^{140} from alpha S2 casein from *Capra hircus* (P33049) that are unique for the two protein sequences, respectively, and are aligned in the multiple alignment of the identified alpha S2 caseins (Fig. V.1.4), thus

confirming the co-presence of milk from cow and goat as assumed on the basis of the peptides of alpha S1 casein reported above.

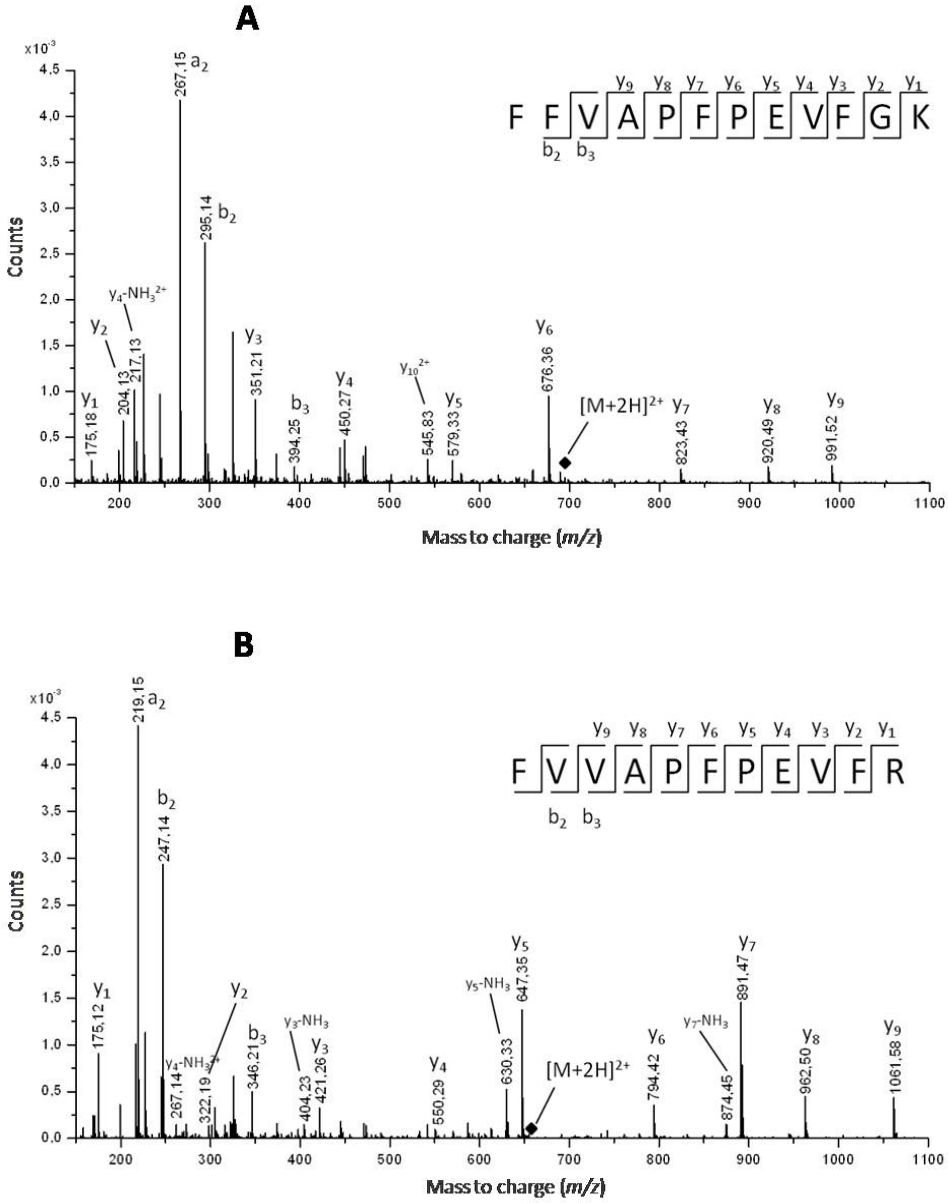


Figure V.1.3. MS/MS spectra acquired in SIM analysis mode. The doubly charged ions at *m/z* 692.87 (A) of the peptide ^{38}FFVAPFPEVFGK^{49} of alpha S1 casein from *Bos taurus* (P02662) and 654.58 (B) of the peptide ^{38}FVVAPFPEVFR^{48} of alpha S1 casein from *Capra hircus* (P18626), respectively, that have identified in the LC-MS/MS analysis of the sample 214-clay.

```
1    MKFFIFTCLLAVALAKNTMEHVSSSEESI-ISQETYKQEKNMAINPSKENLCSTFCKEVV   59  P02663   CASA2_BOVIN
1    MKFFIFTCLLAVALAKHKMEHVSSSEEPINIFQEIYKQEKNMAIHPRKEKLCTTSCEEVV   60  P33049   CASA2_CAPHI
     ****************..********* * * ** *********.* **:**:* *:***

60   RNANEEEYSIGSSSEESAEVATEEVKITVDDKHYQKALNEINQFYQKFPQYLQYLYQGPI  119  P02663   CASA2_BOVIN
61   RNANEEEYSIRSSSEESAEVAPEEIKITVDDKHYQKALNEINQFYQKFPQYLQYPYQGPI  120  P33049   CASA2_CAPHI
     ********** ********** **:***************************** *****

120  VLNPWDQVKRNAVPITPTLNREQLSTSEENSKKTVDMESTEVFTKKTKLTEEEKNRLNFL  179  P02663   CASA2_BOVIN
121  VLNPWDQVKRNAGPFTPTVNREQLSTSEENSKKTIDMESTEVFTKKTKLTEEEKNRLNFL  180  P33049   CASA2_CAPHI
     ************ *:***:***************:************************

180  KKISQRYQKFALPQYLKTVYQHQKAMKPWIQPKTKVIPYVRYL  222  P02663   CASA2_BOVIN
181  KKISQYYQKFAWPQYLKTVDQHQKAMKPWTQPKTNAIPYVRYL  223  P33049   CASA2_CAPHI
     ***** ***** ******* ********* ****:.*******
```

Figure V.1.4. Alignment of the sequences of Alpha S2 caseins. The aminoacidic sequences from *Bos taurus* (P02663), *Capra hircus* (P33049). The experimentally identified peptides are highlighted in grey. Strikethrough sequence is the signal peptide that is absent in the mature form of the protein. Protein sequences were aligned using Align tool available online at UniProt [14], which uses Clustal-Omega program[15]. * (asterisk) indicates positions which have a single, fully conserved residue; (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix; . (period) indicates conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix.

V.1.4 Discussion

The data provide important technical information relatively to the Asian art of the time. While, in fact, the stylistic relations between the Hellenistic and Roman world (and their successors) in the West, and the Sogdian and Persian art, as well as Indian and even Chinese in the East, have been discussed and are still studied, the technical aspects are mainly unknown. The attribution of the organic materials to the different paint layers was possible thanks to the analysis of a high number of samples and their division in sub-samples. Egg, detected in the inner layers by GC-MS analysis, carried out by the research group of Prof. M.P. Colombini, produces bright paint layers, insoluble and elastic, well resisting to mechanical stress due to daily and seasonal thermo-hygrometric variations. This was well known in the antiquity, both in the East – see for example the polychromy of the Terracotta Army, Xi'an China (3rd century BC [16]) - and the West – see for example the murals of the Palace of Nestor, Pylos Greece (13th century BC)[17]. Considering the dimensions of the Buddhas, the results indicate that the knowledge of the quality of the egg paint and its durability was available to the artists of the Gandharan School in central Asia at the time.

The presence of milk in the statues as painting material in external layers was in a first stage quite surprising. The use of milk as paint binder in open-air paintings seems to be an arguable choice due to the brittleness of the film formed and its consequent sensitivity to the harsh

climatic conditions of the area [18]. In Xuanzang's report of the site in 630 AD, it is mentioned also that the region produced wheat, fruits and flowers and provided rich pasture for sheep, cattle and horses [19]. The choice of milk must have been based on practical considerations such as the availability of the material in the area.

Proteomics data could help us also to understand the origin of the milk found in the fragments. In the case of the samples from the Giant Buddhas, establishing the origin of milk was quite demanding, given the age and the degradation undergone by the proteins. Moreover, the high degree of similarity between homologous proteins can provide multiple possible hits with very similar scores, as can be seen in Table 1 for the case herein reported, where we identified four possible sources for the milk (*Bos taurus*, *Bubalis bubalis*, *Capra hircus* and *Ovis aries*).

Only peptides that are unique for *Bos taurus* and *Capra hircus* sequences were detected and since no peptide that was unique to protein sequences from *Bubalis bubalis* and *Ovis aries* has been identified, we ruled out the possibility of assessing them as the source of milk. Goat hairs identified in both the undercoat and the finish coat, to strengthen the cohesiveness of the clay layers, further supports the presence of goat livestock in the area [20].

Data also clearly showed that one type of milk cannot be associated to a specific restoration layer, and in general, both milks could be present in the same sub-samples. A possible explanation could be that one type of milk was used for the first restoration and another type for the second, and the binder of the second intervention has penetrated the layers underneath. Another possible and more likely explanation is that, considering the gigantic dimensions of the surfaces to be painted, and the multi-layered structure of the paints, restorers used whatever they had, mixing milk from different animals, or using any milk they had available when they were painting the different layers.

V.1.5 Conclusions

The organic materials used as binders and their source were identified with in some of the remaining fragments of the Giant Buddas by analytical approaches based on the combined use of two mass spectrometric techniques: gas chromatography mass spectrometry (GC-MS), by the research group of Prof. M.P. Colombini, and liquid chromatography tandem mass spectrometry LC-MS/MS. Puzzling together these data with some results of previously performed investigations allowed us to improve our understanding of the appearance of the Buddhas along the centuries and to relate it to the materials used but also to obtain unprecedented information from the historic-artistic point of view.

The robustness of the data obtained is in some cases validated by the agreement between the source of the binders identified and remains of other organic materials used in the building of the statues. In fact, the identification of goat milk is in agreement with the presence of goat hair in the clay layers meaning that livestock of goats were present in the area along the 7th to 9th century when the overpaintings were peformed. The use of tragacanth gum (obtained from Astragalus gummifer) as binder in the original paint layers is supported by the identification of Astragalus cuneifolius Bunge as the origin of the material to make the ropes used as core of the fold ridges of the Western Buddha.

The two statues were painted using the same materials in each historical moments. The use of egg in the original paint layers confirms its use in polychrome artworks all around the Indo-European continent in the antiquity. However, data also highlighted a change in the materials used in the overpaintings, as milk and some pigment of lesser quality were used. Though there is no explanation why the original paint layers were reduced to mere traces, the brittleness of paint layers made by using milk as binder in the upper layers explains the reason for the repeated overpainting. Nevertheless, the change of the materials used opens room for debate about several important historic-artistic issues. The choice of materials of less quality could be related to a loss of importance of Bāmiyān as a Buddhist pilgrimage site, to a change of the livestock in the area or to the trading along the Silk Road.

However, the lack of any description of the statues by travellers in the area from Xuanzang description, from the 7th to the 11th century, could support the loss of importance of the Buddhas. In this case, either the technical level of the artists in charge of the overpainting was lower than the ones painting the original layers or a choice of a ready to use material was made on purpose in order to simplify the work. Moreover, the fact that a different painting technique - oil painting - was evidenced in the mural paintings from some of the caves at Bāmiyān could point to different artists travelling in Central Asia using different painting techniques.

Results herein presented on these lost invaluable giant statue demonstrate that, despite the destruction of the war, pieces of human history can still be learned by careful inspection of small fragments of the original masterpieces.

V.2 A Milk and Ochre Paint Mixture Used 49,000 Years Ago at Sibudu, South Africa.

V.2.1 Introduction

Paint is a mixture of solid pigment and a liquid vehicle that can be applied to a surface or to a body for decorative or protective purposes [21]. Here is reported the application of proteomic strategies for the analysis of the oldest case of tempera paint, preserved as a mineral and organic residue on the working edge of a stone flake from a Middle Stone Age (MSA) layer of Sibudu (South Africa), dated to c. 49,000 years ago [22]. Data obtained indicate the presence of bovine milk and collagen, and the results obtained by Gas chromatography/mass spectrometry (GC/MS) and SEM/EDS analyses indicate that it was present in mixture with ochre. Several bone samples of equid and bovid fauna from layer MOD were analysed in order to investigate the effective origin of the source of collagen The presence of bovid milk is surprising at this time since domestic cattle are documented in South Africa only in the Iron Age, about 300 AD [23]. However, hunting practices in the MSA suggest that milk could be obtained by killing a lactating female wild bovid [24].

V.2.2 Material and methods

V.2.2.1 Samples

MOD flake sample Sibudu is a large rock shelter, approximately 40 km north of Durban and 15 km inland from the Indian Ocean. It has a 2.7 m deep MSA sequence spanning from ca. 77,000 to 38,000 years ago excavated since 1998 under the direction of Lyn Wadley [25]. The stone specimen, an unretouched flake of dolerite (length =26.7 mm; width = 17.5 mm; thickness = 5.1 mm) was recovered from layer MOD, a mottled brown silty sand with flecks of white ash, gypsum and charcoal, about 10-15 cm thick, was excavated on a surface of 11 square meters. The residue is along the thin, unretouched, right lateral edge (Figure V.2.1). Dolerite is an igneous rock that occurs in the vicinity of the site [26]. The flake was found in square D6, quadrant d, where 98 flakes of dolerite were also found. Square D6 yielded 19 pieces of unutilized and 8 pieces of utilized ochre [27] 405 blades and flakes, 3 cores and 32 retouched pieces. In layer MOD there is a total of 192 ochre pieces. Supporting information are reported in Online Text (http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0131273#sec015).

Bone samples. A set of four bovid size II and III (50) and one equine (*Equus quagga*) bone samples from Sibudu, layer MOD, were selected for the analysis and catalogued (86, 91,1117, 1120, and 69).

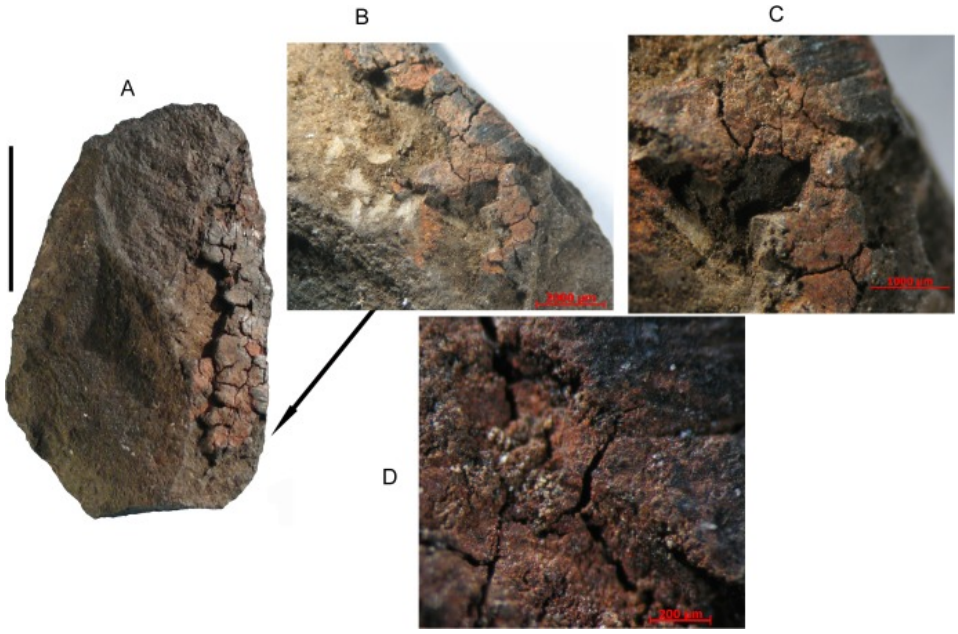


Figure V.2.1. (A) MOD flake before sampling, scale bar = 1 cm. (B) Detail of residue left after sampling for chemical and proteomic analyses. (C) View at 40 x. (D) View at 128 x.

V.2.2.2 Proteomic analyses

Stone flake sample. Proteomic analyses was carried following a minimally invasive proteomic procedure out as described in the section IV without any pretreatment aimed at preserving the integrity of the archeological piece. The micro-black sample of ca. 4 mg was directly submerged in 100 µl ammonium bicarbonate 10 mM pH 8.0, and trypsin digested in heterogeneous phase enzymatic by adding proteomics-grade trypsin 0.1µg/µl at 37 °C for 16 h. The supernatant was then recovered by centrifugation 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). The eluate was analyzed by LC-MS/MS as desctibed in the chapter IV section 2.4

Bone samples. Proteomic analysis was carried out with decalcification pretreatment as described in the chapter IV section 2.3. Approximately 3 mg of bone powder from each sample was weighed into a 1,5 mL microtube and demineralised with extensive incubation in 400 µl 0.5 M EDTA pH 8 at 4°C for 24 h. Insoluble pellets were then extensively washed with MilliQ water, and then submerged in 200 µl ammonium bicarbonate 10 mM pH 8.0, and trypsin digested in heterogeneous phase was carried out.

Data handling. Preliminary data obtained by nanoLC-MSMS analysis both for MOD flake than to bone samples were used to query the SwissProt database 2012_08 (537505 sequences; 190795142 residues), without any taxonomy restriction. Mascot search parameters were: trypsin as enzyme; 3, as allowed number of missed cleavage;10 ppm MS tolerance and 0.6 Da MSMS tolerance; peptide charge from +2 to +3. No fixed chemical modification was inserted, but possible oxidation of methionine and the formation of pyroglutammic acid from glutamine residues at the N-terminal position of peptides were considered as variable modification. As soon as collagen proteins were identified, a new identification run was carried out, with the insertion of hydroxylation on Lysine and Proline as variable modifications, since more confident identifications are commonly obtained with these proteins by taking into consideration their extensive post-translational modifications [28,29]. Individual ion score threshold provided by software to evaluate the quality of matches in MSMS data was 41. Spectra with score of <20, having low quality were rejected. Error tolerant searches were also carried out for each sample using as variable modifications proline and lysine hydroxylation to allow for single amino acid substitutions, enzyme non-specificity, as well as to check for unexpected modifications potentially induced by aging.

V.2.3 Results

Alpha-S1-casein, (P02662), Alpha-S2-casein, (P02663), Beta-casein, (P02666), and Kappa-casein, (P02668), and two collagens [Collagen alpha-1(I) chain, (P02453), and Collagen alpha-2(I) chain, (P02465)] from bovine species were identified in the MOD flake sample (Table V.2.1).

Although casein proteins are highly divergent across mammalian milks [30], they are conserved enough and allowed identification of the old protein homologue. Divergence, moreover, allowed genus-specific sequence information, since several of the matched peptides are specific enough to allow us to confidently identify the bovid origin of the caseins. In Figure V.2.2, the alpha-S1 casein sequences from *Bos taurus* were aligned to those from some other potential sources of milk, i.e. human, sheep and goat. The matched peptides in the MOD flake (highlighted in grey and bold) clearly indicate that the casein belongs to an animal in the bovid family. Likewise, alignments of the other identified caseins with homologous proteins are similarly reported by [31] showing that several of the identified peptides are proteotipic of sequences of bovine caseins.

Table V.2.1. Identification of the proteins in the MOD sample from Sibudu by overnight trypsin digestion in heterogeneous phase and LCMSMS analysisa. Proteins were identified searching UniprotSprot database, without any taxonomy restriction, with MS/MS Ion search Mascot software (Matrix Science).

Identified protein (Accession number)	Total score[a]	Sequence coverage (%)	Individual ion score	Matched sequence (Oxidation of methionine, hydroxylation of proline and lysine, pyro-Glu formation at Gln at the N-terminus of peptides were inserted as variable modifications).
Collagen alpha-1(I) chain (P02453)	832	40	20	GAPGPAGPK + Hydroxy (P)
			26	GAAGLPGPK+ Hydroxy (P)
			34	GPAGPQGPR
			27	GFSGLDGAK
			44	GSEGPQGVR
			26	GFPGADGVAGPK + Hydroxy (P)
			33	GADGAPGKDGVR
			29	GVQGPPGPAGPR + Hydroxy (P)
			28	GQAGVMGFPGPK+ Oxidation (M); Hydroxy (K)
			40	GLTGSPGSPGPDGK+ 2 Hydroxy (P)
			24	GFPGLPGPSGEPGK + Hydroxy (P)
			26	GPSGPQGPSGPPGPK + Hydroxy (P)
			24	GETGPAGPAGPIGPVGAR
			21	GSPGEAGRPGEAGLPGAK + Hydroxy (K); 2 Hydroxy (P)
			20	EGAPGAEGSPGRDGSPGAK + Hydroxy (K); 2 Hydroxy (P)
			22	GPPGPMGPPGLAGPPGESGR + Oxidation (M); 2 Hydroxy (P)
			36	GEPGPTGIQGPPGPAGEEGK + 2 Hydroxy (P)
			31	SGDRGETGPAGPAGPIGPVGAR
			36	GEPGPTGIQGPPGPAGEEGKR + 2 Hydroxy (P)
			41	GDRGETGPAGPPGAPGAPGAPGPVGPAGK + 4 Hydroxy (P)
			26	GAPGDRGEPGPPGPAGFAGPPGADGQPGAK+ 4 Hydroxy (P)
			58	GFSGLQGPPGPPGSPGEQGPSGASGPAGPR + 3 Hydroxy (P)
			29	GLTGPIGPPGPAGAPGDKGEAGPSGPAGPTGAR + Hydroxy (K); Hydroxy (P)
Collagen alpha-2(I) chain (P02465)	575	24	35	GLPGADGR + Hydroxy (P)
			21	GATGPAGVR
			26	GVVGPQGAR
			46	VGAPGPAGAR+ Hydroxy (P)
			24	GEPGNIGFPGPK + 2 Hydroxy (P)
			36	GPAGPSGPAGKDGR
			32	GIPGEFGLPGPAGAR + 2 Hydroxy (P)
			21	GDGGPPGATGFPGAAGR + 2 Hydroxy (P)
			25	GAAGLPGVAGAPGLPGPR+ 3 Hydroxy (P)
			35	GPPGESGAAGPTGPIGSR + Hydroxy (P)
			30	GSTGEIGPAGPPGPPGLR + 2 Hydroxy (P)
			20	EGPVGLPGIDGRPGPIGPAGAR+ 2 Hydroxy (P)
			65	GLPGVAGSVGEPGPLGIAGPPGAR + 3 Hydroxy (P)
Alpha-S1-casein (P02662)	249	33	40	EDVPSER
			20	YLGYLEQLLR
			31	FFVAPFPEVFGK
			24	VPQLEIVPNSAEER
			59	HQGLPQEVLNENLLR
Alpha-S2-casein (P02663)	160	23	34	TKVIPYVR
			30	FALPQYLK
			28	NAVPITPTLNR
			28	ALNEINQFYQK

Beta-casein (P02666)	99	12	24 25 21	VLPVPQK AVPYPQR VLPVPQKAVPYPQR
Kappa-casein (P02668)	47	14	24 23	YIPIQYVLSR SPAQILQWQVLSNTVPAK

```
1    MKLLILTCLVAVALARPKHPIKHQGLPQEVLNENLLRFFVAPFPEVFGKEKVNELSKDIG   60  P02662 CASA1_BOVIN
1    MKLLILTCLVAVALARPKQPIKHQGLPQGVLNENLLRFFVAPFPEVFGKEKVNELSTDIG   60  O62823 CASA1_BUBBU
1    MKLLILTCLVAVALARPKHPINHRGLSPEVPNENLLRFVVAPFPEVFRKENINELSKDIG   60  P18626 CASA1_CAPHI
1    MKLLILTCLVAVALARPKHPIKHQGLSSEVLNENLLRFVVAPFPEVFRKENINELSKDIG   60  P04653 CASA1_SHEEP
     ******************:**:*:**   * *******.******** **::****.***

61   SESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYK  120  P02662 CASA1_BOVIN
61   SESTEDQAMEDIKQMEAESISSSEEIVPISVEQKHIQKEDVPSERYLGYLEQLLRLKKYN  120  O62823 CASA1_BUBBU
61   SESTEDQAMEDAKQMKAGSSSSSEEIVPNSAEQKYIQKEDVPSERYLGYLEQLLRLKKYN  120  P18626 CASA1_CAPHI
61   SESIEDQAMEDAKQMKAGSSSSSEEIVPNSAEQKYIQKEDVPSERYLGYLEQLLRLKKYN  120  P04653 CASA1_SHEEP
     *** ******* ***:* * ******** *.***:************************:

121  VPQLEIVPNSAEERLHSMKEGIHAQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSGAWY  180  P02662 CASA1_BOVIN
121  VPQLEIVPNLAEEQLHSMKEGIHAQQKEPMIGVNQELAYFYPQLFRQFYQLDAYPSGAWY  180  O62823 CASA1_BUBBU
121  VPQLEIVPKSAEEQLHSMKEGNPAHQKQPMIAVNQELAYFYPQLFRQFYQLDAYPSGAWY  180  P18626 CASA1_CAPHI
121  VPQLEIVPKSAEEQLHSMKEGNPAHQKQPMIAVNQELAYFYPQLFRQFYQLDAYPSGAWY  180  P04653 CASA1_SHEEP
     ********: ***:*******  *:**:***.**********:*****************

181  YVPLGTQYTDAPSFSDIPNPIGSENSEKTTMPLW  214  P02662 CASA1_BOVIN
181  YVPLGTQYPDAPSFSDIPNPIGSENSGKTTMPLW  214  O62823 CASA1_BUBBU
181  YLPLGTQYTDAPSFSDIPNPIGSENSGKTTMPLW  214  P18626 CASA1_CAPHI
181  YLPLGTQYTDAPSFSDIPNPIGSENSGKITMPLW  214  P04653 CASA1_SHEEP
     *:****** ***************** * *****
```

Figure V.2.2. Alignment of the sequences of Alpha S1 caseins. The aminoacidic sequences from *Bos taurus* (P02662), *Bubalus bubalis* (O62823), *Capra hircus* (P18626) and *Ovis aries* (P04653). The experimentally identified peptides are highlighted in grey. Strikethrough sequence is the signal peptide that is absent in the mature form of the protein. Protein sequences were aligned using Align tool available online at UniProt [14], which uses Clustal-Omega program[15]. * (asterisk) indicates positions which have a single, fully conserved residue; (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix; . (period) indicates conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix.

Collagen was not identified in the fraction of the sample that was analyzed by GC/MS by the research group of Prof. Colombini given the absence in the amino-acidic profile of hydroxyproline, which is a marker of animal glue. This indicates that, possibly, collagen was not uniformly mixed in the sample as one would have expected if collagen was used as animal

glue and that the collagen source was most likely a minute bone fragment or a fragment of connective or epithelial tissue.

Type I collagen is sufficiently variable between mammal genera to be taxonomically useful. Collagen from several bone samples of equid and bovid fauna from layer MOD was extracted and analyzed. Peptide matches from each sample and from the MOD flake were manually examined and aligned to the collagen sequences from modern species. Significantly, comparisons with samples of equid and bovid fauna from layer MOD indicate that the collagens were from a bovid and not from an equid (Figure V.2.3).

```
MOD flake

P02453 Bos taurus                                                 QLSYGYDEKSTGISVPGPM 180
B9VR88 Equus asinus                                               QLSYGYDEKSAGISVPGPM 180
                                                                  **********:********

#86   Bovid                                                               STGISVPGPM
#69   Equus quagga
#1120 Bovid
#1117 Bovid                                                               STGISVPGPM
#91   Bovid


MOD flake

P02453 Bos taurus   GPSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRP 240
B9VR88 Equus asinus GPSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRP 240
                    ************************************************************

#86   Bovid         GPSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRP
#69   Equus quagga                                          GPPGPPGKNGDDGEAGKPGRP
#1120 Bovid                                                 GPPGPPGK
#1117 Bovid         GPSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRP
#91   Bovid                                                 GPPGPPGK


MOD flake                                    GFSGLDGAK

P02453 Bos taurus   GERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQM 300
B9VR88 Equus asinus GERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQM 300
                    ************************************************************

#86   Bovid         GERGPPGPQGARGLPGTAGLPGMK   GFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQM
#69   Equus quagga  GERGPPGPQGAR                                 GEPGSPGENGAPGQM
#1120 Bovid            GPPGPQGAR               GFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQM
#1117 Bovid         GERGPPGPQGARGLPGTAGLPGMK   GFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQM
#91   Bovid            GPPGPQGAR


MOD flake

P02453 Bos taurus   GPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEGGPQGPR 360
B9VR88 Equus asinus GPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGAR 360
                    ************************************************.**** *

#86   Bovid         GPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAK
#69   Equus quagga  GPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGAR
#1120 Bovid         GPRGLPGERGRPGAPGPAGAR
#1117 Bovid         GPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAK
#91   Bovid                  GRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAK
```

```
MOD flake            GSEGPQGVR                                                GFPGARGPSGPQ

P02453 Bos taurus    GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 420
B9VR88 Equus asinus  GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 420
                     ************************************************************

#86   Bovid          GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ
#69   Equus quagga   GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ
#1120 Bovid          GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ
#1117 Bovid          GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ
#91   Bovid          GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ

MOD flake            GPSGPPGPK                    GEPGPTGIQGPPGPAGEEGKR

P02453 Bos taurus    GPSGPPGPKGNSGEPGAPGSKGDTGAKGEPGPTGIQGPPGPAGEEGKRGARGEPGPAGLP 480
B9VR88 Equus asinus  GPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGPTGIQGPPGPAGEEGKRGARGEPGPTGLP 480
                     ******************.************************************:***

#86   Bovid          GPSGPPGPKGNSGEPGAPGSKGDTGAKGEPGPTGIQGPPGPAGEEGKRGARGEPGPAGLP
#69   Equus quagga   GPSGPPGPK                  GEPGPTGIQGPPGPAGEEGK    GEPGPTGLP
#1120 Bovid          GPSGPPGPKGNSGEPGAPGSKGDTGAK                        GEPGPAGLP
#1117 Bovid          GPSGPPGPK            GDTGAKGEPGPTGIQGPPGPAGEEGKR   GEPGPAGLP
#91   Bovid          GPSGPPGPKGNSGEPGAPGSKGDTGAKGEPGPTGIQGPPGPAGEEGKR   GEPGPTGLP

MOD flake                        GFPGADGVAGPK      GAPGPAGPKGSPGEAGRPGEAGLPGAKGLT

P02453 Bos taurus    GPPGERGGPGSRGFPGADGVAGPKGPAGERGAPGPAGPKGSPGEAGRPGEAGLPGAKGLT 540
B9VR88 Equus asinus  GPPGERGGPGARGFPGADGVAGPKGPAGERGAPGPAGPKGSPGEAGRPGEAGLPGAKGLT 540
                     **********:*************************************************

#86   Bovid          GPPGER      GFPGADGVAGPK      GAPGPAGPKGSPGEAGRPGEAGLPGAKGLT
#69   Equus quagga   GPPGER      GFPGADGVAGPK      GAPGPAGPKGSPGEAGRPGEAGLPGAKGLT
#1120 Bovid          GPPGER      GFPGADGVAGPK      GAPGPAGPKGSPGEAGRPGEAGLPGAKGLT
#1117 Bovid          GPPGER      GFPGADGVAGPK      GAPGPAGPKGSPGEAGRPGEAGLPGAKGLT
#91   Bovid          GPPGER      GFPGADGVAGPK      GAPGPAGPKGSPGEAGRPGEAGLPGAKGLT

MOD flake            GSPGSPGPDGK                        GQAGVMGFPGPK

P02453 Bos taurus    GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 600
B9VR88 Equus asinus  GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 600
                     ************************************************************

#86   Bovid          GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGK    GVP
#69   Equus quagga   GSPGSPGPDGK                        GQAGVMGFPGPKGAAGEPGK    GVP
#1120 Bovid          GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPK
#1117 Bovid          GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPK            GVP
#91   Bovid          GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP

MOD flake

P02453 Bos taurus    GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 660
B9VR88 Equus asinus  GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGESGKPGEQ 660
                     ************************************************:******

#86   Bovid          GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ
#69   Equus quagga   GPPGAVGPAGKDGEAGAQGPPGPAGPAGER
#1120 Bovid          GPPGAVGPAGK
#1117 Bovid          GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ
#91   Bovid          GPPGAVGPAGKDGEAGAQGPPGPAGPAGER
```

```
MOD flake                                    GVQGPPGPAGPR

P02453 Bos taurus   GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 720
B9VR88 Equus asinus GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGSNGAPGNDGAKGDAGAPGAPGSQ 720
                    ************************************:**********************

#86   Bovid         GVPGDLGAPGPSGAR   GFPGERGVQGPPGPAGPRGSNGAPGNDGAKGDAGAPGAPGSQ
#69   Equus quagga                         GVQGPPGPAGPR            GDAGAPGAPGSQ
#1120 Bovid                             GFPGERGVQGPPGPAGPR            GDAGAPGAPGSQ
#1117 Bovid         GVPGDLGAPGPSGAR   GFPGERGVQGPPGPAGPR            GDAGAPGAPGSQ
#91   Bovid                           GFPGERGVQGPPGPAGPR

MOD flake                          GAAGLPGPK         GADGAPGKDGVRGLTGPIGPPGPAGAPGDK

P02453 Bos taurus   GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGAPGKDGVRGLTGPIGPPGPAGAPGDK 780
B9VR88 Equus asinus GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 780
                    **********************************:**********************

#86   Bovid         GAPGLQGMPGER                                GLTGPIGPPGPAGAPGDK
#69   Equus quagga GAPGLQGMPGERGAAGLPGPK                       GLTGPIGPPGPAGAPGDK
#1120 Bovid         GAPGLQGMPGER                    GADGAPGKDGVR
#1117 Bovid         GAPGLQGMPGERGAAGLPGPKGDR                    GLTGPIGPPGPAGAPGDK
#91   Bovid                   GAAGLPGPK         GADGAPGKDGVRGLTGPIGPPGPAGAPGDK

MOD flake           GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAK

P02453 Bos taurus   GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP 840
B9VR88 Equus asinus GETGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP 840
                    **:**********************************************************

#86   Bovid         GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP
#69   Equus quagga GETGPSGPAGPTGAR      GEPGPPGPAGFAGPPGADGQPGAK         GDAGPP
#1120 Bovid
#1117 Bovid         GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAK         GDAGPP
#91   Bovid         GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP

MOD flake

P02453 Bos taurus   GPAGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 900
B9VR88 Equus asinus GPAGPAGPPGPIGSVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPV 900
                    *************.*********************************************.

#86   Bovid         GPAGPAGPPGPIGNVGAPGPK   GSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA
#69   Equus quagga GPAGPAGPPGPIGSVGAPGPK   GSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPV
#1120 Bovid                               GSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA
#1117 Bovid         GPAGPAGPPGPIGNVGAPGPK   GSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA
#91   Bovid         GPAGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA

MOD flake

P02453 Bos taurus   GKEGSKGPRGETGPAGRPGEVGPPGPPGPAGEKGAPGADGPAGAPGTPGPQGIAGQRGVV 960
B9VR88 Equus asinus GKEGGKGPRGETGPAGRPGEAGPPGPPGPAGEKGSPGADGPAGAPGTPGPQGIAGQRGVV 960
                    ****.***************.*************:**************************

#86   Bovid         GK    GPRGETGPAGRPGEVGPPGPPGPAGEKGAPGADGPAGAPGTPGPQGIAGQRGVV
#69   Equus quagga GK       GETGPAGRPGEAGPPGPPGPAGEKGSPGADGPAGAPGTPGPQGIAGQRGVV
#1120 Bovid         GK       GETGPAGRPGEVGPPGPPGPAGEK                        GVV
#1117 Bovid         GK    GPRGETGPAGRPGEVGPPGPPGPAGEKGAPGADGPAGAPGTPGPQGIAGQRGVV
#91   Bovid         GK       GETGPAGRPGEVGPPGPPGPAGEKGAPGADGPAGAPGTPGPQGIAGQRGVV
```

```
MOD flake                    GFPGLPGPSGEPGK          GPPGPMGPPGLAGPPGESGREGAPGAE

P02453 Bos taurus   GLPGQRGERGFPGLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAE 1020
B9VR88 Equus asinus GLPGQRGERGFPGLPGPSGEPGKQGPSGASGERGPPGPVGPPGLAGPPGESGREGSPGAE 1020
                    ************************************:****************:****

#86   Bovid          GLPGQRGERGFPGLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAE
#69   Equus quagga   GLPGQRGERGFPGLPGPSGEPGKQGPSGASGERGPPGPVGPPGLAGPPGESGR
#1120 Bovid          GLPGQR   GFPGLPGPSGEPGK                                EGAPGAE
#1117 Bovid          GLPGQR   GFPGLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAE
#91   Bovid          GLPGQRGERGFPGLPGPSGEPGK                                EGAPGAE


MOD flake           GSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGPVGPAGKSGDRGETGPAGPAGPIGPV

P02453 Bos taurus   GSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGPVGPAGKSGDRGETGPAGPAGPIGPV 1080
B9VR88 Equus asinus GSPGRDGSPGPKGDRGETGPAGPPGAPGAPGAPGPVGPAGKSGDRGEAGPAGPAGPIGPV 1080
                    ********** ************************************:************

#86   Bovid          GSPGR       GDRGETGPAGPPGAPGAPGAPGPVGPAGKSGDRGETGPAGPAGPIGPV
#69   Equus quagga                  GETGPAGPPGAPGAPGAPGPVGPAGKSGDRGEAGPAGPAGPIGPV
#1120 Bovid          GSPGRDGSPGAK   GETGPAGPPGAPGAPGAPGPVGPAGK    GETGPAGPAGPVGPV
#1117 Bovid          GSPGR       GDRGETGPAGPPGAPGAPGAPGPVGPAGKSGDRGETGPAGPAGPIGPV
#91   Bovid          GSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGPVGPAGKSGDRGETGPAGPAGPIGPV


MOD flake           GAR                              GFSGLQGPPGPPGSPGEQGPSGASGPAGPR

P02453 Bos taurus   GARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPR 1140
B9VR88 Equus asinus GARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPR 1140
                    ************************************************************

#86   Bovid          GARGPAGPQGPR                    GFSGLQGPPGPPGSPGEQGPSGASGPAGPR
#69   Equus quagga   GARGPAGPQGPR                            GPPGSPGEQGPSGASGPAGPR
#1120 Bovid          GAR
#1117 Bovid          GARGPAGPQGPR                    GFSGLQGPPGPPGSPGEQGPSGASGPAGPR
#91   Bovid          GARGPAGPQGPRGDK


MOD flake

P02453 Bos taurus   GPPGSAGSPGKDGLNGLPGPIGPPGPRGRTGDAGPAGPPGPPGPPGPPGPPSGGYDLSFL 1200
B9VR88 Equus asinus GPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGPPSAGFDFSFL 1200
                    *******:***************************.****************.*:*:***

#86   Bovid          GPPGSAGSPGKDGLNGLPGPIGPPGPR
#69   Equus quagga
#1120 Bovid
#1117 Bovid          GPPGSAGSPGKDGLNGLPGPIGPPGPR
##91  Bovid                    DGLNGLPGPIGPPGPR


MOD flake

P02453 Bos taurus     PQPPQEKAHDGGRYYR 1260
B9VR88 Equus asinus   PQPPQEKSHDGGRYYR 1260
                      *******:********

#86   Bovid
#69   Equus quagga
#1120 Bovid
#1117 Bovid
##91  Bovid
```

Figure V.2.3. **Alignment of the matched peptides in the MOD flake to the Collagen alpha-2(I) chain from *Bos taurus* (P02465) and *Equus asinus* (B9VR89).** Peptides identified in the MOD flake sample were manually aligned over the automatically generated alignment (CLUSTAL O(1.2.0) multiple sequence alignment, of the two collagen alpha-2(I) chain from *Bos taurus* (P02465) and *Equus asinus* (B9VR89). Peptides from MOD flake are reported in bold. * (asterisk) indicates positions which have a single, fully conserved residue; (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix; . (period) indicates conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix.
Under the alignment, peptides matched in samples from bones of some Bovids and *Equus quagga* remains recovered from Sibudu layer MOD, are reported. Underlined sequences are those identified in the error tolerant database search and the modified residue is indicated in red.
Numbering refers to the precursor polypeptide chain while only the sequence of the mature protein is reported. X stands for either Ile or Leu that cannot be distinghished as substitution since are isobaric.

V.2.4 Discussion

The question whether milk was used as glue or protein binder as in tempera painting is not questionable. Milk is an emulsion of fats within a water-based fluid, which also contains carbohydrates, proteins and minerals. Casein is the major protein component of milk and is extracted by heating milk and adding acid [32]. Casein then precipitates into small, solid globules. Use of casein as a glue requires mixing with quicklime or hydrated lime (calcium hydroxide, also called slaked lime) [33]. Lime is obtained by crushing and heating limestone. However, recent geological mapping shows that there is no limestone in the Sibudu area. For use as a binder for pigments, lumps of casein can also be dissolved and thinned with water and wood ashes (mainly composed of calcium carbonate or calcium oxide). The absence of calcium (Ca) was observed in the EDS analyses. Given the absence of lime and wood ashes in the Sibudu residue, we reject an interpretation of the casein as either glue or a complex form of binder. Instead, we conclude that powdered ochre was simply mixed with milk in its liquid form and casein is the residue left after evaporation, since milk is a water-based liquid.

The identification of a milk and ochre mixture on a 49,000 year old flake from a late MSA layer of Sibudu raises a number of questions indicated below.

How was the milk obtained from wild bovids?

Some African bovids separate from the herd when giving birth and hide their calves until they are strong enough to keep up with the herd as a protection against carnivores [34]. This is the case with the eland (*Taurotragus oryx*), the kudu (*Tragelaphus strepsiceros*), and the impala (*Aepicerus melampus*). The bushbuck (*Tragelaphus scriptus*) and the red and blue duiker (*Cephalophus natalensis* and *Philantomba monticola*) are solitary animals; they hide their

young and go off to browse alone. All these species and the buffalo (*Syncerus caffer*, a herd animal) are represented in the MOD fauna.

Klein [24] notes that in the MSA deposits of Klasies River Cave I the giant buffalo (*Pelorovis antiquus*) a very large bovid with a horn span of 3 m or more, falls into two age groups: newborn or perhaps fetal individuals and physically mature individuals with worn molars. According to Klein, the explanation for this peculiar distribution may be that the Klasies people, faced with such a formidable prey, focused on females in advanced pregnancy or in the process of giving birth. Cows are easy prey when giving birth and would already have milk. It would not be difficult for hunters to locate lactating cows, particularly among seasonal breeders. This hunting pattern would result in access to milk and might explain the inhomogeneous presence of collagen observed by proteomic analyses. In the process of collecting milk from the mammary glands of a dead animal, a small amount of its epithelial tissue could have been taken as well.

Was the mixture a hafting adhesive?

The use of adhesives made from resin or plant gum combined with ochre has previously been suggested for the hafting of stone points and backed pieces to be used as hunting tools at Sibudu. Howiesons Poort and later MSA tools were studied by optical microscopy in combination with spatial distribution of residues on the tool [35,36]. Replication experiments [37,38] demonstrated the usefulness of ochre as a loading agent with products such as plant gum. Microscopic analysis showed that ground ochre was mixed with a plant exudate; experiments suggested that ochre-loaded adhesives are robust, dry fast and are not hygroscopic. Most archaeological residues occurred at the base of the point or on the back of backed pieces implying that the mixture was used for hafting. However optical microscopy alone cannot securely distinguish between plant gum (polysaccharides) and resin (terpenoid secretions from trees) [39] thus new analyses using GC/MS are appropriate.

As mentioned before, the use of casein as a glue requires mixing with hydrated lime [33] which is obtained by crushing and heating limestone. Recent geological mapping shows that there is no limestone in the Sibudu area so; in conclusion, there is no evidence that the MOD flake residue was a hafting adhesive.

Was the MOD flake used to treat animal skin?

Three scrapers from layer SS at Sibudu, dated to 59.6 ± 2.3 ka by OSL [22], have ochre and animal products (fat, muscle tissue) on their edges. Ochre and fat were quite probably used for processing hides and some of the animal products were the result of use on animal skin [40]. The presence of milk is however not consistent with use for processing hides.

What was the use of the MOD flake mixture?

Our analyses show that this ochre-based mixture was neither a hafting adhesive nor a residue left after treating animal skins, but a liquid mixture consisting of a solid (powdered) pigment mixed with milk; in other words, a paint medium that could have been applied to a surface or on the skin.

The edge of the flake may have been used as a mixing implement for combining the substances. A mixture of milk and ochre may have been used for body painting, skin protection or for painting on a rock surface like a stone slab. The oldest representational rock art in Africa is on seven small quartzite slabs from the Apollo 11 rock shelter in Namibia. The date of 27,500 years ago is an average of 3 radiocarbon samples of charcoal in hearths from the same layer [41]. The pigments and binders of the slabs have not yet been analyzed. Traces of red coloring on ostrich eggshell fragments from the older MSA layers in Apollo 11 [42] imply that there may have been simpler precursors to the representational art. Milk or casein has never been documented as media for pigment in San rock art [43]. In fact, little is known about the methods or ingredients that San used for making paint and even less is known about paint binders [43–45]. Blood was identified in rock art paint at Rose Cottage [46]. Some ethnographies and historic sources mention milk and egg as a binder for rock paint but there is as yet no direct evidence for its use [43].

V.2.5 Conclusions

Animal or vegetable fat were the most commonly reported media for body paint [47]. Clarified butter, in fact, is mixed with ochre for body paint used by the Himba of Namibia [21]. Only further research on pigments and binders of rock art in South Africa will allow us to identify similarities or differences that may support one hypothesis over the other.

Nevertheless, obtaining milk by killing a lactating wild bovid and then mixing it with ochre shows that MSA people experimented with colouring materials in creative ways and may have attributed a special significance and value to that product. In the present state of research, we cannot demonstrate conclusively whether paint such as that found at Sibudu was a precursor to rock art painting or used for body painting. In either case this find suggests that toward the end of the Middle Stone Age new techniques of symbolic communication were evolving that became, some millennia later, key elements in the social life of hunter-gatherers in South Africa.

V.3 Different organic-based binders found in public outdoor and private indoor Cuma wall paintings

V.3.1 Introduction

Archeometry allows to investigate a wide range of materials shedding light on the history of the men who built and used them, their socio-economic relationships, their needs. The archaeological and artistic material are very often diverse and therefore requires a physical-chemical multi-methodological diagnostic approach to make it possible to investigate fully all its components. Moreover, a multi-disciplinary approach is required to place the information gained in terms of material composition in their historical, cultural, geographical, economic and social context. The integration between the humanities and scientific studies of the subject allows a proper analysis so that the data obtained are evaluated in the best way.

Herein, we show the results obtained by the application of chemical strategies to have a general picture of the organic nature of binders, with a special attention to implications on the technology of wall painting preparation of samples from the archaeological site of Cuma.

The chemical analysis of wall paintings was performed for the cases of study (*Tempio con Portico* - TCP and *domus*). In particular, the discussion was addressed to possible implications of differences between these two cases according to different conservative requirements (outdoor versus indoor) and different destinations of the building (public versus private).

These experiments are part of a collaborative multidisciplinary research project carried out within the Department of Chemical Sciences and with the research group of Prof. G. Greco from the Department of Humanities, University of Naples 'Federico II'.

V.3.2. Material and Methods

V.3.2.1. Samples

Cuman findings were obtained by prof. G. Greco (Deptartment of Human studies, University of Naples 'Federico II'). The Cuman samples come from two distinct zones located inside the archaeological area of the lower city of Cuma: a niche of the *Tempio con Portico* (TCP) (Figure V.3.1) and the floor of one of the houses (*domus*) located on the west side of the TCP.

Samples coming from the niche TCP were extracted directly from the wall of the niche, which is exposed to the weather elements. Samples coming from the *domus*, located in the west area

of the TCP, originate from the excavation of the domus floor where they were kept sealed for centuries. Both kind of samples were not treated with ethyl silicate.

V.3.2.2. Proteomic analyses

Reagents. Ammonium hydrogen carbonate (Ambic), urea, TPCK-treated trypsin, KOH ethanol, trifluoroacetic acid, N,O-Bis(trimethylsilyl)trifluoroacetamide and ammonia were from Sigma; Formic acid and Acetonitrile (ACN) were purchased from Baker; recombinant Peptide N-Glycosidase F (PNGaseF) was from Roche. Deionized water was obtained from Millipore cartridge equipment.

S*ample treatment.* A deglycosylation pre-treatment step was performed as described by Vinciguerra *et al* [1]. Briefly, 50 µL of AMBIC 50 mM containing 60mU/µL of PNGaseF solution was added to microsamples (ca 600-800 µg), and incubated at 37°C for two hours. The reaction was stopped by boiling the sample for 2 minutes. Subsequently, enzymatic digestion was carried out with the urea 6 M pretreatment as described in the chapter IV [29].

LC-MS/MS analyses was carried out as described in the section V.1.2.2. An *ad hoc* exclusion list of the peptides from keratins and other contaminating proteins was adopted to overcome the problems of contaminations.

Data handling. The acquired MS/MS spectra were transformed in Mascot Generic files (.mgf) format and used to query the SwissProt database 2015_04 (548208 sequences; 195282524 residues), without any taxonomy restriction for protein identification with licensed version of MASCOT software (www.matrixscience.com) version 2.4.0 with the same parameters as reported in the chapter IV, section 2.5[29]. No fixed chemical modification was inserted, but possible oxidation of methionine residues, formation of pyroglutamic acid from glutamine residues at the N-terminal position of peptides, and deamidation at asparagines and glutamines were considered as variable modifications[48].

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 40. Spectra with MASCOT score of <10, having low quality were rejected.

Figure.V.3.1. The building of *Tempio con Portico*(A) has been in the Julio-Claudian age. The second context is the *domus* (B), inhabited between the end of the III century b.C. to the IV century a.C.; Some wall painting fragments (C) of this site have been investigated using a mutimethodological chemical approaches.

V.3.2.3. Lipid-resinous analyses

Lipid-resinous extraction. Approximately 800 µg of samples were subjected to ammonia extraction as described by Lluveras *et al*.(2010)[49]. Concisely, 200µL of 2.5 M NH3 is added to the sample in an ultrasonic bath at 60 °C for 120 min. The residue containing insoluble organic and inorganic species was subjected to saponification with 300 µL of 10 wt % KOH in ETOH at 60 °C for 120 min. After saponification, the alcoholic solution is diluted in bidistilled water and acidified with trifluoroacetic acid (aqueous solution 1:1), and then extracted with hexane 200 µL and diethyl ether 200 µL. The extract was dried under a nitrogen flow and derivatized with 20 µL of N,O-Bis(trimethylsilyl)trifluoroacetamide60 °C for 30 min. Afterwards, the solution was redissolved in acetone and hexane (300 µL 1:1) andanalyzed by GC/MS.

GC-MS analysis. Lipid-resinous analyses were performed with an Agilent 6890 Series GC, coupled to a detector MSD quadrupole mass spectrometer 5973, equipped with a gas chromatograph by 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 stationary phase used consists of polydimethylsiloxane containing 5% of a phenyl derivative. The injection temperature was 250 °C. The oven temperature was increased from 90 °C to 140 °C at 10 °Cmin^{-1} held for 1 min, increasing to 280 °C at 20 °Cmin^{-1} and held for 3 min, and

finally to 300 °C at 10 °C min^{-1} held for 5 min and remaining at this temperature for 4 min; carrier gas (He) at a constant flow rate of 1 mL min-1. MS interface was kept at 250 °C; electron ionization source was set at290 °C, +70eV (30-450 *m/z*) and electron multiplier was at 1.7 kV. Each species was interpreted on the basis of electron impact spectra (NIST library 2008).

V.3.3. Results and discussion

Some of the wall paintings from *domus* exhibit proteins based binders, while drying oils were identified in TCP wall paintings. In all the samples, as could be expected taking into account the extensive exposure of the samples to external contaminations during and after the excavation, keratin proteins were identified. Two different type of binders were identified in two of the samples from *domus*, as reported table V.3.1 (keratins are not reported in the Table). In the case of the sample named Cuma dark red (12M080) four proteins from egg yolk were identified, thus clearly indicating the use of egg yolk as binder, while in the Cuma blue (12M081) animal glue was used, as demonstrated by the identification of collagen proteins with high confidence. Details of the proteins identification are reported in Table V.3.2.

Table V.3.1. Protein composition of samples from *domus*

	LC-MS/MS Analysis		
Sample	**Proteins**	**Score**	**N° of Peptides**
Cuma dark-red (purple) 12M080	Vitellogenin-2 [*Gallus gallus*]	1163	35
	Vitellogenin-1 [*Gallus gallus*]	626	20
	Apolipoprotein B (Fragment) [*Gallus gallus*]	107	4
	Apovitellenin-1 [*Gallus gallus*]	90	2
Cuma blue 12M081	CO1A1 [*Bos taurus*]	798	27
	CO1A2 [*Bos taurus*]	340	11

Table V.3.2: Details of the identification of proteins in Cuma paint samples by LC-MSMS. An aliquot of the *domus* fragment "Purple 12M080" and "Blue 12M081" were subjected to trypsin digestion after PNGaseF pre-treatment and the resulting peptide mixtures analysed by LC-MSMS. Proteins were identified searching UniprotSprot database with MSMS Ion search Mascot software (Matrix Science) without any taxonomy restriction, with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications and collagen proteins were identified with the insertion of hydroxylation on Lysine and Proline. Only proteins identified with at least two peptides were considered as significative.

Purple pigment 12M080				
Identified protein (Accession number)	**Total score[a]**	**Sequence coverage (%)**	**Individual ion scores[b]**	**Matched sequence**
Vitellogenin-2 (P02845)	1163	24	13	K.ALQGWK.E
			29	R.IANQIR.N
			14	K.TFNEVK.F
			40	R.AAVSVEGK.M
			25	K.TVVEPADR.N
			24	R.YVIQEDR.K
			29	R.NIGELGVEK.R
			35	K.VSTELVTGR.F
			32	R.MVVALTSPR.T
			37	K.ELLQQVMK.T + Oxidation (M)
			52	K.TVQLAGVDSK.C
			64	R.QQLTLVEVR.S
			13	R.TMFPSAIISK.L + Oxidation (M)
			28	R.NIGELGVEKR.T
			22	K.MTPPLTGDFR.L
			(22)	K.MTPPLTGDFR.L + Oxidation (M)
			32	R.ADTYFDNYR.V
			18	R.VEGLADVIMKR.N + Oxidation (M)
			14	K.LEWPKVPSNVR.S
			16	R.NIPFAEYPTYK.Q
			67	R.IANADNLESIWR.Q
			(34)	R.LSQLLESTMQIR.S
			53	R.LSQLLESTMQIR.S + Oxidation (M)
			15	R.DASFIQNTYLHK.L
			33	R.VGATGEIFVVNSPR.T
			46	K.ILGQEVAFININK.E
			106	K.QSDSGTLITDVSSR.Q
			13	K.LEISGLPENAYLLK.V
			58	R.SPQVEEYNGVWPR.D
			26	K.ELPTETPLVSAYLK.I
			13	K.IVLMPVHTDADIDK.I
			15	R.KKPMDEEENDQVK.Q + Oxidation (M)
			35	R.WLLSAVSASGTTETLK.F
			32	K.LKQSDSGTLITDVSSR.Q
			35	K.EALQPIHDLADEAISR.G
			33	R.NSIAGQWTQPVWMGELR.Y
			(18)	R.NSIAGQWTQPVWMGELR.Y + Dea (NQ)
			36	R.QVYQISPFNEPTGVAVMEAR.Q
Vitellogenin-1 (P87498)	626	13	10	K.FIITTR.K
			38	K.ANLVDVTK.S
			14	K.QVAEVPPK.E
			29	K.LALIEVQK.Q
			36	K.ALLLSEIR.C

			24	R.NIEDLAASK.M
			18	K.KANLVDVTK.S
			21	K.NVNFDGEILK.V
			40	R.LTELLNSNVR.L
			18	K.ALGNVGHPASIK.H
			34	R.IAALATTGQMAR.K + Oxidation (M)
			57	K.SNIEEVLLALK.A
			19	K.ILDDTDNQATR.N
			57	K.VAGNVQAQITPSPR.S
			77	R.AANEENYESVWK.Q
			33	R.YLLDLLPAAASHR.S
			25	R.MSFKPVYSDVPIEK.I
			26	K.MTPVLLPEAVPDIMK.M
			(26)	K.MTPVLLPEAVPDIMK.M + Oxidation (M)
			(18)	K.MTPVLLPEAVPDIMK.M + 2 Oxidation (M)
			12	R.GYSPDKDWETNYDFR.E
			27	R.VADPIEVGIAAEGLQEMFVR.G + Dea (NQ); Oxidation (M)
Apolipoprotein B (Fragment) (P11682)	107	13	12	R.YWSAVAK.R
			26	R.VPASETILR.G
			52	R.QTFVTLQEADFAGK.L
			9	K.YTGEELYLMTTEK.A + Oxidation (M)
Apovitellenin-1 (P02659)	90	16	58	K.LAEQLMEK.I + Oxidation (M)
			33	R.NFLINETAR.L

Blue pigment 12M081				
Identified protein (Accession number)	**Total score[a]**	**Sequence coverage (%)**	**Individual ion scores[b]**	**Matched sequence**
Collagen alpha-1(I) chain (P02453)	798	28	14	R.GPPGPPGK.N + Hyd (K); Hyd (P)
			29	R.GPAGPQGPR.G
			18	R.GFSGLDGAK.G
			15	R.GPPGPQGAR.G + Hyd (P); Dea (NQ)
			45	R.GSEGPQGVR.G
			31	R.GVVGLPGQR.G + Hyd (P); Dea (NQ)
			20	K.QGPSGASGER.G + Dea (NQ)
			34	R.GFPGADGVAGPK.G + Hyd (P)
			15	R.GRPGAPGPAGAR.G + 2 Hyd (P)
			24	K.GADGAPGKDGVR.G
			35	R.GVQGPPGPAGPR.G + Hyd (P); Dea (NQ)
			45	R.GLPGTAGLPGMK.G + Oxi (M); Hyd (K); 2 Hyd (P)
			16	R.GVPGPPGAVGPAGK.D + Hyd (P)
			39	R.GQAGVMGFPGPK.G + Oxi (M); Hyd (K); Dea (NQ)
			12	R.GLTGSPGSPGPDGK.G + 2 Hyd (P)
			40	K.GEAGPSGPAGPTGAR.G
			15	R.GFPGLPGPSGEPGK.Q + 3 Hyd (P)
			26	R.GEPGPAGLPGPPGER.G + 3 Hyd (P)
			79	R.GSAGPPGATGFPGAAGR.V + 2 Hyd (P)
			85	R.GETGPAGPAGPIGPVGAR.G
			25	K.DGLNGLPGPIGPPGPR.G + 3 Hyd (P); Dea (NQ)
			19	K.GANGAPGIAGAPGFPGAR.G + 3 Hyd (P); Dea (NQ)
			28	R.GLTGPIGPPGPAGAPGDK.G + 2 Hyd (P)
			50	K.GEPGPTGIQGPPGPAGEEGKR.G + 2 Hyd (P); Dea (NQ)

			19	R.GETGPAGRPGEVGPPGPPGPAGEK.G + Hyd (K); 2 Hyd (P)
			33	R.GETGPAGPPGAPGAPGAPGPVGPAGK.S + 6 Hyd (P)
Collagen alpha-2(I) chain (P02465)	340	11	18	R.GLPGADGR.A + Hyd (P)
			34	R.GVVGPQGAR.G + Dea (NQ)
			43	R.VGAPGPAGAR.G + Hyd (P)
			12	R.GPAGPSGPAGK.D
			50	T.IGQPGAVGPAGIR.G + Hyd (P); Dea (NQ)
			42	R.GIPGPVGAAGATGAR.G + Hyd (P)
			36	R.GIPGEFGLPGPAGAR.G + 2 Hyd (P)
			17	R.GEPGPAGAVGPAGAVGPR.G + Hyd (P)
			24	K.GAAGLPGVAGAPGLPGPR.G + 3 Hyd (P)
			25	R.GPPGESGAAGPTGPIGSR.G + Hyd (P)
			35	R.GSTGEIGPAGPPGPPLR.G + 2 Hyd (P)

[a]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*Log(P), where P is the probability that the observed peptide match is a random event. [b]in brackets ion scores that Mascot did not consider for deriving total protein score, but that we reported for the sake of completeness to show that .

Given the lack of protein binders, a GC-MS analysis was performed on the TCP samples to look for organic substances. Figure V.3.1 showed the total ion current chromatogram obtained for the sample Cuma 1 (red pigment). As reported in the Table V.3.3 different fatty acids were detected. As a whole, a fatty acids composition can be achieved at qualitative and semi quantitative levels. Myristic, palmitic and stearic acids were detected in all the samples (highlighted in Table V.3.3). These organic acids are present in the composition of both plant and animal oils. However, animal fats generally contain less palmitic acid than stearic acid, while palmitic acid predominates over stearic acid in plant oils [32]. Semiquantitative analysis showed an higher amount of palmitic acid with respect to stearic acid, thus suggesting the occurrence of vegetable oils.

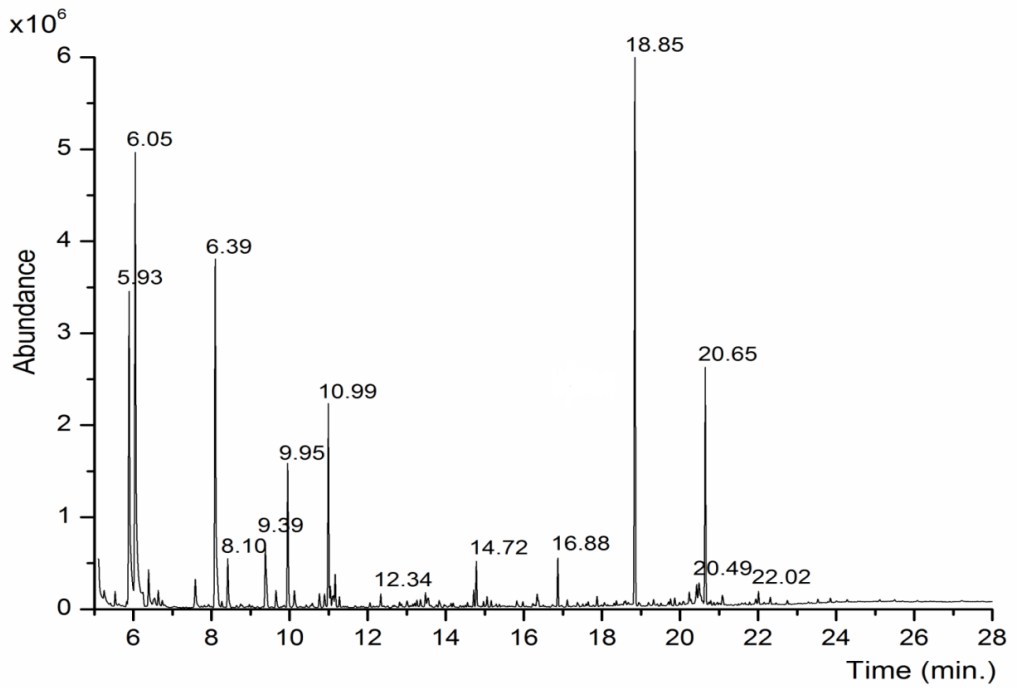


Figure V.2.1. Cuman (red pigment) sample GC/MS TIC profile of the organic acids fraction.

Table V.2.3. Fatty acidic composition of the pigmented Cuman samples from TCP

Sample	Compound	Retention time	Compound	Retention time
Cuma1 (Red pigment.)	2,3-butanediol	6.05	Decanoic acid (capniric)	12.34
	2-propionic acid (lactic) acid)	6.39	Dodecanoic acid (lauric)	14.72
	3-hydroxybutyric acid	8.01	Tetradecanoic acid (myristic)	16.87
	Benzoic acid	9.39	Hexadecanoic acid (palmitic)	18.85
	Octanoic acid (caprylic)	9.65	Oleic acid	20.49
	Glycerol	9.96	Octadecanoic acid (stearic acid)	20.64
	Nonanoic acid (pelargonic)	10.99	Acid deidroabielico	22.01
Cuma2 (Yellow pigment)	2,3-butanediol	6.04	Dodecanoic acid (lauric)	14.72
	3-hydroxybutyric acid	8.02	2,3-butanediol	6.07
	Benzoic acid	9.39	Tetradecanoic acid (myristic)	16.88
	Octanoic acid (caprylic)	9.65	Hexadecanoic acid (palmitic)	18.85
	Glycerol	9.96	Oleic acid	20.48
	Nonanoic acid (pelargonic)	11.04	Octadecanoic acid (stearic acid)	20.65
	Decanoic acid (caprinic)	12.34	Acid deidroabielico	22.02
	Hydroxybutyric acid	13.36		
Cuma 4 (Black pigment)	2-propionic acid	6.39	Hydroxybutyric acid	13.36
	3-hydroxybutyric acid	8.04	Dodecanoic acid (lauric)	14.72
	Benzoic acid	9.4	Tetradecanoic acid (myristic)	16.88
	Octanoic acid (caprylic)	9.66	Hexadecanoic acid (palmitic)	18.85
	Glycerol	9.96	Oleic acid	20.49
	Nonanoic acid (pelargonic)	10.99	Octadecanoic acid (stearic acid)	20.65
	Decanoic acid (caprinic)	12.34	Acid deidroabielico	22.02

V.3.4. Conclusions

This work is part of a multimethodological characterization of wall paintings (mortars pigments and their binders) from Cuman Archeological sites. The results suggested us information on a) how possibly Romans responded to the preservation of outdoor wall paintings compared to indoor ones, and on b) how they prepared a public temple dedicated to an emperor compared to a private *domus*.

Differences were observed not only in binders composition, but also in starting materials used for mortars. Particularly, the multi-methodological analysis performed on about thirty Cuman artifacts from *Tempio con Portico* (TCP) and housing district (*domus*), in collaboration with Proff. Vergara, M. Trifuoggi and L.Paduano at the Department of Chemical Sciences, University of Naples 'Federico II' offers comparisons on the wall paintings fabrication for distinct uses and/or conservation. Raman, X-ray diffraction and Ionic Chromatography studies revealed a highly variable unusual dolomite content only in the TCP mortars, and not in *domus* mortars where only calcite is present. The reason for such a peculiar composition in Cuman mortar has been discussed, and possibly related to an engineer's choice for conservative and aesthetic reasons. The possible sea-water contamination has been ruled out. Furthermore, considering the geological maps we can speculate that Romans of Cuma have taken the construction materials from the mountains between Caserta and Benevento and that they have been transporting them via the Volturno river to their city.

Differences between TCP and *domus* extend also to different painting binders: protein based binders in *domus* (egg and animal glue), whereas in TCP the use of drying vegetable oil is suggested.

As far as pigments are concerned, they are all comparable to other Roman sites: vegetable carbon (ivory black) for black pigments and cinnabar (HgS) for bright red; hematite (α-Fe_2O_3) for dark red, yellow pigments were Yellow ochre (α-FeOOH), blue pigment was *Egyptian blue*, whereas green probably comes from *terrae verte*.

The presence of mortar in pigments of Cuma (and occasionally the presence of the pigment, HgS, in the mortar) suggests possible *a fresco* procedure, though the different types of binders suggests that different kind of techniques might also have been used.

V.4. References

1. Francioni, F. & Lenzerini, F. The Destruction of the Buddhas of Bamiyan and International Law. **14,** 619–651 (2003).
2. Margottini, C. Instability and geotechnical problems of the Buddha niches and surrounding cliff in Bamiyan Valley , central Afghanistan. 41–51 (2004). doi:10.1007/s10346-003-0010-0
3. Pfeffer, S. & Thiemann, L. Dating of the Buddha Statues – AMS 14 C Dating of Organic Materials. 231–236 (2004).
4. Hackin, J. The Colossal Buddhas at Bamiyan Their Influence on Buddhist Sculpture. *East. Art* **1,** 109–116 (1928).
5. Takayasu Higuchi, G. B. Bamiyan: Buddhist Cave Temples in Afghanistan. *World Archaeol.* **27,** 282–302 (1995).
6. Rees-Jones, S. G. *The Organic Chemistry of Museum Objects*. (Elsevier Science, 1987).
7. Colombini, M. P., Andreotti, A., Bonaduce, I., Modugno, F. & Ribechini, E. Analytical strategies for characterizing organic paint media using gas chromatography/mass spectrometry. *Acc. Chem. Res.* **43,** 715–727 (2010).
8. Bonaduce, I. *et al.* New Insights into the Ageing of Linseed Oil Paint Binder: A Qualitative and Quantitative Analytical Study. *PLoS One* **7,** e49333 (2012).
9. Emilia Ghezzi, L. *et al.* Interactions between inorganic pigments and rabbit skin glue in reference paint reconstructions. *J. Therm. Anal. Calorim.* **122,** 315–322 (2015).
10. Duce, C. *et al.* Physico-chemical characterization of protein-pigment interactions in tempera paint reconstructions: Casein/cinnabar and albumin/cinnabar. *Anal. Bioanal. Chem.* **402,** 2183–2193 (2012).
11. Duce, C. *et al.* Interactions between inorganic pigments and proteinaceous binders in reference paint reconstructions. *Dalt. Trans.* **42,** 5975–5984 (2013).
12. Pellegrini, D. *et al.* Fourier transform infrared spectroscopic study of rabbit glue/inorganic pigments mixtures in fresh and aged reference paint reconstructions. *Microchem. J.* **124,** 31–35 (2016).
13. Nadolny, J. The first century of published scientific analyses of the materials of historical painting and polychromy, circa 1780-1880. *Rev. Conserv.* 39–51 (2003).
14. Consortium, T. U. UniProt : a hub for protein information. **43,** 204–212 (2015).
15. Sievers, F. *et al.* Fast , scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. (2011). doi:10.1038/msb.2011.75
16. Bonaduce, I., Blaensdorf, C., Dietemann, P. & Colombini, M. P. The binding media of the polychromy of Qin Shihuang's Terracotta Army. *J. Cult. Herit.* **9,** 103–108 (2008).
17. Brecoulaki, H., Andreotti, A., Bonaduce, I., Colombini, M. P. & Lluveras, A. Characterization of organic media in the wall-paintings of the 'Palace of Nestor' at Pylos, Greece: evidence for a secco painting techniques in the Bronze Age. *J. Archaeol. Sci.* **39,** 2866–2876 (2012).
18. Delmonaco, G. and Margottini, C. in *After the Destruction of Giant Buddha Statues in Bāmiyān (Afghanistan) in 2001* (ed. Springer) 69–100 (2014).
19. Samad, R. U. *The Grandeur of Gandhara: The Ancient Buddhist Civilization of the Swat, Peshawar, Kabul and Indus Valleys*. *Algora Publishing* (2011). doi:10.1017/CBO9781107415324.004
20. Blaensdorf, C. *et al.* in *The Giant BUddhas of Bāmiyān* (ed. ICOMOS) 229–230 (hendrik Babler, 2009).
21. Rifkin, R. F., D'Errico, F., Dayet-Boulliot, L. & Summers, B. Assessing the photoprotective effects of red ochre on human skin by in vitro laboratory experiments. *S. Afr. J. Sci.* **111,** 1–8 (2015).
22. Jacobs, Z., Wintle, A. G., Duller, G. A. T., Roberts, R. G. & Wadley, L. New ages for the post-

Howiesons Poort, late and final Middle Stone Age at Sibudu, South Africa. *J. Archaeol. Sci.* **35,** 1790–1807 (2008).
23. Robbins, L. H. *et al.* The Advent of Herding in Southern Africa: Early AMS Dates on Domestic Livestock from the Kalahari Desert. *Curr. Anthropol.* **46,** 671–677 (2005).
24. Klein, R. G. Middle Stone Age Man-Animal Relationships in Southern Africa: Evidence from Die Kelders and Klasies River Mouth. *Science (80).* **190,** 265–267 (1975).
25. Wadley, L. & Jacobs, Z. Sibudu Cave: background to the excavations, stratigraphy and dating. *South African Humanit.* **18,** 1–26 (2006).
26. Wadley, L. & Kempson, H. A review of rock studies for archaeologists, and an analysis of dolerite and hornfels from the Sibudu area, KwaZulu-Natal. *South. African Humanit.* **23,** 87–107 (2011).
27. Hodgskiss, T. An investigation into the properties of the ochre from Sibudu. Southern African Humanities **24**: 99–120 (2012).
28. Leo, G. *et al.* Proteomic strategies for the identification of proteinaceous binders in paintings. *Anal. Bioanal. Chem.* **395,** 2269–2280 (2009).
29. Vinciguerra, R., De Chiaro, A., Pucci, P., Marino, G. & Birolo, L. Proteomic strategies for cultural heritage: form bones to paintings. *Microchem. J.* **126,** 341–348 (2015).
30. Lemay, D. G. *et al.* The bovine lactation genome: insights into the evolution of mammalian milk. *Genome Biol.* **10,** 1–18 (2009).
31. Kawasaki, K., Lafont, A. G. & Sire, J. Y. The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol. Biol. Evol.* **28,** 2053–2061 (2011).
32. Colombini, M. P. and Modugno, F. *Organic Mass Spectrometry in Art and Archaeology*. *The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference.* John Wiley & Sons, Ltd.**1,** (2009).
33. Laboratory, F. P. Casein Glues : Their Manufacture , Preparation , and Application. *U. S. Dep. Agric. For. Serv. For. Prod. Lab. MADlSON, WIS* 1–17 (1967).
34. Stuart, C. & M. *Field Guide to Mammals of Southern Africa*. (1997).
35. Lombard, M. Finding resolution for the Howiesons Poort through the microscope: micro-residue analysis of segments from Sibudu Cave, South Africa. *J. Archaeol. Sci.* **35,** 26–41 (2008).
36. Lombard, M. The gripping nature of ochre: The association of ochre with Howiesons Poort adhesives and Later Stone Age mastics from South Africa. *J. Hum. Evol.* **53,** 406–419 (2007).
37. Wadley, L. Putting ochre to the test: replication studies of adhesives that may have been used for hafting tools in the Middle Stone Age. *J. Hum. Evol.* **49,** 587–601 (2005).
38. Wadley, L., Hodgskiss, T. & Grant, M. Implications for complex cognition from the hafting of tools with compound adhesives in the Middle Stone Age, South Africa. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 9590–9594 (2009).
39. Burger, P., Charrié-Duhaut, A., Connan, J. & Albrecht, P. Taxonomic characterisation of fresh Dipterocarpaceae resins by gas chromatography-mass spectrometry (GC-MS): Providing clues for identification of unknown archaeological resins. *Archaeol. Anthropol. Sci.* **3,** 185–200 (2011).
40. Wadley, L. Y. N. & Langejans, G. Preliminary Study of Scrapers Around Combustion Features in Layer Ss , Sibudu , 58 000 Years Ago. **69,** 19–33 (2014).
41. Wendt, W. E. 'Art Mobilier' from the Apollo 11 Cave, South West Africa: Africa's Oldest Dated Works of Art. *South African Archaeol. Bull.* **31,** 5–11 (1976).
42. Vogelsang, R. *Middle-Stone-Age-Fundstellen in Sidwest-Namibia*. (Heinrich-Barth-Institut, 1998).
43. Rudner, I. Paints of the Khoisan Rock Artists. *Goodwin Ser.* **4,** 14–20 (1983).

44. Arocena, J. M., Hall, K. & Meiklejohn, I. Minerals provide tints and possible binder/extender in pigments in san rock paintings (South Africa). *Geoarchaeology* **23,** 293–304 (2008).
45. Prinsloo, L. C., Tournié, A., Colomban, P., Paris, C. & Bassett, S. T. In search of the optimum Raman/IR signatures of potential ingredients used in San/Bushman rock art paint. *J. Archaeol. Sci.* **40,** 2981–2990 (2013).
46. Williamson, B. S. Direct Testing of Rock Painting Pigments for Traces of Haemoglobin at Rose Cottage Cave, South Africa. *J. Archaeol. Sci.* **27,** 755–762 (2000).
47. Rudner, I. *Khoisan Pigments and Paints and Their Relationship to Rock Paintings*. (South African Museum, 1982).
48. Vinciguerra, R. *et al.* Deglycosylation Step to Improve the Identification of Egg Proteins in Art Samples. *Anal. Chem.* 10178–10182 (2015). doi:10.1021/acs.analchem.5b02423
49. Lluveras, A., Bonaduce, I., Andreotti, A. & Colombini, M. P. 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. **82,** 376–386 (2010).

VI

A targeted proteomic approach for Cultural Heritage

The fascinating potential of standard proteomic strategies definitively lies in the possibility of uncovering novel, or at least unexpected proteins, in what can be called a "blind" analysis, without any a priori knowledge or hypothesis. As we already underlined, despite the tremendous success of these standard proteomics approaches in the field of cultural heritage, a principal disadvantage lies in the complexity of these samples, in terms of chemical composition, presence of contaminating protein-based materials, as well as the high levels of molecular damage found in ancient samples, that can often overwhelm even the most modern mass spectrometer.

As a result of the analytical disadvantages due to the complexity of unusual samples from cultural heritage, the development of a targeted MS method (Multiple Reaction Monitoring, MRM), to be applied to binders in paintings that are protein-based, is described in this chapter.

This method can be considered the first example of the application of targeted proteomics to samples in the field of cultural heritage that will allow an increase in sensitivity of protein detection in highly complex and rather contaminated samples.

To make this approach rational and efficient, we created a database of commonly used protein-based binders (milk, eggs and animal glue) from the results we have obtained in the years in our laboratory by the standard proteomic approaches that have been described in the previous chapters.

In particular, the set-up of the database has been an essential starting point for a statistical analysis aimed to identify marker peptides of specific proteins contained in each binder.

VI.1 Introduction

The most frequently used proteomics strategies rely on mass spectrometry technologies and can be broadly divided in two main categories: discovery and targeted proteomics. The most common MS-based proteomics pipelines aimed at identifying proteins follow the classical bottom up approach, where protein samples are enzymatically digested to peptides, which are then separated by liquid chromatography and eventually analyzed by tandem mass spectrometry. In discovery proteomics (or shotgun proteomics, global profiling proteomics),

peptides and thereof proteins are identified by correlating experimental spectra to virtual ones in a protein sequence database. In this operating mode, the mass spectrometer continuously repeats the full scan of the mass spectrum, and selects and fragments the *n* (typically *n* = 1-10) most abundant ions (information-dependent acquisition, IDA).

So far, to the best of our knowledge, all the applications to the cultural heritage field are discovery proteomics based experiments ([1] and references herein cited).

As we said, the fascinating potential of this operating mode lies in the possibility of uncovering novel, or at least unexpected proteins, in what can be called a "blind" analysis, without any "a priori" knowledge or hypothesis. It is therefore compatible with both well and poorly characterized systems in cultural heritage, the only requirements being the actual presence of proteins in the sample and the presence of the protein sequences in a properly selected database.

However, despite the tremendous success of these discovery, standard based proteomics approaches so far in the field of cultural heritage objects [2–12], a chief disadvantage lies in the automated peak selection and to the linked instrument's bias towards repeatedly selecting and fragmenting the species with the most intense MS signal: the complexity of these unusual samples, in terms of chemical composition, presence of contaminating protein-based materials, as well as the high levels of molecular damage found in ancient samples, can often overwhelm even the most modern mass spectrometer.

Moreover, a "blind" analysis is not always absolutely required in this kind of samples, where some hypotheses can in principle be worked out and the results are expected in a limited range of possibilities. This is for instance the case of proteinaceous binders in paintings, where generally the choices are among milk, egg or animal glue based binders. In such cases, the basic analytical challenge is the unequivocal discrimination, with the highest sensitivity, among a discrete number of possibilities, overcoming the unavoidable problems strictly connected to the peculiar nature of the samples: the unusual physical state, the complex chemical composition (i.e. proteins are only minor, although essential, component in the actual preparation of the colour, which is full of inorganic pigments), the unavoidable presence of contaminating proteins coming from the environmental dust, and the chemical modifications induced by aging processes.

In such cases, alternate strategies that allow the selectively recognition of protein components are welcome. Some interesting immunological methods, based on the antigen/antibody interaction, offer a technique with high specificity and sensitivity [13,14]. The application of immune-based techniques to the study of proteins in ancient samples dates back to the early

1960 [15], while, nowadays, the immune-based techniques are mainly proposed for imaging studies of ancient samples [16,17].

Alternatively, a mass spectrometric strategy, that selectively and non-redundantly detects the ions of one or, more desirably, a few peptides leading to unequivocally identify a protein among a restricted range of possibilities would thus represent a significant advance in the solution of the analytical challenge.

As an alternative to a discovery, data-dependent operation mode of the mass spectrometer, it is possible to operate in a targeted mode, by programming the instrument to continuously monitor selected ions that upon fragmentation generate specific species, and these ions transitions can be diagnostic of the presence of a specific peptide, and, thereof, can constitute a signature of the presence of a protein.

Targeted proteomics is recognized as the most sensitive and selective way to detect pre-selected components in a complex matrix such as a proteolytic digest of a plasma or tissue extract. Adoption of targeted mass spectrometry approaches such as Multiple Reaction Monitoring (MRM) to study biological and biomedical questions is well underway in the proteomics community, and is gradually complementing large shotgun studies with hypothesis-driven experiments [18–22], prompted by the growing demand for reliable quantification methods that could compete with immune-based strategies in terms of sensitivity, specificity and reliability, that however suffer for the lack of easy availability of antibodies with high specificity for any desired proteins.

In this context, the targeted MS methods (Multiple Reaction Monitoring, MRM, and Selected Reaction Monitoring, SRM, are terms that can be used interchangeably) can be set-up very rapidly, although the developing of a robust and precise assay can be quite demanding, the main challenge being the proper selection of the target peptides that would ensure reliable, high quality data.

There is no report so far of any application of targeted proteomics in the field of cultural heritage, that would be desirable to increase sensitivity of protein identification in highly complex and rather contaminated samples, beside paving the way to the fascinating perspective of quantitate proteins and/or, even more attractive, modifications on proteins in artworks.

In this chapter, we report, to the best of our knowledge, the first example of application of targeted proteomics to samples in the field of cultural heritage. Proteinaceous binders in painting are mainly derived from egg, milk casein, and animal glue made up of collagen, and constitute a perfect benchmark for testing the application of targeted proteomics to cultural

heritage samples. A targeted proteomics approach could rapidly answer with sensitivity comparable to immune-based strategies to the analytical question of identifying the type of binders in paintings, a critical aspect for art conservationists and restorers that can plan a proper restoration campaign on the basis of a scientific diagnosis. We therefore developed an MRM assay to detect and distinguish proteins in painting binders and tested it on samples that had already produced positive identifications and, later, on samples for which no protein identification was achieved in standard discovery proteomic analyses.

Selection of the proteotypic peptides, a critical step in the assay set up, was preliminarily carried out on an experimental basis, from the statistical analysis of the most recurrent peptides in a large dataset of samples that had been analyzed by classical discovery strategies.

Results herein presented demonstrate the improvement in selectivity and sensitivity achieved with the targeted approach in respect to classical profiling experiments and pave the way to possible application of this most advanced analytical strategy in the field of archaeology and artworks.

VI.2. Materials and Methods

VI.2.1. Painting samples

Replicas and historical samples analysed and used for database building consist of pigments and protein-based binders as reported in Table VI.1 (Appendix B). Aliquots of each sample were treated in heterogeneous phase with trypsin with the minimally invasive protocol either without any sample pre-treatment or with a particular pre-treatment as indicated in the chapters IV and V, and the resulting peptide mixtures were analysed by LC-MS/MS and/or MALDI-TOF[23,24,8,25,26].

In particular, the peptide mixture (1µl) was analyzed by nanoLC-MS/MS, on a CHIP MS 6520 QTOF equipped with a capillary 1200 HPLC system and a chip cube (Agilent Technologies, Palo Alto, CA) as indicated for proteomics analysis in the chapters IV and V. For some samples, however, the peptide mixtures were analyzed by Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) on a 4800 Plus MALDI TOF/TOF mass spectrometer (Applied Biosystems, Framingham, MA) equipped with a nitrogen laser (337 nm). The peptide mixture (1ml) was mixed (1:1, v/v) with a 10 mg/ml solution of a-cyano-4-hydroxycinnamic acid in acetonitrile/50mM citrate buffer (70:30 v/v). Mass calibration was performed using external peptide standards purchased from Applied Biosystems. Spectra were acquired using a mass (*m/z*) range of 400–4,000amu and raw data

were analyzed using Data Explorer Software provided by the manufacturer. Mass spectra were analysed by manual inspection against the theoretical list of peptides, providing a proteins identification.

VI.2.2. MS database building and statistical analysis

A collection of 116 mass spectrometry analyses, carried out on 107 independent samples, were screened to identify the proteins most frequently detected in standard global profiling experiments.

The choice of the collected samples was mainly based on the binder generally used in the past and the animal species of the substance used: milk and animal glue from *Bos taurus* and eggs from *Gallus gallus*.

The statistical analysis was performed by screening the samples proteins and peptides most frequently detected in standard global profiling experiments in a large dataset of samples. A relational database was designed on PivotTables (https://support.office.com) to simplify the data processing. Moreover, data evaluation was carried out regardless the treatment procedures they have undergone and the mass spectrometers that was used in the specific analysis.

VI.2.3. MRM assay design

The MRM assay requires the setup of specific mass spectrometry acquisition parameters that was carried out by Prof. Angela Amoresano and Dr. Anna Illiano.

Skyline Software drew the choice of specific transitions and parameters related to ionization and fragmentation for each analyte. Subsequently, LC-MRM/MS analysis was carried out using a Xevo TQ-S (Waters) equipped with an ionKey UPLC Microflow Source coupled to an UPLC Acquity System (Waters).

VI.3. Results and Discussion

The basic scheme of the development of a targeted proteomic assay starts with the selection of a target list of proteins, based on previous experiments, literature or knowledge. Then, target peptides that optimally represent the target proteins are selected, and this step is followed by the selection of a set of suitable transitions for each selected peptide. In our assay development, schematically described in FigureVI.1, we opted for an empirical identification of marker proteins and peptides based on our previous experience with painting samples. The

designed assay was then validated on samples that had already been characterized in profiling experiments and eventually tested on unknown ones. Moreover, we validated the use of an MRM assay for some specific requests that could be relevant in specific proteinaceous binders analyses such as species discrimination and identification of modifications.

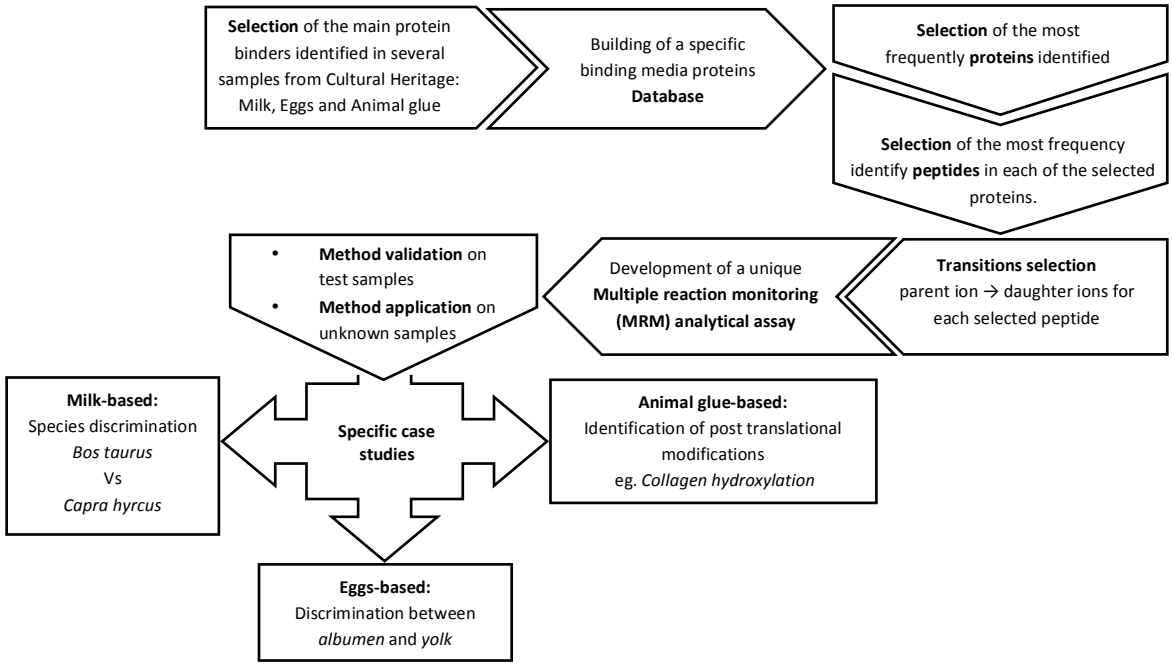


Figure VI.1. Diagram of the developed assay steps.

VI.3.1. Database assembling and proteins selection

Selection of diagnostic peptides that can constitute the signature of proteins and, in our case, thereof of a material, is the critical step in the experimental design of the assay because it will affect both the specificity and the sensitivity of the targeted analysis. Predictive models based on experimental data [27] have proven to be extremely powerful in the selection of proteotypic peptides in targeted proteomics, and several software packages exist that facilitate this step of the assay development [28]. However, among the factors that could affect the effective detection of a peptide, there is for sure the physical state of the sample. The peculiar, solid and aged nature of the samples we aim to develop the assay for, prompted us to approach the selection in a definitively empirical mode. We therefore screened the results obtained on the collection of painting samples that we have been analysing throughout years in our lab, including both paint replicas and historical samples, regardless the treatment procedures they have undergone and the mass spectrometers that have been used in the specific analysis, in the perspective of picking the most experimental procedure independent markers that could be used in any

laboratory and of minimizing the influence of sample pre-treatment procedures in the selection of the most influential and discriminative peptides.

Figure VI.2 (Appendix B)reports the summary of the results of the screening of the samples that contained milk (43 samples, detailed in Table VI.2 of Appendix B related to milk proteins): data clearly indicates that, whenever milk was used alpha S1 casein from *Bos taurus* (P02662), was identified, as could be expected considering its relative abundance [29]. Moreover, beta casein (P02666) (77%) and alpha S2 casein (P02663) (72%) could be detected in 33 samples and in 31 samples out of 43 total milk containing samples, respectively. When animal glue was used, (Figure VI.3 in Appendix B) COL1α1 (collagen alpha-1(I)) from *Bos taurus* (P02453) was detected in 29 samples out of 29 samples and COL1α2 (collagen alpha-2(I)) (P02465) in 28 samples. The case of egg containing samples is more complex, since we have to distinguish samples where only albumen was used from samples with only yolk and samples containing both yolk and albumen. It's worth to underline that while is technically quite easy to have "pure" albumen from an egg, the yolk is likely contaminated by the egg white. Therefore, in the sake of simplicity, on the basis of experiments carried out on eggs, that will be discussed below, we grouped the results obtained with the different egg containing samples (43), as reported in Figure VI.4 (Appendix B), according to the presence of albumen only (27 samples), and albumen plus yolk (16 samples), without considering yolk only containing samples separately, for which contamination from albumen proteins is very likely. As expected chicken ovalbumin (P01012), the main protein found in egg white, making up 60-65% of the total protein, is identified in almost all the samples: 38 out of 43 total samples and 27 out of 27 albumen only samples, accounting for 90% and 100%, respectively. Similarly, ovotransferrin (P02789) is frequently identified (29 out of 43 total samples, 67%, that increases up to 78% if we consider the albumen only samples, since it was identified in 21 out of 27 albumen only samples). Vitellogenin-2 (P02845) and apovitellenin-1 (P02659) are the most frequently identified proteins from yolk (both of them have been identified in 14 samples out of 15 yolk and albumen containing ones, counting for 87%). Moreover, Vitellogenin-1 (P87498) can be identified in 12 out of 15 samples (75%). It is worth noting that neither vitellogenin-2 nor apovitellenin-1 have been detected in any albumen only containing sample.

From the results presented, we selected alpha S1 casein, alpha S2 casein and beta casein as the representative proteins for milk based binder, collagen alpha-1(I) collagen alpha-2(I) as animal glue representative proteins, and ovalbumin and ovotranferrin as generic representative

of egg containing binder, while, more specifically, presence of vitellogenin-2, apovitellenin-1 and vitellogenin-1 can be used as indicator of yolk presence.

VI.3.2. Selection of the proteotypic peptides

Once we have identified the target proteins as the most frequently detected in the different binders, we deepened our analysis to identify the peptides to be used in the MRM assay. Our strategy of selection was essentially similar to that carried out in the selection of the proteins, i.e., we listed the peptides of the selected proteins that allowed identification, and screened the list for the most frequently observed ones. The frequency of observation was calculated from the detailed analysis reported in Table VI.2 for milk- and eggs-containing (Appendix B). The selection of the peptides was carried out with the same rules also for collagen-based samples, which is part of the research project of Dr. A. De Chiaro. We opted to select 3/4 target peptides per target protein.

It is worth mentioning that in most cases, we carried out both the MALDI-TOF and the LC-MSMS analyses on the same peptide mixtures, in order to evaluate the independence of the observation from the specific mass spectrometric analysis. However, frequency of observation has been calculated considering the detection of the specific peptide in any of the analyses only once per sample. Moreover, in the perspective of selecting specific ions to develop an appropriate MRM assay, modified peptides were considered separately.

VI.3.2.1. Milk proteotypic peptides

Figure VI.5 (Appendix B) reports the frequency of observation of the alpha S1 casein peptides in the milk containing samples. Peptide ^{106}YLGYLEQLLR^{115} of the sequence of alpha S1 casein from *Bos taurus*, has been observed in 37 samples out of 43, accounting for 86% of the cases. Peptide ^{38}FFVAPFPEVFGK^{49} has been identified in 34 samples, 79% of the cases. Peptide HQGLPQEVLNENLLR, position from 23 to 37, has been observed in 26 samples, accounting for 60% of the cases.

As shown in Figure VI.6 (Appendix B), in the case of alpha S2 casein, peptide ^{189}FALPQYLK^{196}, has been observed in 23 samples (53%), peptide ^{130}NAVPITPTLNR^{140} in 20 samples (46%) and its singly deamidated form in 19 samples (44%), while peptide ^{204}AMKPWIQPK^{212} in 15 samples (35%), and ^{96}ALNEINQFYQK^{106} in 17 samples (39%).

Moreover, always in Appendix B, Figure VI.7 reports the results for Beta casein. We could select peptide ^{192}AVPYPQR^{198} that was detected in 29 samples (67%), peptide

^{199}DMPIQAFLLYQEPVLGPVR^{217} identified in 19 samples (44%), and its oxidized form that was identified in 17 cases (39%).

VI.3.2.2. Eggs proteotypic peptides

The frequency of observed peptides for ovalbumin and ovotransferrin was calculated on 43 samples that containing albumen out of 51 samples analyzed. Figure VI.8 shown in Appendix B reports the frequency of observation of the ovalbumin peptides in the eggs containing samples. Peptide ^{265}LTEWTSSNVMEER^{277}, of the sequence of ovalbumin from *Gallus gallus*, has been observed in 25 samples out of 43, accounting for 57% of the cases. Peptide ^{128}GGLEPINFQTAADQAR^{143} has been identified in 31 samples, 73% of the cases. Peptides ^{324}ISQAVHAAHAEINEAGR^{340} and ^{144}ELINSWVESQTNGIIR^{159} has been observed in 28 samples, accounting for 65% of the cases.

In the case of ovotransferrin (Figure VI.9 in Appendix B), peptides ^{595}ANVMDYR^{601} and ^{120}GTEFTVNDLQGK^{131} has been observed in 22 samples (51%), while peptides ^{289}AQSDFGVDTK^{298}, ^{275}DDNKVEDIWSFLSK^{288} and ^{155}GAIEWEGIESGSVEQAVAK^{173} in 23 samples (53%) and peptide ^{141}SAGWNIPIGTLLHR^{154} in 25 samples (57%).

In a similar way, the frequency of observed peptides for the yolk proteins (vitellogenin-1 and vitellogenin-2) was calculated on 15 total samples. Figure VI.10 reports the frequency of observation for vitellogenin-1.

The peptides ^{831}LTELLNSNVR^{840}, ^{472}SNIEEVLLALK^{482}, ^{365}YLLDLPAAASHR^{377}, ^{315}LQDLVETTYEQLPSDAPAK^{333}, and the oxidized peptide ^{674}VADPIE-VGIAAEGLQEMFVR^{693} has been detected in 4 out of 15 total samples (28%). Peptide ^{813}VAGNVQAITPSPR^{826}, instead, has been identified in 6 samples, accounting for 40% of the cases.

In the case of vitellogenin-2 (Figure VI.11 in Appendix B), peptides ^{1515}MVVALTSPR^{1523} and ^{456}EALQPIHDLADEAISR^{471}, has been observed in 10 out of 14 samples (67%), peptide in 20 samples (46%), peptide ^{260}QQLTLVEVR^{268} and ^{642}VGATGEIFWNSPR^{655} in 12 samples (80%) in 17 samples (39%). We could select peptides ^{226}QSDSGTLITDVSSR^{239} and ^{240}QVYQISPENEPTGVAVMEAR^{259} in 11 samples (73%), while ^{919}NIGELGVEKR^{928} and ^{1543}LPLSLPVGPR^{1552} in 7 and 8 samples equal to 47% and 53% respectively.

The selection of more than 3/4 proteotypic peptides for ovotransferrin, vitellogenin-1 and vitellogenin-2 was determined by the observation of a more heterogeneous distribution of frequencies in respect to other proteins. This result is related to the fact that differently from

milk containing samples, identification of eggs proteins can differ in terms of peptides from samples to samples, making the selection of the marker peptides more complicated.

VI.3.3. MRM Analysis

The development of the MRM assay required the setup of acquisition parameters that was carried out by Prof. Angela Amoresano and Dr. Anna Illiano, Department of Chemical Sciences, University of Naples 'Federico II'.

Taking advantage of the software Skyline, the developed MRM assay was used to detect and distinguish proteins in painting binders and tested in a first instance on samples that had already produced positive identifications and later on samples for which no protein identification was achieved in discovery proteomic analyses. Table VI.5 (Appendix B) reports the proteotypic peptides for each selected protein, the selected transitions and the corresponding calculated collision energy.

On pictorial specimens containing milk as a binder, the MRM analysis showed a positive result in the identification of all selected peptides to support thereby the validity of the method. Subsequently the MRM method was also tested on ancient samples previously analyzed by LC-MS/MS, and whose results are reported and discussed in the chapter V. In standard analyses, we observed a mixture of cow and goat milk in 16 out of 30 fragments of decorations of the Buddha's statues. With the MRM assay, we obtained positive results for all the 30 samples. As an example, in Figure VI.12 are shown the transitions related to the peptide YLGYLEQLLR of alpha S1 casein in one of the Buddha's sample that had not produced identification in a standard analysis carried out as indicated in the Chapter V. The different transitions perfectly co-eluted at retention time of 20 min, thus demonstrating the identification of the target protein and then of the specific binder with high sensitivity.

In a similar way, the MRM assay was tested on some samples whose historical references and artistic investigations indicated the presence of egg- and animal glue-based binders. Preliminary results on samples containing egg showed positive outcome: the ovalbumin and ovotransferrin pre-selected peptides have been identified in pictorial specimens that had produced positive and negative identification in discovery proteomics analysis. Conclusive analyses are still in progress.

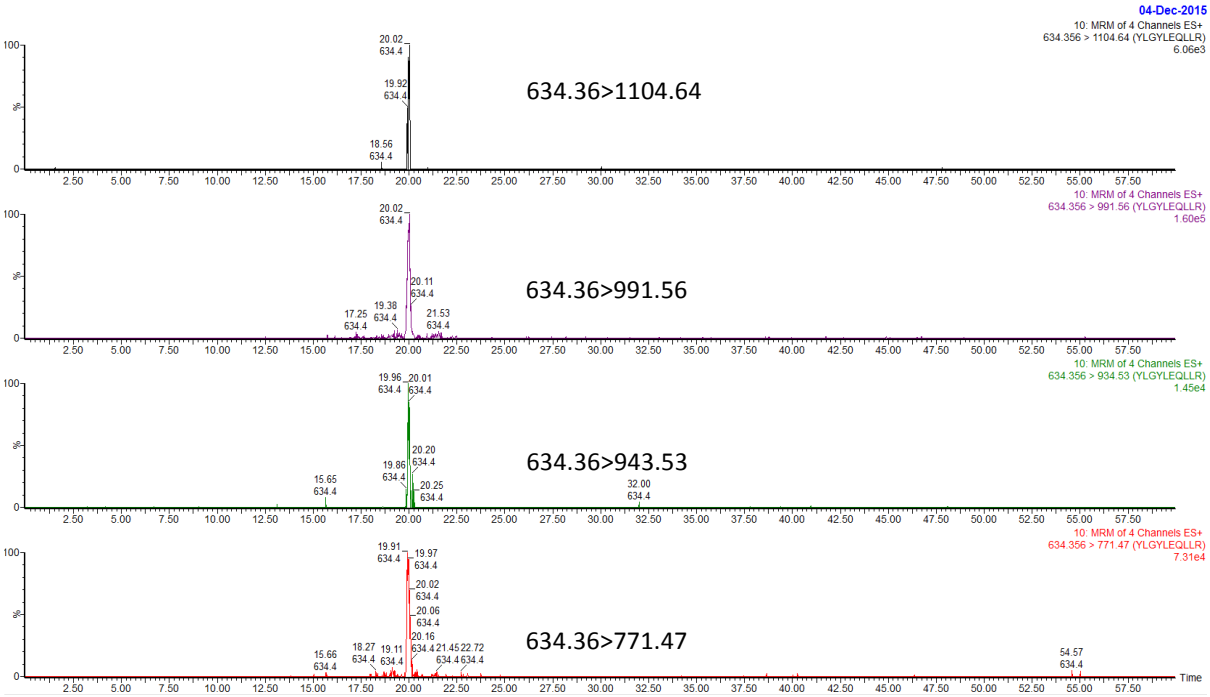


Figure VI.12. MRM traces chromatogram of four transitions (644.36>771.47; 644.36>943.53; 644.36>991.56; 644.36>1104.64) related to YLGYLEQLLR peptide of alpha s1 casein that co-eluted at 20 min. The analysis was carried out in one of the 14 Buddha's fragments for which no protein identification was achieved in discovery proteomic analysis (see chapter V).

VI.3.3.1. Specific applications

The same MRM procedure as described above can also be used to answer to specific requests in addition to the identification of the type of binder. For instance, a specific request in case of milk binders can be the discrimination of the milk origin. Among the peptides that we selected for the identification of milk, some can be used to distinguish the animal origin at different levels. The "best" observed milk peptide, YLGYLEQLLR from alpha S1 casein, for instance, is not suitable for specie discrimination, since it is shared by several alpha S1 caseins from different mammals origin, all from the Bovidae family, namely *Bos taurus*, *Bos mutus*, *Bubalus bubalis*, *Ovis aries* and *Capra hircus* (Figure VI.13), similarly to what can be observed with AMKPWIQPK and ALNEINQFYQK peptides from alpha S2 casein, and DMPIQAFLLYQEPVLGPVR peptide from beta casein (Figure VI.14). Conversely, FFVAPFPEVFGK is specific of Bovinae subfamily, since it is shared by alpha S1 casein from *Bos taurus* and *Bos mutus* (*Bos* genus) but also by alpha S1 casein from *Bubalus bubalis* (*Bubalus* genus), as peptide NAVPITPTLNR from alpha S2 casein, and peptide AVPYPQR from beta casein. HQGLPQEVLNENLLR is instead specific of *Bos* genus, as peptide FALPQYLK from alpha S2 casein.

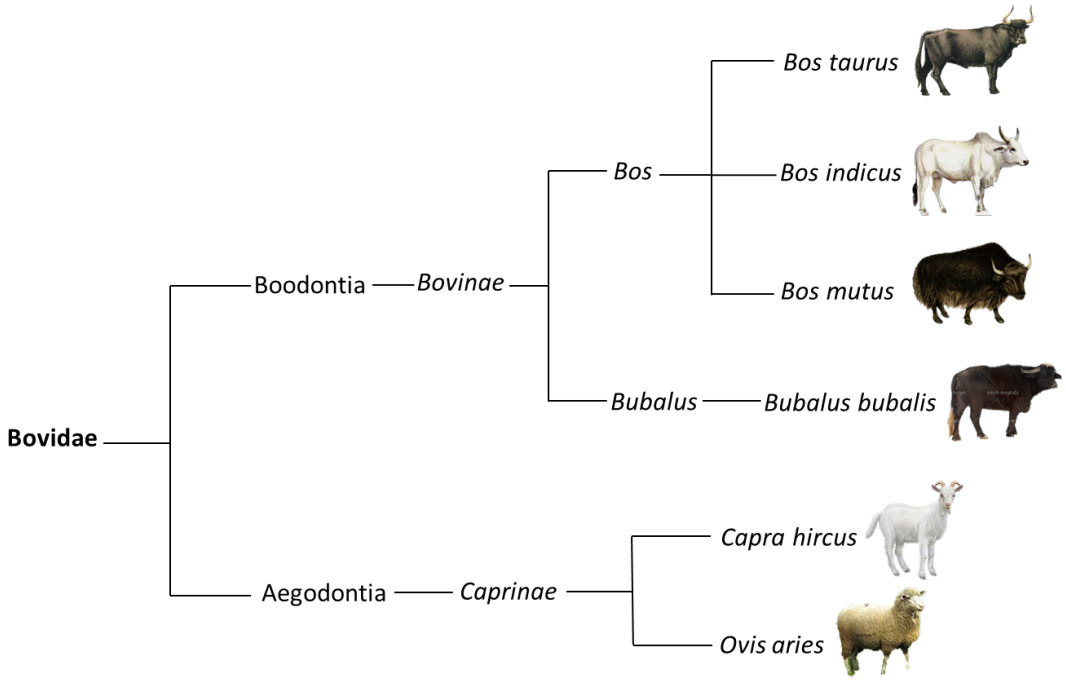


Figure VI.13. A portion of phylogenetic tree of the Bovidae family. Subfamilies are typically arranged into two clades: Boodontia (containing the single subfamily *Bovinae*), and Aegodontia (containing the other seven subfamilies).

While in discovery proteomics identification of more than one peptides of the single proteins helps a lot in assigning the matches to a specie by the so-called principal of parsimony, in MRM operation mode peptides should be selected properly in the assay development, if we want to answer questions like that.

An example is given by the species-specific issue that has been addressed in the chapter V. In particular, milk proteins from *Bos taurus* and *Chapra hircus* were identified in different fragments recovered from the decorations of two statues of Buddha from the Bamiyan Valley in Afghanistan, and alignment tools allowed us to identify peptides that may be considered species specific and that will therefore discriminate milks.

We selected FFVAPFPEVFGK and FVVAPFPEVFR peptides of alpha S1 casein and tested the assay on the sample of Buddha's decoration that was negative in standard discovery analysis (Figure VI.15). Identification of the pre-selected four perfectly co-eluted transitions for both of peptides, clearly demonstrates the presence of *Bos taurus* and *Capra hircus* milk in mixture, successful obtained by the improved sensitivity of the MRM assay.

Although still in the process of optimization, the MRM method will also be used to obtain various supporting information in the case of eggs and animal glue. In the case of egg-based binders it could be used to answer the question whether albumen or yolk has been used in

different ancient paintings; and in the case of animal glue-based binders, it could be used in the identification of the hydroxylation sites, very useful to evaluate the conservation state of the cultural heritage objects.

This demonstrates that MRM analysis can also be used for answering to questions more specific than simply identificative purposes.

Moreover, the improvement in selectivity and sensitivity achieved with the targeted approach in respect to classical profiling experiments could pave the way to possible application of this most advanced analytical strategy in the field of archaeology and artworks.

```
                          ----------------------YLGYLEQLLR-----
TR|L8I5S0|L8I5S0_9CETA  SEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYK 120
SP|P04653-1|CASA1_SHEEP SEEIVPNSAEQKYIQKEDVPSERYLGYLEQLLRLKKYN 120
SP|P18626|CASA1_CAPHI   SEEIVPNSAEQKYIQKEDVPSERYLGYLEQLLRLKKYN 120
SP|P04653|CASA1_SHEEP   SEEIVPNSAEQKYIQKEDVPSERYLGYLEQLLRLKKYN 120
SP|P02662|CASA1_BOVIN   SEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRLKKYK 120
SP|O62823|CASA1_BUBBU   SEEIVPISVEQKHIQKEDVPSERYLGYLEQLLRLKKYN 120
                                           **********

                          ---------------FFVAPFPEVFGK-----------
TR|L8I5S0|L8I5S0_9CETA  HQGLPQEVLNENLLRFFVAPFPEVFGKEKVNELSKDIG 60
SP|P02662|CASA1_BOVIN   HQGLPQEVLNENLLRFFVAPFPEVFGKEKVNELSKDIG 60
SP|O62823|CASA1_BUBBU   HQGLPQGVLNENLLRFFVAPFPEVFGKEKVNELSTDIG 60
                                    ************

                          ---------HQGLPQEVLNENLLR---------------
TR|L8I5S0|L8I5S0_9CETA LARPKHPIKHQGLPQEVLNENLLRFFVAPFPEVFGKDIG 60
SP|P02662|CASA1_BOVIN  LARPKHPIKHQGLPQEVLNENLLRFFVAPFPEVFGKDIG 60
                               ***************

                          --------------FALPQYLK-----------------
TR|L8I6J3|L8I6J3_9CETA RLNFLKKISQRYQKFALPQYLKTVYQHQKAMKPWIQPKT 213
SP|P02663|CASA2_BOVIN  RLNFLKKISQRYQKFALPQYLKTVYQHQKAMKPWIQPKT 213
                                  ********

                          ---------NAVPITPTLNR-------------------
TR|Q3Y443|Q3Y443_BUBBU LNPWDQVKRNAVPITPTLNREQLSTSEENSKKTVDMEST 159
TR|L8I6J3|L8I6J3_9CETA LNPWDQVKRNAVPITPTLNREQLSTSEENSKKTVDMEST 159
SP|P02663|CASA2_BOVIN  LNPWDQVKRNAVPITPTLNREQLSTSEENSKKTVDMEST 159
                                   ***********

                          -------AMKPWIQPK-----------------------
TR|L8I6J3|L8I6J3_9CETA TVYQHQKAMKPWIQPKTKVIPY----------------- 219
SP|P02663|CASA2_BOVIN  TVYQHQKAMKPWIQPKTKVIPYVRYL------------- 222
TR|Q3Y443|Q3Y443_BUBBU TVYQYQKAMKPWTQPKTKVIPYVRYL------------- 222
SP|P33049|CASA2_CAPHI  TVDQHQKAMKPWTQPKTNAIPYVRYL------------- 223
SP|P04654|CASA2_SHEEP  TVDQHQKAMKPWTQPKTNAIPYVRYL------------- 223
                                 *********

                          ----------------------ALNEINQFYQK------
TR|Q3Y443|Q3Y443_BUBBU EESAEVATEEVKITVDDKHYQKALNEINQFYQKFPQYLQ 112
TR|L8I6J3|L8I6J3_9CETA EESAEVATEEVKITVDDKHYQKALNEINQFYQKFPQYLQ 112
SP|P33049|CASA2_CAPHI  EESAEVAPEEIKITVDDKHYQKALNEINQFYQKFPQYLQ 113
SP|P04654|CASA2_SHEEP  EESAEVAPEEVKITVDDKHYQKALNEINQFYQKFPQYLQ 113
SP|P02663|CASA2_BOVIN  EESAEVATEEVKITVDDKHYQKALNEINQFYQKFPQYLQ 112
                                           ***********
```

```
                          -----AVPYPQR---------------------------
TR|U5Y4C2|U5Y4C2_BOSIN PVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFP---- 154
TR|L8I8G5|L8I8G5_9CETA PVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV- 224
TR|F5BHA8|F5BHA8_BOSMU PVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIFVS 155
SP|Q9TSI0|CASB_BUBBU   PVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV- 224
SP|P02666|CASB_BOVIN   PVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV- 224
                            *******

                          ----------------DMPIQAFLLYQEPVLGPVR---
TR|U5Y4C2|U5Y4C2_BOSIN QSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPF 153
TR|L8I8G5|L8I8G5_9CETA QSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPF 220
TR|F5BHA8|F5BHA8_BOSMU QSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPF 150
SP|Q9TSI0|CASB_BUBBU   QSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPF 220
SP|P33048|CASB_CAPHI   QPKVLPVPQKAV--PQRDMPIQAFLLYQEPVLGPVRGPF 218
SP|P11839|CASB_SHEEP   QPKVLPVPQKAV--PQRDMPIQAFLLYQEPVLGPVRGPF 218
SP|P02666|CASB_BOVIN   QSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPF 220
                                     ******************
```

Figure VI.14. **Alignment of the selected peptides (in bold) of alpha s1 casein (P02662), alpha s2 casein (P02663) and beta casein (P02666) chains from *Bos taurus*.** A Blast search in mammalia was carried out with the single peptides to detected proteins that shared them and subsequently multiple sequence alignment was carried out with Clustal O(1.2.0) with alpha S1 casein from *Bos taurus*, *Bos mutus*, *Bubalus bubalis*, *Ovis aries* and *Capra hircus*. An * (asterisk) indicates positions which have a single, fully conserved residue.

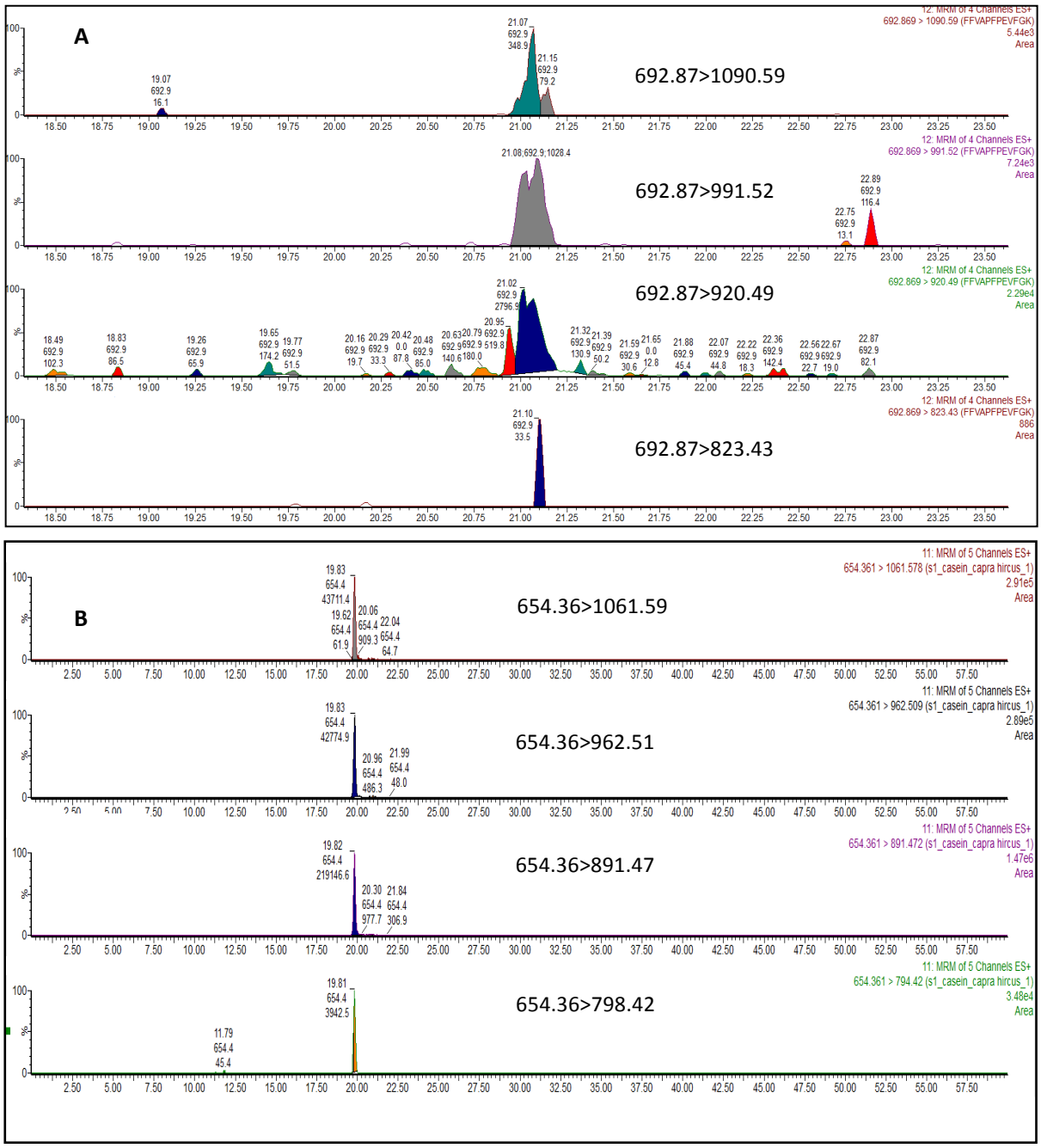


Figure VI.15. MRM traces chromatogram. Four pre-selected transitions (692.87>823. 692.8743; 692.87>920.49; 692.87>991.52; 692.87>1090.59) of FFVAPFPEVFGK peptide related to alpha s1 casein from *Bos taurus* (A) and transitions (654.36>798.42; 654.36>891.47; 654.36>962.51; 654.36>1061.59) of FVVAPFPEVFR peptide from *Capra hircus* (B) co-eluted at 21 and 19 min, respectively.

VI.4. Conclusions

Diagnostic methods represent a crucial aspect of the scientific investigation of artworks and proteomics procedures are increasingly applied for the identification of protein binders in samples from archeological objects and works of art. Proteomics strategies applied to artworks has begun to be accepted as the gold standard analytical technique when proteins have to be identified and characterized because of the sensitivity, the capability of identifying specie-specific proteins and detecting degradation processes.

Ongoing developments in the analysis of Cultural Heritage samples, in terms of adaptation of well-established proteomics protocols to the peculiar requirements of studying the protein-based binders, are rapidly putting proteomics in a worthy position in the wide panorama of studies that include archaeology, paleontology and molecular evolution for any kind of cultural heritage that we may think of.

However, methodological adaptations to specific analytical problems of samples from cultural heritage still need to be developed to tailor appropriate approaches.

Highly complex and rather contaminated samples require high selectivity and sensitivity methods to obtain good protein detection and to answer to more specific questions.

Based on these demands, a targeted MS method (multiple reaction monitoring, MRM) was developed and successfully applied to cultural heritage samples. A method based on these analytical criteria can be considered the first example of the application of targeted proteomics to samples in the field of cultural heritage.

A MS peptidic database of commonly used protein-based binders was created on the base of results obtained in the years by the standard proteomic approaches suitably adapted.

As an essential part for the developed assay, peptides collected in the database were subjected to statistical analysis in order to select diagnostic peptides, which constitute the signature of the proteins and consequently of a material used as binder.

The assay was used not only to detect and distinguish proteins in historical painting binders but also to answer to specific requests. For instance, a specific request, in case of milk binders, was the discrimination of milk originating from different animal sources in some Buddha's decoration fragments that had not produced identification in a standard analysis.

VI.5. References

1. Dallongeville, S., Garnier, N., Rolando, C. & Tokarski, C. Proteins in Art, Archaeology, and Paleontology: From Detection to Identification. *Chem. Rev.* acs.chemrev.5b00037 (2015). doi:10.1021/acs.chemrev.5b00037
2. Tokarski, C., Martin, E., Rolando, C., Post-traductionnelles, M. & Cedex, A. Identification of Proteins in Renaissance Paintings by Proteomics. *Anal. Chem.* **78,** 1494–1502 (2006).
3. Dallongeville, S. *et al.* Proteomics applied to the authentication of fish glue: application to a 17th century artwork sample. *Analyst* **138,** 5357–5364 (2013).
4. Dallongeville, S. *et al.* Dealing with the identification of protein species in ancient amphorae. *Anal. Bioanal. Chem.* **399,** 3053–3063 (2011).
5. Solazzo, C., Fitzhugh, M. W. W., Rolando, C. & Tokarskl, C. Identification of Protein Remains in Archaeological Potsherds by Proteomics. **80,** (2008).
6. Dallongeville, S. *et al.* Identification of animal glue species in artworks using proteomics: Application to a 18th century gilt sample. *Anal. Chem.* **83,** 9431–9437 (2011).
7. Chambery, A. *et al.* Improved procedure for protein binder analysis in mural painting by LC-ESI/Q-q-TOF mass spectrometry: Detection of different milk species by casein proteotypic peptides. *Anal. Bioanal. Chem.* **395,** 2281–2291 (2009).
8. Leo, G. *et al.* Proteomic strategies for the identification of proteinaceous binders in paintings. *Anal. Bioanal. Chem.* **395,** 2269–2280 (2009).
9. Fremout, W. *et al.* Development of a dedicated peptide tandem mass spectral library for conservation science. *Anal. Chim. Acta* **728,** 39–48 (2012).
10. Fremout, W. *et al.* Tryptic peptide analysis of protein binders in works of art by liquid chromatography-tandem mass spectrometry. *Anal. Chim. Acta* **658,** 156–162 (2010).
11. Rao, H., Yang, Y., Abuduresule, I., Li, W. & Hu, X. Proteomic identi fi cation of adhesive on a bone sculpture-inlaid wooden artifact from the Xiaohe Cemetery , Xinjiang , China. *J. Archaeol. Sci.* **53,** 148–155 (2015).
12. Villa, P. *et al.* A Milk and Ochre Paint Mixture Used 49,000 Years Ago at Sibudu, South Africa. *PLoS One* **10,** e0131273 (2015).
13. Palmieri, M., Vagnini, M., Pitzurra, L., Brunetti, B. G. & Cartechini, L. Identification of animal glue and hen-egg yolk in paintings by use of enzyme-linked immunosorbent assay (ELISA). *Anal. Bioanal. Chem.* **405,** 6365–6371 (2013).
14. Gambino, M. *et al.* A simple and reliable methodology to detect egg white in art samples. *J. Biosci.* **38,** 397–408 (2013).
15. Jones, P. L. Some Observations on Methods for Identifying Proteins in Paint Media. *Stud. Conserv.* **7,** 10–16 (1962).
16. Dolci, L. S., Sciutto, G., Guardigli, M., Rizzoli, M. & Prati, S. Ultrasensitive chemiluminescent immunochemical identification and localization of protein components in painting cross-sections by microscope low-light imaging. 29–35 (2008). doi:10.1007/s00216-008-2023-y
17. Sciutto, G., Dolci, L. S., Guardigli, M. & Zangheri, M. Single and multiplexed immunoassays for the chemiluminescent imaging detection of animal glues in historical paint cross-sections. 933–940 (2013). doi:10.1007/s00216-012-6463-z
18. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Resource Full Dynamic Range Proteome Analysis of S . cerevisiae by Targeted Proteomics. *Cell* **138,** 795–806 (2009).
19. Pan, S. *et al.* Mass spectrometry based targeted protein quantification: methods and applications. *J Proteome Res* **8,** 787–797 (2009).
20. Kiyonami, R. *et al.* Increased Selectivity, Analytical Precision, and Throughput in Targeted Proteomics. *Mol. Cell. Proteomics* **10,** M110.002931–M110.002931 (2011).

21. Carr, S. A. *et al.* Targeted Peptide Measurements in Biology and Medicine: Best Practices for Mass Spectrometry-based Assay Development Using a Fit-for-Purpose Approach. *Mol. Cell. Proteomics* **13,** 907–917 (2014).
22. Picotti, P. & Aebersold, R. Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9,** 555–566 (2012).
23. Vinciguerra, R., De Chiaro, A., Pucci, P., Marino, G. & Birolo, L. Proteomic strategies for cultural heritage: form bones to paintings. *Microchem. J.* **126,** 341–348 (2015).
24. Vinciguerra, R. *et al.* Deglycosylation Step to Improve the Identification of Egg Proteins in Art Samples. *Anal. Chem.* 10178–10182 (2015). doi:10.1021/acs.analchem.5b02423
25. Leo, G. *et al.* Deamidation at asparagine and glutamine As a major modification upon deterioration/aging of proteinaceous binders in mural paintings. *Anal. Chem.* **83,** 2056–2064 (2011).
26. Bonaduce, I., Colombini, M. P. & Diring, S. Identification of garlic in old gildings by gas chromatography-mass spectrometry. *J. Chromatogr. A* **1107,** 226–232 (2006).
27. Mallick, P. *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. **25,** 125–131 (2007).
28. Colangelo, C. M., Chung, L., Bruce, C. & Cheung, K. Review of software tools for design and analysis of large scale MRM proteomic datasets. *Methods* **61,** 287–298 (2013).
29. D, P. C. P. P. Milk nutritional composition and its role in human health. *NUT* **30,** 619–627 (2014).

Appendix A

Table V.1.2. Identification of the proteins in the samples by LC-MSMS after trypsin digestion in heterogeneous phase. Proteins were identified searching UniprotSprot database with MSMS Ion search Mascot software (Matrix Science) with Chordata as taxonomy restriction. with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications, and a peptide mass tolerance of 10 ppm. Individual ions scores > 32 indicate identity or extensive homology ($p<0.05$). Only proteins identified with at least two peptides were considered as significative. Peptide sequences are given with preceding and following residues. Underlined peptide sequences are unique to the specific species, while the others are in common between homologous proteins.

SAMPLE 206-4-3			
Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S1-casein *Bos taurus* (P02662)	**21**	27 23 13	R.YLGYLEQLLR.L + Deamidated (NQ) R.FFVAPFPEVFGK.E K.HQGLPQEVLNENLLR.F + 3 Deamidated (NQ)
SAMPLE 206-6-7-5-4			
Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S2-casein *Capra hircus* (P33049)	**13**	20 34 13 17 20	K.TNAIPYVR.Y + Deamidated (NQ) R.NAGPFTPTVNR.E + Deamidated (NQ) R.NAGPFTPTVNR.E + 2 Deamidated (NQ) R.NANEEEYSIR.S R.NANEEEYSIR.S + Deamidated (NQ)
SAMPLE 188			
Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S1-casein *Bos taurus* (P02662) or *Bubalis bubalis* (O62823)	**13**	24 28 22 17	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FFVAPFPEVFGK.E
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	24 28 22 12	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FVVAPFPEVFR.K
SAMPLE 235-grey-blue			
Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**

Alpha-S1-casein *Bos taurus* (P02662) or *Bubalis bubalis* (O62823)	**13**	24 19 18 8	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R.FFVAPFPEVFGK.E
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	24 19 18 20	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R.FVVAPFPEVFR.K

SAMPLE 235-bulk

Identified protein (Accession number)	Sequence coverage (%)	Individual Ion Score	Matched sequence
Alpha-S1-casein *Bos taurus* (P02662)	**27**	16 19 28 28 21 17 16 47	K.EKVNELSK.D R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FFVAPFPEVFGK.E K.VPQLEIVPNSAEER.L K.VPQLEIVPNSAEER.L + Deamidated (NQ) K.VPQLEIVPNSAEER.L + 2 Deamidated (NQ) K.HQGLPQEVLNENLLR.F + Deamidated (NQ)
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**9**	19 28 19	R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FVVAPFPEVFR.K
Alpha-S2-casein *Bos taurus* (P02663)	**26**	25 31 19 24 16 11 16	K.FALPQYLK.T K.FALPQYLK.T + Deamidated (NQ) R.NAVPITPTLNR.E R.NAVPITPTLNR.E + Deamidated (NQ) R.EQLSTSEENSK.K K.ALNEINQFYQK.F + Deamidated (NQ) K.TVDMESTEVFTK.K + Oxidation (M)
Beta-casein *Bos taurus*(P02666) or *Bubalis bubalis* (Q9TSI0)	**17**	14 12 18 20 16 20 24	K.EMPFPK.Y + Oxidation (M) K.VLPVPQK.A K.AVPYPQR.D K.AVPYPQR.D + Deamidated (NQ) R.DMPIQAFLLYQEPVLGPVR.G + Oxidation (M) R.DMPIQAFLLYQEPVLGPVR.G +Oxidation(M); Deamidated (NQ) R.DMPIQAFLLYQEPVLGPVR.G +Oxidation(M); 2 Deamidated (NQ)
Beta-lactoglobulin *Bos taurus* (P02754) or *Bubalis bubalis* (P02755)	**19**	15 57 31	K.IDALNENK.V + Deamidated (NQ) K.VLVLDTDYK.K R.TPEVDDEALEK.F

SAMPLE 235--grey-blue II

Identified protein (Accession number)	Sequence coverage (%)	Individual Ion Score	Matched sequence

Alpha-S1-casein *Bos taurus* (P02662) or *Bubalis bubalis* (O62823)	**13**	40 32 33 26	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FFVAPFPEVFGK.E
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	40 32 33 15	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FVVAPFPEVFR.K
Alpha-S2-casein *Bos taurus* (P02663)	**10**	27 30 14	R.NAVPITPTLNR.E R.NAVPITPTLNR.E + Deamidated (NQ) R.NAVPITPTLNR.E + 2 Deamidated (NQ)
Beta-casein *Bos taurus*(P02666) or *Bubalis bubalis* (Q9TSI0)	**5**	16 10 14	K.EMPFPK.Y + Oxidation (M) K.AVPYPQR.D K.AVPYPQR.D + Deamidated (NQ)

SAMPLE 235-black

Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S1-casein *Bos taurus* (P02662) or *Bubalis bubalis* (O62823)	**13**	40 18 30 23	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R.FFVAPFPEVFGK.E
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	40 18 30 23	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R.FVVAPFPEVFR.K

SAMPLE 214-blue

Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S1-casein *Bos taurus* (P02662)	**20**	27 30 38 18 16 26	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R.FFVAPFPEVFGK.E K.HQGLPQEVLNENLLR.F K.HQGLPQEVLNENLLR.F + Deamidated (NQ)
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	27 30 38 17	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FVVAPFPEVFR.K
Alpha-S2-casein *Bos taurus* (P02663)	**13**	14 21 12	K.FALPQYLK.T + Deamidated (NQ) R.NAVPITPTLNR.E R.NAVPITPTLNR.E + Deamidated (NQ)
Beta-casein *Bubalis bubalis* (Q9TSI0)	**30**	13 20 14	K.VLPVPQK.A K.AVPYPQR.D K.AVPYPQR.D + Deamidated (NQ)

		22	R.DMPIQAFLLYQEPVLGPVR.G+2 Deamidated (NQ); Oxidation (M)
Beta-lactoglobulin ***Bos taurus* (P02754)** **or *Bubalis bubalis*** **(P02755)**	**15**	32 39	K.VLVLDTDYK. K R.TPEVDDEALEK.F

SAMPLE 18-2

Identified protein (Accession number)	Sequence coverage (%)	Individual Ion Score	Matched sequence
Alpha-S1-casein ***Bos taurus* (P02662)** **or *Bubalis bubalis*** **(O62823)**	**8**	42 17	K.EDVPSER.Y R.FFVAPFPEVFGK.E
Beta-casein ***Bos taurus*(P02666)** **or *Bubalis bubalis*** **(P02755)**	**6**	25 18	K.VLPVPQK.A K.AVPYPQR.D

SAMPLE 214-sup

Identified protein (Accession number)	Sequence coverage (%)	Individual Ion Score	Matched sequence
Alpha-S1-casein ***Bos taurus* (P02662)** **or *Bubalis bubalis*** **(O62823)**	**13**	30 34 10	R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FFVAPFPEVFGK.E
Alpha-S1-casein ***Capra hircus* (P18626)** **or *Ovis aries* (P04653)**	**13**	30 34 20	R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FVVAPFPEVFR.K
Alpha-S2-casein ***Bos taurus* (P02663)**	**8**	11 24 13 15	K.FALPQYLK.T + Deamidated (NQ) R.NAVPITPTLNR.E R.NAVPITPTLNR.E + Deamidated (NQ) R.NAVPITPTLNR.E + 2 Deamidated (NQ)
Beta-casein ***Bos taurus*(P02666)** **or *Bubalis bubalis*** **(Q9TSI0)**	**5**	19 17 16	K.EMPFPK.Y + Oxidation (M) K.AVPYPQR.D K.AVPYPQR.D + Deamidated (NQ)
Beta-lactoglobulin ***Bos taurus* (P02754)** **or *Bubalis bubalis*** **(P02755)** **or *Ovis aries* (P67976)**	**11**	32 15	K.VLVLDTDYK.K R.TPEVDDEALEK.F

SAMPLE 18-clay

Identified protein (Accession number)	Sequence coverage (%)	Individual Ion Score	Matched sequence

Alpha-S1-casein *Bos taurus* (P02662) or *Bubalis bubalis* (O62823)	**8**	20 17	K.EDVPSER.Y R.FFVAPFPEVFGK.E

SAMPLE 22-2

Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S1-casein *Bos taurus* (P02662) or *Bubalis bubalis* (O62823)	**8**	28 16	K.EDVPSER.Y R.FFVAPFPEVFGK.E

SAMPLE 16—b

Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S1-casein *Bos taurus* (P02662) or *Bubalis bubalis* (O62823)	**8**	28 11	K.EDVPSER.Y R.FFVAPFPEVFGK.E
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	28 18	K.EDVPSER.Y R. FVVAPFPEVFR.K

SAMPLE 235—clay1

Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S1-casein *Bos taurus* (P02662) or *Bubalis bubalis* (O62823)	**13**	17 31 39 17	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FFVAPFPEVFGK.E
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	17 31 39 11	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R. FVVAPFPEVFR.K

SAMPLE 16-a

Identified protein (Accession number)	**Sequence coverage** (%)	**Individual Ion Score**	**Matched sequence**
Alpha-S1-casein *Bos taurus* (P02662)	**13**	30 16 16	K.EDVPSER.Y R.YLGYLEQLLR.L R.FFVAPFPEVFGK.E
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	30 16 23	K.EDVPSER.Y R.YLGYLEQLLR.L R.FVVAPFPEVFR.K

Beta-casein *Bos taurus*(P02666)	**5**	14 11 16	K.VLPVPQK.A K.AVPYPQR.D K.AVPYPQR.D + Deamidation (NQ)

SAMPLE 214-clay

Identified protein (Accession number)	Sequence coverage (%)	Individual Ion Score	Matched sequence
Alpha-S1-casein *Bos taurus* (P02662)	**20**	44 18 13 29 20 18	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R.FFVAPFPEVFGK.E K.VPQLEIVPNSAEER.L + Deamidated (NQ) K.VPQLEIVPNSAEER.L + 2 Deamidated (NQ)
Alpha-S1-casein *Capra hircus* (P18626) or *Ovis aries* (P04653)	**13**	44 18 13 35	K.EDVPSER.Y R.YLGYLEQLLR.L R.YLGYLEQLLR.L + Deamidated (NQ) R.FVVAPFPEVFR.K
Alpha-S2-casein *Bos taurus* (P02663)	**21**	10 22 16 27 20 20 17 12 15	K.VIPYVR.Y K.FALPQYLK.T K.FALPQYLK.T + Deamidated (NQ) R.NAVPITPTLNR.E R.NAVPITPTLNR.E + Deamidated (NQ) R.NAVPITPTLNR.E + 2 Deamidated (NQ) K.ALNEINQFYQK.F + Deamidated (NQ) K.ALNEINQFYQK.F + 2 Deamidated (NQ) K.TVDMESTEVFTK.K + Oxidation (M)
Alpha-S2-casein *Capra hircus* (P33049)	**14**	29 12 18 17 12	R.NAGPFTPTVNR.E R.NAGPFTPTVNR.E + 2 Deamidated (NQ) R.NANEEEYSIR.S K.ALNEINQFYQK.F + Deamidated (NQ) K.ALNEINQFYQK.F + 2 Deamidated (NQ)
Beta-casein *Bos taurus*(P02666) or *Bubalus bubalis* (Q9TSI0)	**17**	27 17 14	K.VLPVPQK.A K.AVPYPQR.D R.DMPIQAFLLYQEPVLGPVR.G + 2Deamidated (NQ)
Beta-lactoglobulin *Bos taurus* (P02754) or *Bubalus bubalis* (P02755)	**15**	19 24 61	K.IDALNENK.V + Deamidated (NQ) K.VLVLDTDYK.K R.TPEVDDEALEK.F

Table V.1.3. Frequency of detection of peptides/milk proteins in the samples from Buddha statua by LC-MS/MS. The comprehensive list of the matched peptides that have been observed in the whole set of samples. The position of the starting and the ending residues of each peptide are indicated as superscript. In brackets the accession number in the UniProt database. Frequency of the peptides and proteins identified in the sample are also reported below and on the side of each peptide/protein as assigned by Mascot. Peptides systematically detected as deamidated are also indicated (dea).

Protein	Peptide	*Bos taurus* (P02662)	*Bubalis bubalis* (O62823)	*Capra hircus* (P18626)	*Ovis aries* (P04653)	Number of samples in which the peptide has been matched
Alpha-S1-casein	^{19}HQGLPQEVLNENLLR^{37}	X	X			**3**
	^{29}EVLNENLLR^{37}	X			X	**4**
	^{38}FFVAPFPEVFGK^{49}	X	X			**15**
	^{38}FFVAPFPEVF^{47}	X	X			**1**
	^{40}VAPFPEVFGK^{49}	X	X			**1**
	^{38}FVVAPFPEVFR^{48}			X	X	**11**
	^{39}VVAPFPEVFR^{48}			X	X	**2**
	^{50}EKVNELSK^{57}	X				**1**
	^{99}EDVPSER^{105}	X	X	X	X	**12**
	^{106}YLGYLEQLLR^{115}	X	X	X	X	**11**
	^{107}LGYLEQLLR^{115}	X	X	X	X	**4**
	^{121}VPQLEIVPNSAEER^{134}	X				**2**
	^{158}AYFYPELFR^{166}	X	x as dea	x as dea	x as dea	**1**
	^{160}FYPELFR^{166}	X	x as dea	x as dea	x as dea	**2**
Number of samples in which the protein has been identified by Mascot		**15**	**10**	**11**	**11**	
Alpha-S2-casein		***Bos taurus* (P02663)**		***Capra hircus* (P33049)**		
	^{62}NANEEEYSIR^{72}			X		**2**
	^{96}ALNEINQFYQK^{106}	X		X		**2**
	^{130}NAVPITPTLNR^{140}	X				**5**
	^{130}NAGPFTPTVNR^{140}			X		**2**
	^{141}EQLSTSEENSK^{151}	X		X		**1**
	^{153}TVDMESTEVFTK^{164}	X				**2**
	^{189}FALPQYLK^{196}	X				**4**
	^{213}TNAIPYVR^{220}			X		**1**
	^{215}VIPYVRYL^{222}	X				**1**
Number of samples in which the protein has been identified by Mascot		**5**		**2**		
Beta-casein		***Bos taurus (P02666)***	***Bubalis bubalis* (Q9TSI0)**			
	^{123}EMPFPK^{128}	X	X			**3**
	^{129}YPVEPFTESQSL^{140}	X	X			**1**
	^{185}VLPVPQK^{191}	X	X			**5**
	^{192}AVPYPQR^{198}	X	X			**7**
	^{199}DMPIQAFLLYQEPVLGPVR^{217}	X	X			**3**
	^{206}LLYQEPVLGPVR^{217}	X	X			**3**
	^{207}LYQEPVLGPVR^{217}	X	X			**2**

	^{208}YQEPVLGPVR^{217}	X	X			**1**
Number of samples in which the protein has been identified by Mascot		**7**	**7**			
Beta-lactoglobulin		***Bos taurus* (P02754)**	***Bubalis bubalis* (P02755)**			
	^{37}SLAMAASDISLLDAQSAPLR^{56}	X	X			**4**
	^{41}AASDISLLDAQSAPLR^{56}	X	X			**1**
	^{50}AQSAPLR^{56}	X	X			**2**
	^{100}IDALNENK^{107}	X	X			**2**
	^{108}VLVLDTDYK^{118}	X	X			**4**
	^{141}TPEVDDEALEK^{151}	X	X			**4**
Number of samples in which the protein has been identified by Mascot		**4**	**4**			

Appendix B

Table VI.1. Samples used in the database construction. A random numbering and a letter that indicate the binders are shown: milk (M), eggs (E) and animal glue (G). For each sample, the pictorial pigment and some specific characteristics such as the state of aging, special pre-treatment that has been used in the classical proteomic approaches and the mixing with other protein-based binders were reported. Proteins identified in unpublished paints replica are shown in table VI.2. In apex is indicated, if present, the reference number corresponding to the paper where results of singles sample are shown.

Sample	Pigment	Characteristic	Pre-treatment	Mass Spectrometry analysis
M1[23]	Minium ($Cu_2CO_3(OH)_2$)	-	-	MALDI-TOF; LC-MS/MS
M2[23]	Cinnabar (HgS)	-	-	MALDI-TOF; LC-MS/MS
M3[23]	-	-	-	MALDI-TOF; LC-MS/MS
M4[23]	$CaCO_3$	-	-	MALDI-TOF; LC-MS/MS
M5	Malachite ($Cu_2CO_3(OH)_2$)	-	-	LC-MS/MS
M6	Yellow ochre ($Fe_2O_3H_2O$)	-	-	LC-MS/MS
M7	Biacca $(PbCO_3)_2 \cdot Pb(OH)_2$	-	-	LC-MS/MS
M8	Biacca $(PbCO_3)_2 \cdot Pb(OH)_2$	-	-	LC-MS/MS
M9	Biacca $(PbCO_3)_2 \cdot Pb(OH)_2$	-	-	LC-MS/MS
M10	-	-	-	LC-MS/MS
M11[23]	Minium ($Cu_2CO_3(OH)_2$)	Aged sample	-	MALDI-TOF; LC-MS/MS
M12[23]	Cinnabar (HgS)	Aged sample	-	MALDI-TOF; LC-MS/MS
M13[23]	-	Aged sample	-	MALDI-TOF; LC-MS/MS
M14[23]	$CaCO_3$	Aged sample	-	MALDI-TOF; LC-MS/MS
M15	Red ochre (Fe_2O_3)	-	-	MALDI-TOF; LC-MS/MS
M16[23]	$CaCO_3$	Mixed with animal glue	-	MALDI-TOF; LC-MS/MS
M17[23]	Azurite ($Cu_3(CO_3)_2(OH)_2$)	Mixed with animal glue	-	MALDI-TOF; LC-MS/MS
M18[23]	Cinnabar (HgS)	Mixed with animal glue	-	MALDI-TOF; LC-MS/MS
M19[23]	Minium ($Cu_2CO_3(OH)_2$)	Mixed with animal glue	-	MALDI-TOF; LC-MS/MS
M20[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M21[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS

M22[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M23[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M24[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M25[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M26[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M27[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M28[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M29[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M30[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M31[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M32[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M33[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M34[30]	-	Buddha's decoration	Urea 6M	LC-MS/MS
M35[24]	Red ochre (Fe_2O_3)	-	N-Glycosidase F	MALDI-TOF; LC-MS/MS
M36[24]	Red ochre (Fe_2O_3)	-	N-Glycosidase F	MALDI-TOF; LC-MS/MS
M37[25]	-	-	-	LC-MS/MS
M38[25]	-	-	-	LC-MS/MS
M39[25]	-	Camposanto Monumentale (Pisa)	-	LC-MS/MS
M40[8]	Hematite (α-Fe_2O_3)	-	-	LC-MS/MS
M41[8]	Malachite ($Cu_2CO_3(OH)_2$)	-	-	LC-MS/MS
M42[8]	Malachite ($Cu_2CO_3(OH)_2$)	Paraloid B72	-	LC-MS/MS
M43[8]	Red pigment (GA-07)	-	-	LC-MS/MS
E1	-	-	-	LC-MS/MS
E2	Biacca $(PbCO_3)_2 \cdot Pb(OH)_2$		-	LC-MS/MS
E3	Malachite ($Cu_2CO_3(OH)_2$)	-	-	LC-MS/MS
E4	Yellow ochre ($Fe_2O_3H_2O$)	-	-	LC-MS/MS
E5	White St. John's ($CaCO_3$)	-	-	LC-MS/MS

E6[8]	Hematite (α-Fe_2O_3)$_2$)	-	-	LC-MS/MS
E7	-	-	-	MALDI-TOF; LC-MS/MS
E8	-	Aged sample	-	MALDI-TOF; LC-MS/MS
E9	$CaCO_3$	-	-	MALDI-TOF; LC-MS/MS
E10	$CaCO_3$	Aged sample	-	MALDI-TOF; LC-MS/MS
E11[24]	Cinnabar (HgS)	-	N-Glycosidase F	MALDI-TOF; LC-MS/MS
E12	Cinnabar (HgS)	Aged sample	-	MALDI-TOF; LC-MS/MS
E13[24]	Minium ($Cu_2CO_3(OH)$	-	N-Glycosidase F	MALDI-TOF; LC-MS/MS
E14	Minium ($Cu_2CO_3(OH)$	Aged sample	-	MALDI-TOF; LC-MS/MS
E15[24]	Azurite ($Cu_3(CO_3)_2(OH)_2$)	-	N-Glycosidase F	MALDI-TOF; LC-MS/MS
E16	Azurite ($Cu_3(CO_3)_2(OH)_2$)	Aged sample	-	MALDI-TOF; LC-MS/MS
E17[24]	Red ochre (Fe_2O_3)	Mixed with milk	N-Glycosidase F	LC-MS/MS
E18	Red ochre (Fe_2O_3)	Aged sample	-	LC-MS/MS
E19[24]	Red ochre (Fe_2O_3)	-	N-Glycosidase F	LC-MS/MS
E20[8]	GA04	-	-	LC-MS/MS
E21	-	Gilding	-	LC-MS/MS
E22	-	Gilding	-	LC-MS/MS
E23	-	Gilding	-	LC-MS/MS
E30	-	Gilding		LC-MS/MS
E31	Cn/tg Cinabro (HgS)	Fat tempera and linseed oil	-	LC-MS/MS
E32	Biacca $(PbCO_3)_2 \cdot Pb(OH)_2$	-	-	LC-MS/MS
E33	Verdigri (Cu)	UV	-	LC-MS/MS
E36	Malachite ($Cu_2CO_3(OH)_2$)	-	-	LC-MS/MS
E37	White St. John's ($CaCO_3$)	-	-	LC-MS/MS
E38	Yellow ochre ($Fe_2O_3H_2O$)	-	-	LC-MS/MS
E39	Biacca $(PbCO_3)_2 \cdot Pb(OH)_2$	-	-	LC-MS/MS
E40	-	-	-	LC-MS/MS
E41[8]	Malachite ($Cu_2CO_3(OH)_2$)	-	-	LC-MS/MS

E42[8]	Hematite (α-Fe_2O_3)	Microwave	-	LC-MS/MS
E43[8]	Malachite ($Cu_2CO_3(OH)_2$)	Paraloid B72	-	LC-MS/MS
E44[8]	Malachite ($Cu_2CO_3(OH)_2$)	Paraloid B72 and microwave	-	LC-MS/MS
E45	Azurite ($Cu_3(CO_3)_2(OH)_2$)	Mixed with animal glue	-	MALDI-TOF; LC-MS/MS
E46	Minium ($Cu_2CO_3(OH)$	Mixed with animal glue	-	MALDI-TOF; LC-MS/MS
E47	Cinnabar (HgS)	Mixed with animal glue	-	MALDI-TOF; LC-MS/MS
E48[24]	Red ochre (Fe_2O_3)	Mixed with animal glue	N-Glycosidase F	MALDI-TOF; LC-MS/MS
E49	$CaCO_3$	-	-	MALDI-TOF; LC-MS/MS
E50	Purple pigment	-	N-Glycosidase F	LC-MS/MS
E51[26]	DSFL6	Old gilding sample	-	LC-MS/MS
G1	-	-	Urea 6M	MALDI-TOF; LC-MS/MS
G2	Minium ($Cu_2CO_3(OH)_2$)	-	Urea 6M	MALDI-TOF; LC-MS/MS
G3[24]	Yellow ochre ($Fe_2O_3H_2O$)	-	Urea 6M	MALDI-TOF; LC-MS/MS
G4	Cinnabar (HgS)	-	Urea 6M	MALDI-TOF; LC-MS/MS
G5	$CaCO_3$	-	Urea 6M	MALDI-TOF; LC-MS/MS
G6	Azurite ($Cu_3(CO_3)_2(OH)_2$)	-	Urea 6M	MALDI-TOF; LC-MS/MS
G7[24]	Yellow ochre ($Fe_2O_3H_2O$)	-	Urea 6M and N-Glycosidase F	MALDI-TOF; LC-MS/MS
G8	Minium ($Cu_2CO_3(OH)$	Mixed with milk	Urea 6M	MALDI-TOF; LC-MS/MS
G9	Yellow ochre ($Fe_2O_3H_2O$)	Mixed with milk	Urea 6M	MALDI-TOF; LC-MS/MS
G10	Cinnabar (HgS)	Mixed with milk	Urea 6M	MALDI-TOF; LC-MS/MS
G11	$CaCO_3$	Mixed with milk	Urea 6M	MALDI-TOF; LC-MS/MS
G12	Azurite ($Cu_3(CO_3)_2(OH)_2$)	Mixed with milk	Urea 6M	MALDI-TOF; LC-MS/MS
G13	Minium ($Cu_2CO_3(OH)$	Mixed with eggs	Urea 6M	MALDI-TOF; LC-MS/MS
G14	Yellow ochre ($Fe_2O_3H_2O$)	Mixed with eggs	Urea 6M	MALDI-TOF; LC-MS/MS
G15	Cinnabar (HgS)	Mixed with eggs	Urea 6M	MALDI-TOF; LC-MS/MS
G16	$CaCO_3$	Mixed with eggs	Urea 6M	MALDI-TOF; LC-MS/MS

G17	Azurite ($Cu_3(CO_3)_2(OH)_2$)	Mixed with eggs	Urea 6M	MALDI-TOF; LC-MS/MS
G18[25]	-	Camposanto Monumentale (Pisa)	-	LC-MS/MS
G19[25]	-	-	-	LC-MS/MS
G20[25]	-	-	-	LC-MS/MS
G21[8]	Hematite (α-Fe_2O_3)	-	-	LC-MS/MS
G22[8]	Malachite ($Cu_2CO_3(OH)_2$)	-	-	LC-MS/MS
G23	Blue pigment	-	-	LC-MS/MS
G24	-	Gilding	-	LC-MS/MS
G25	Yellow ochre ($Fe_2O_3H_2O$)	-	-	LC-MS/MS
G26	-	-	-	LC-MS/MS
G27	Malachite ($Cu_2CO_3(OH)_2$)	-	-	LC-MS/MS
G28	White St. John's ($CaCO_3$)	-	-	LC-MS/MS
G29	Biacca (PbCO3*$Pb(OH)_2$)	-	-	LC-MS/MS

8. Leo, G. *et al.* Proteomic strategies for the identification of proteinaceous binders in paintings. *Anal. Bioanal. Chem.* **395,** 2269–2280 (2009).
23. Vinciguerra, R., De Chiaro, A., Pucci, P., Marino, G. & Birolo, L. Proteomic strategies for cultural heritage: form bones to paintings. *Microchem. J.* **126,** 341–348 (2015).
24. Vinciguerra, R. *et al.* Deglycosylation Step to Improve the Identification of Egg Proteins in Art Samples. *Anal. Chem.* 10178–10182 (2015). doi:10.1021/acs.analchem.5b02423
25. Leo, G. *et al.* Deamidation at asparagine and glutamine As a major modification upon deterioration/aging of proteinaceous binders in mural paintings. *Anal. Chem.* **83,** 2056–2064 (2011).
26. Bonaduce, I., Colombini, M. P. & Diring, S. Identification of garlic in old gildings by gas chromatography-mass spectrometry. *J. Chromatogr. A* **1107,** 226–232 (2006).
30. Lluveras-Tenorio, A., Vinciguerra, R., Galano, E., Blaensdorf, C., Emmerling, E., Colombini, M.P., Birolo, L., Bonaduce, I. The Painting Technique of the Giant Buddhas of Bamiyan (Afghanistan). Submitted. (Peptides identification is reported in Table V.1.2 (Appendix A)).

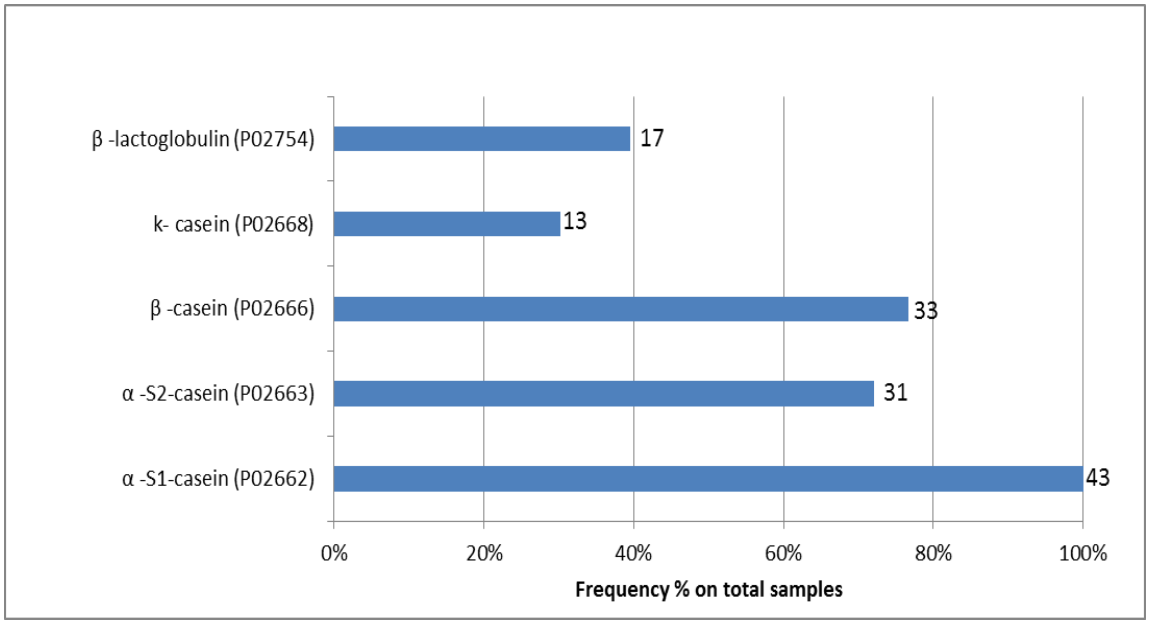


Figure VI.2. Graphical representation of the identified proteins in 43 samples containing milk from *Bos taurus* as protein-based binder.

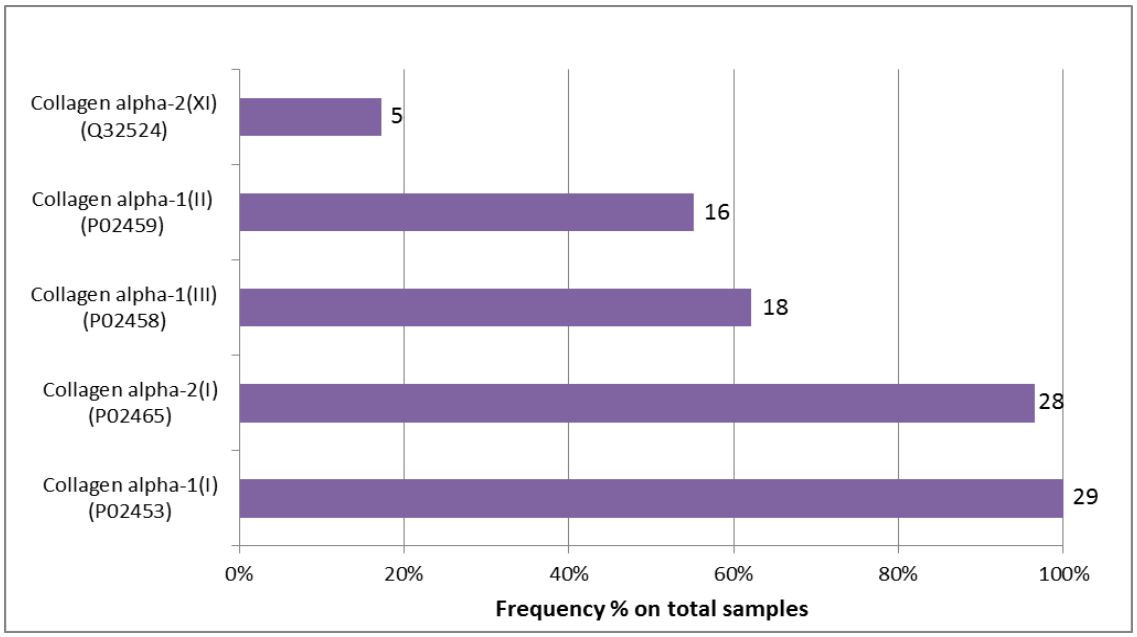


Figure VI.3. Graphical representation of the identified proteins in 29 samples containing animal glue from *Bos taurus* as protein-based binder.

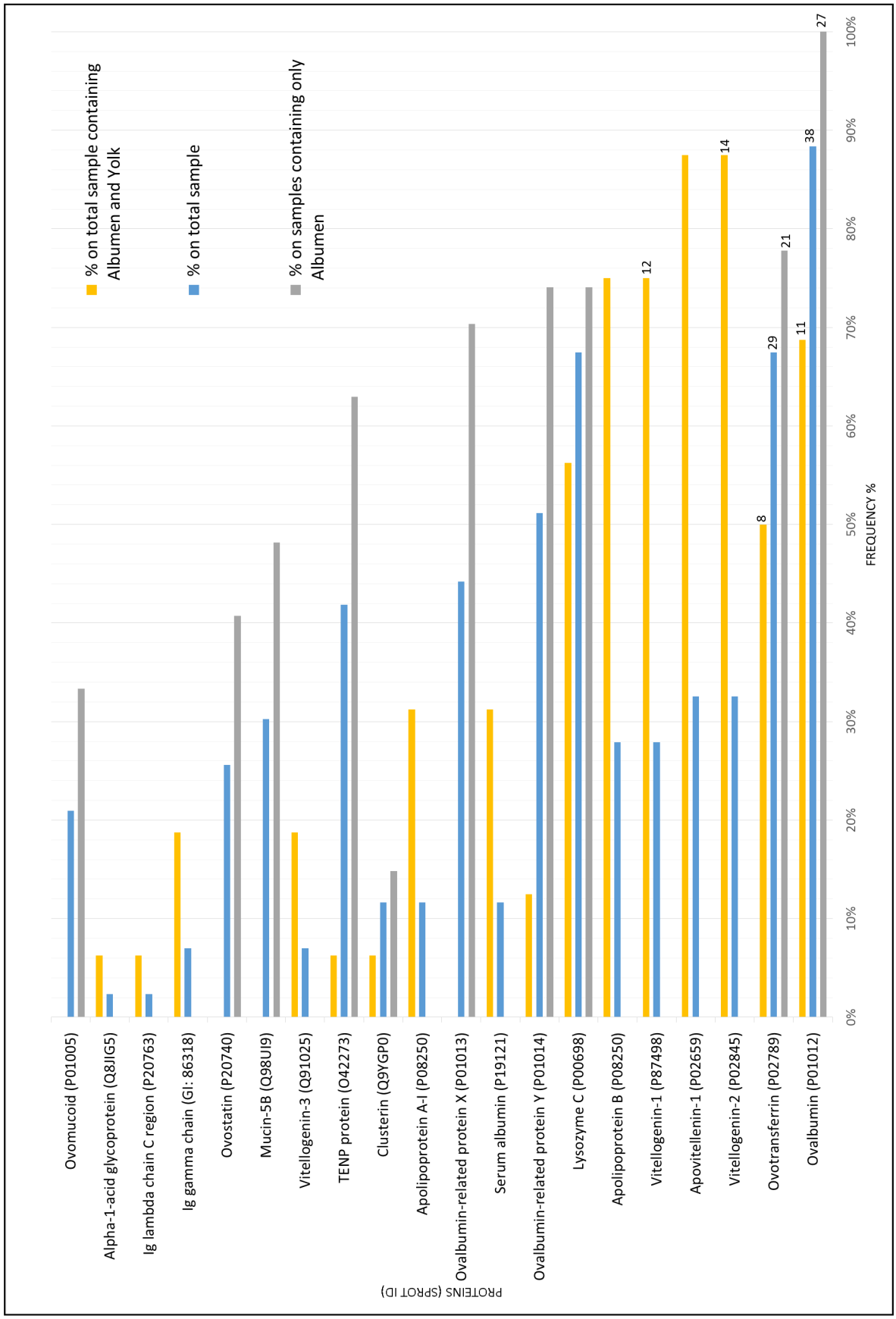


Figure VI.4. Graphical representation of the identified proteins in 43 samples containing egg from *Gallus gallus* as protein-based binder.

Table VI.2. Peptide occurrence in the identification by LC-MS/MS and MALDI-TOF used for database setup. Aliquots of 43 samples (replicas and historical) containing milk from *Bos taurus* (from page 145 to 147) *and* egg from *Gallus gallus* (from page 148 to 160) were treated in heterogeneous phase with trypsin with or without pre-treatment and analysed by LC-MS/MS and MALDI-TOF as indicated in Table VI.1. Proteins were identified searching UniprotSprot database with MS/MS Ion search Mascot software (Matrix Science) with Chordata as taxonomy restriction, with deamidation on Gln and Asn, oxidation on Met, pyro-Glu formation at Gln at the N-terminus of peptides as variable modifications. D indicates that peptides has been identified as deamidated. The frequency is calculated on the peptide number or its deamidated if present, by MALDI-TOF and/or LC-MS/MS analysis.

Protein	MH+	Peptide + Modification	Dea sites	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		19		20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		36		37	38	39	40	41	42	43	Frequency
				MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Alpha S1 Casein (P02662)	689,38	52-57 K.VNELSK.D	1																																																						X										1
	748,37	209-214 K.TTMPLW.-																													X		X				X		X																	X		X			X					7
	764,36	209-214 K.TTMPLW.- + Oxi(M)																															X				X		X																	X										4
	831,38	99- 105 K.EDVPSER.Y																							X		X								X		X			X		X		X	X	X	X		X	X	X	X	X	X							X					17
	892,41	140-147 K.EGIHAQQK.E + Gln->pyro-Glu (N-term Q)	2																											X		X		X		X		X																			X									6
	910,47	140-147 K.EGIHAQQK.E	2																								X																																									1
	946,52	50-57 K.EKVNELSK.D	1																								X		X										X						X																	X	X					6
	1267,70	106-115 R.YLGYLEQLLR.L	1	X		X		X	X-D	X	X-D	X	X		X-D		X	X	X		X-D	X	X	X	X-D	X	X-D	X	X-D	X		X	X	X	X-D	X	X-D	X	X-D	X	X-D	X-D	X - D	X - D	X - D	X - D	X - D	X - D	X - D	X - D	X - D					X			X	X-D	X	X	X-D		X	X		50
	1337,68	95-105 K.HIQKEDVPSER.Y	1	X					X-D	X	X				X-D		X-D		X		X		X		X	X	X	X-D	X	X-D		X							X	X																			X	X-D			X		X	X		23
	1384,73	38-49 R.FFVAPFPEVFGK.E		X		X	X	X	X	X	X													X	X	X	X	X	X	X			X				X		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X		X			X	X	X	X	X	40
	1506,78	135-147 R.LHSMKEGIHAQQK.E	2												X																																																						1
	1508,88	106-117 R.YLGYLEQLLRLK.K	1	X																				X		X		X																																								4
	1509,87	106-117 R.YLGYLEQLLRLK.K	1							X																																																										1
	1523,76	135-147 R.LHSMKEGIHAQQK.E	2							X																																																										1
	1580,83	121-134 K.VPQLEIVPNSAEER.L	2	X		X					X			X				X		X		X		X	X	X	X	X	X	X			X		X						X	D-2D			X-D-2D																	X		X-D			X	X	23
	1636,98	106-118 R.YLGYLEQLLRLKK.Y	2																						X				X																																								2
	1641,87	38-51 R.FFVAPFPEVFGKEK.V			X		X		X																X		X		X				X	X	X	X	X	X	X	X	X																	X	X	X			X					19
	1759,95	23-37 K.HQGLPQEVLNENLLR.F	4	X		X			X-D-2D		X-D		X	X	X-D				X	X	X-D-2D	X	X-D-2D	X	X-D	X	X-D	X	X-D-2D			X	X-D	X	X-D	X	X-D	X	X-D	X	X-D		X-3D		X						X-D								X	X	X-D	X	X-D-2D		X	X		38
	1871,99	119-134 K.YKVPQLEIVPNSAEER.L	2	X		X				X	X-D							X					X-D	X	X-D	X	X-D	X	X-D	X	X-D		X	X	X		X		X		X-D																							X-D					21
	2235,24	19-37 K.HPIKHQGLPQEVLNENLLR.F	4	X		X			X	X	X-D		X		X-D				X				X-D-2D	X	X-D	X	X-D-2D	X	X-D-2D	X	X-D-2D		X-D		D		D	D	X-D		X-D																			X-D			X-D					25
	2316,14	148-166 K.EPMIGVNQELAYFYPELFR.Q	2	X		X		X		X	X-D			X	X-D				X				X-D	X	X	X	X	X	X-D	X	X	X	X	X	X-D	X		X	X-D	X	X-D																		X				X					28
	2332,13	148-166 K.EPMIGVNQELAYFYPELFR.Q + Oxi(M)	2	X		X		X		X	X-D			X	X-D								X	X	X	X	X-D	X	X	X	X-D	X	X	X	X-D	X	X	X	X	X	X-D																		X									27
	2455,10	52-73 K.VNELSKDIGSESTEDQAMEDIK.Q + Oxi(M)	2									X																																																								1
	3084,59	121-147 K.VPQLEIVPNSAEERLHSMKEGIHAQQK.E + Oxi(M)	4																																	X																																1
	3207,59	140- 166 K.EGIHAQQKEPMIGVNQELAYFYPELFR.Q	4							X	X- D		X-D-2D		X-D-2D				X		X-D-2D		X-D-2D-3D	X	X-D	X	X-D-2D		X-D-2D	X			X-D		X-D		X-D		X		X-D																			X			X					20
	3223,59	140-166 K.EGIHAQQKEPMIGVNQELAYFYPELFR.Q + Oxi(M)	4																						X		X-D		X-D				X		X		X-D-2D	X	X		X-D																												9
	4699,15	167-208 R.QFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSENSEK.T + Gln->pyro-Glu (N-term Q)	5																						X		X		X																																								3
	4716,17	167-208 R.QFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSENSEK.T	5																						X-D		X	X	X-D																																							4

Protein	MH$^+$	Peptide + Modification	Dea sites	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		19		20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		36		37	38	39	40	41	42	43	Frequency
				MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Alpha S2 Casein (P02662)	634,39	176-180 R.LNFLK.K	1																																				X		X				X		X																	X-D		X			X-D					7
	690,37	86-91 K.ITVDDK.H																																							X				X																													2
	746,46	215-220 K.VIPYVR.Y					X	X	X										X		X		X	X		X	X		X	X	X																														X			X					14
	748,37	168-173 K.LTEEEK.																																																								X										1
	874,45	40-47 K.NMAINPSK.E	2																																X				X																							D					X					4
	904,54	174-180 K.NRLNFLK.K	2																																																							X-D										1
	977,52	166-173 K.TKLTEEEK.N		X																																																								X								2
	979,56	189-196 K.FALPQYLK.T	1		D	X	X-D	X	X-D	X	X-D				X				X		X-D	X	X	X	X-D	X	X-D		X-D		X-D	X	X-D	X	X-D	X	X	X	X-D	X	X-D	X			X-D					X-D	X-D								X	X-D	X	X	X-D					37
	1022,60	215-222 K.VIPYVRYL.-		X				X	X												X		X																	X																	X								7
	1045,43	48-56 K.ENLCSTFCK.E	1			X		X		X																																																									3
	1098,61	204-212 K.AMKPWIQPK.T	1	X	X	X	X-D	X	X-D		X				X				X		X		X	X	X	X	X-D		X-D		X																													X		X	X-D					20
	1195,68	130-140 R.NAVPITPTLNR.E	2	X	X-D		X-D		X-D		X-D												X-D	X	X-D	X	X-D	X	X-D	X	X-D		X		X-D		X		X-D		X	X-D-2D			X-D		X-D-2D		X-D-2D	X-D											X-D	X-D	X-D-2D					27
	1246,64	86-95 K.ITVDDKHYQK.A	1																														X				X																					X										3
	1247,66	176-185 R.LNFLKKISQR.Y	2																														X				X																														2
	1251,57	141-151 R.EQLSTSEENSK.K	2	X	D	X		X		X	X								D	X					D		D				D																X																								12
	1351,78	129-140 K.RNAVPITPTLNR.E	2																													X								X																			X-D			X-D	X-D	X-D					6
	1367,70	96-106 K.ALNEINQFYQK.F	4		X	X	X-2D	X	X-D	X	X-D												X-2D			X	X-D	X	X-D	X-D	X-D-2D	X	X-D	X	X-D		X-D		X		X-D		X-D	D-2D			XD																X	X	X		X-D					28
	1386,65	153-164 K.TVDMESTEVFTK.K																																																													X					1
	1402,65	153-164 K.TVDMESTEVFTK.K + Oxi(M)																																								X																										1
	1400,74	186-196 R.YQKFALPQYLK.T	2		X					X															X				X																															X								5
	1514,74	152-164 K.KTVDMESTEVFTK.K																																																								X										1
	1514,74	153-165 K.TVDMESTEVFTKK.T																																																								X										1
	1530,74	152-164 K.KTVDMESTEVFTK.K + Oxi(M); K.TVDMESTEVFTKK.T + Oxi(M)						X						X														X				X		X		X		X																					X									9
	1633,89	168-180 K.LTEEEKNRLNFLK.K	2																											X	X					X		X																			X	X									6
	1916,86	40-56 K.NMAINPSKEKENLCSTFCK.E + Oxi(M)														X																																																			1
	2556,33	130-152 R.NAVPITPTLNREQLSTSEENSKK.T	4																											X		X		X		X		X																				X									6
	2689,15	61-85 R.NANEEEYSIGSSSEESAEVATEEVK.I	2									X																																																							1
	2709,41	107-128 K.FPQYLQYLYQGPIVLNPWDQVK.R	5	X			X-D	X	X-2D-4D-5D	X	X				X-D					X	X-D	X	X-D-2D	X	X-D	X	X	X	X-D	X	X		D		X-D		X-D-3D				X-D																			X			X					25

Protein	MH^+	Peptide + Modification	Dea sites	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		19		20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		36		37	38	39	40	41	42	43	Frequency
				MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Beta Casein (P02666)	646,32	115-120 K.EAMAPK.H																														X		X		X		X		X																			X									6
	742,45	218-224 R.GPFPIIV.-		X		X				X				X						X		X				X				X								X		X																												10
	748,37	123-128 K.EMPFPK.Y																																																										X			X					2
	764,36	123-128 K.EMPFPK.Y + Oxi(M)						X		X					X				X		X		X																X						X		X			X																			10
	780,50	185-191 K.VLPVPQK.A	1						X		X										X												X				X	X		X	X	X			X					X	X				X				X	X	X	X					17
	830,45	192-198 K.AVPYPQR.D	1				X-D	X	X	X	X			X	X		X	X	X	X	X-D	X	X-D	X	X	X	X-D	X	X-D	X	X-D	X		X		X	X	X	X	X		X			X-D		X-D			X-D	X-D	X				X-D			X	X	X	X	X		X	X		43
	873,49	113-120 K.VKEAMAPK.H																														X		X		X		X																					X									5
	1013,52	121-128 K.HKEMPFPK.Y																																																									X		X					X	X		4
	1438,91	3-16 K.VLILACLVALALAR.E		X				X																	X		X		X																																				X	X		7
	1622,92	115-128 K.EAMAPKHKEMPFPK.Y + Gln->pyro-Glu (N-term Q)																														X		X		X		X		X																			X									6
	1981,86	48-63 K.FQSEEQQQTEDELQDK.I	5																																																							X		X	X		X					4
	2186,17	199-217 R.DMPIQAFLLYQEPVLGPVR.G	2	X		X		X		X					X-D-2D				X		X-D		X-D	X	X-D	X		X		X	X-D-2D		X				D		X-D			2D																		X-D-2D	X-D	X-D	X		X	X		24
	2909,60	199-225 R.DMPIQAFLLYQEP VLGPVRGPFPIIV.-	2	X		X		X		X			X-2D	X	X-D-2D			X	X		X-D		X-D		X-D				X		X-D		X				X		X																					X-D			X					19
	2202,16	199-217 R.DMPIQAFLLYQEPVLGPVR.G + Oxi(M)	2	X		X		X		X			X-D-2D		X-D				X		X-D-2D		X	X		X		X		X															X-D-2D						X-D									X-D						X	X	18
	2927,56	199-225 R.DMPIQAFLLYQEPVLGPVRGPFPIIV.- + Oxi(M)	2															X					X-D-2D																																							X-D-2D							X	4
	2997,60	192-217 K.AVPYPQRDMPIQAFLLYQEPVLGPVR.G	3																																																											X					1
	3014,57	192-217 K.AVPYPQRDMPIQAFLLYQEPVLGPVR.G + Oxi(M)	3																																																						D			D							2

Protein	MH^+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Ovalbumin (P01012)	552,23	863-866 MEEK + Oxi (M)																						X																																						1
	632,34	220-227 K.GLWEK.A																																																	X				X		X		X			4
	647,39	281-285 K.VYLPR.M								X						X				X		X		X				X																									X			X	X	X	X		X		X	13
	780,39	106-111 R.LYAEER.Y		X	X	X	X	X							X		X		X	X	X		X	X	X		X	X	X	X	X	X												X	X												X		X		X			24
	795,37	286-289 R.MKMEEK.Y																						X								X																														2
	811,37	286-289 R.MKMEEK.Y + Oxi (M)																						X																																						1
	820,40	201-205 R.VTEQESK.P	1							X						X		X				X		X		X		X																																		7
	822,40	220-227 R.VASMASEK.M		X	X	X	X	X							X		X		X		X		X		X		X		X	X	X	X																				X								X			18
	827,36	286-289 R.MKMEEK.Y + 2 Oxi (M)														X																																														1
	838,40	220-227 R.VASMASEK.M + Oxi (M)																			X		X	X				X	X																																		5
	888,57	279-285 K.IKVYLPR.M								X		X		X		X		X		X		X			X	X		X		X		X																															12
	906,52	281-287 K.VYLPRMK.M																						X								X																														2
	922,51	281-287 K.VYLPRMK.M + Oxi (M)																						X																																						1
	957,58	52-59 R.TQINKVVR.F	2																	X				X	X		X	X		X	X	X-D																				X				X-D				X			11
	978,54	183-190 K.GLWEKAFK.D																		X																																											1
	1016,66	278-285 R.KIKVYLPR.M								X		X		X				X		X		X				X		X																						X		X											10
	1081,54	220-229 R.VASMASEKMK.I		X		X		X																			X			X	X	X																				X				X							9
	1097,53	220-229 R.VASMASEKMK.I + Oxi (M)																												X	X	X																															3
	1190,60	361-370 R.ADHPFLFCIK.H																																																														1
	1209,52	191-200 K.DEDTQAMPFR.V	1	X	X	X	X	X																																	X	X				X																8
	1225,52	191-200 K.DEDTQAMPFR.V + Oxi (M)	1												X		X		X		X				X-D		X		X		X															X																9
	1275,80	278-287 R.KIKVYLPRMK.M								X				X				X								X																																					4
	1345,73	371-382 K.HIATNAVLFFGR.C	1	X	X	X	X	X									X		X		X	X	X		X	X	X		X	X	X							X					X		X	X				X	X		X	X	X	X	X	X		X-D	29
	1347,77	52-62 R.TQINKVVRFDK.L	2							X				X				X		X		X				X																																				6
	1355,65	94-105 E.PNDVYSFSLASR.E	1																					X																																						1
	1416,79	48-59 K.DSTRTQINKVVR.F	2																	X																																										1
	1455,73	281-291 K.VYLPRMKMEEK.Y + 2 Oxi (M)																		X								X																																		2
	1465,77	112-123 R.YPILPEYLQCVK.E	1					X																																																						1
	1555,72	188-200 K.AFKDEDTQAMPFR.V	1	X	X	X	X	X						X	X		X-D	X	X-D		X		X		X-D		X-D		X-D	X	X-D										X	X		X	X		X			X	X	X		X		X		X	X			29
	1571,71	188-200 K.AFKDEDTQAMPFR.V + Oxi (M)	1												X		X		X		X		X		X-D		X		X	X	X							X								X		X				X		X				X			16

Protein	MH^+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Ovalbumin (P01012)	1581,72	265-277 K.LTEWTSSNVMEER.K	1		X	X	X	X		X				X			X	X	X-D		X	X	X-D		X		X		X	X	X										X	X	X	X	X		X		X	X		X		X	X		X		X			30
	1597,72	265-277 K.LTEWTSSNVMEER.K + Oxi (M)	1														X		X-D		X				X		X		X	X	X								X								X		X		X						X		X		X	15
	1687,84	128-143 R.GGLEPINFQTAADQAR.E	3	X	X	X	X	X		X					X-D		X-D	X-D	X-D	X	X-D	X	X		X	X	X		X	X	X	X-D		X-D		X			X		X	X	X	X	X	X	X		X		X	X	X		X	X			X-D		X	39
	1709,82	265-278 K.LTEWTSSNVMEERK.I	1	X		X														X										X	X																		X										X		X	8
	1731,90	371-386 K.HIATNAVLFFGRCVSP.-	1																	X																																										1
	1773,90	324-340 K.ISQAVHAAHAEINEAGR.E	2	X	X	X	X	X		X					X-D		X-D	X	X-D-2D	X	X-D	X			X-D		X-D		X-D	X	X-D	X-D					X				X	X	X	X	X	X	X		X		X-D-3D		X		X	X	X-D		X-D		X	35
	1858,97	144-159 R.ELINSWVESQTNGIIR.N	3	X	X	X	X	X							X		X-D-2D		X-D-2D		X-D		X		X-D		X-D		X-D	X	X-D-2D	X-D-2D-3D						X-D	X-D		X	X	X	X	X	X	X		X	X	X-3D	X	X	X	X	X	X	X	X-D		X-D-2D	37
	1950,99	265-280 K.LTEWTSSNVMEERKIK.V																										X				X																														2
	2008,94	341-360 R.EVVGSAEAGVDAASVSEEFR.A			X	X	X	X							X		X		X		X		X		X		X		X	X	X												X																			15
	2008,94	341-360 R.EVVGSAEAGVDAASVSEEF.A																																											X																1
	2027,07	112-127 R.YPILPEYLQCVKELYR.G	1											X																																																1
	2102,20	311-328 K.DPVLKDLLFKDSAIMLKR.V			X		X																																																							2
	2169,04	183-200 K.GLWEKAFKDEDTQAMPFR.V	1																												X																				X											2
	2227,15	106-123 R.LYAEERYPILPEYLQCVK.E	1																													X																														1
	2249,13	124-143 K.ELYRGGLEPINFQTAADQAR.E	3			X		X		X				X				X				X	D			X					D	X																	X	X	X		X				X	X	X			17
	2281,18	86-105 R.DILNQITKPNDVYSFSLASR.L	3		X	X	X	X		X							X-D	X	X-D		X-D	X	X		X-D	X	X-D		X-D	X	X-D	X-D-2D		X				D			X		X	X		X	X		X		X-D			X	X-D		X-D	X	D-2D		X-D	33
	2284,15	201-219 R.VTEQESKPVQMMYQIGLFR.V	3			X	X			X				X			X		X-D		X	X	X		X-D		X		X-D	X	X	X-D							X								X		X		X-D				X				X-D		D	22
	2300,14	201-219 R.VTEQESKPVQMMYQIGLFR.V + Oxi (M)	3																								X		X																		X		X		X			X				X				7
	2316,14	201-219 R.VTEQESKPVQMMYQIGLFR.V + 2Oxi (M)	3																		X				X																						X		X								X		X			6
	2460,32	160-182 R.NVLQPSSVDSQTAMVLVNAIVFK.G	4	X	X	X	X	X							X-D		X-D		X-D		X-D		X		X-D		X-D		X-D	X	X-D								D		X	X		X	X	X	X		X		X		X		X		X-D		X			28
	2476,31	160-182 R.NVLQPSSVDSQTAMVLVNAIVFK.G+ Oxi (M)	4												D		D		X-D		X				X-D		X		X		X-D-2D			X				D								X									X		X-D					13
	2476,32	160-182 R.NVLQPSSVDSQTAMVLVNAIVFK.G + Oxi (M)	4																																														X													1
	2788,44	106-127 R.LYAEERYPILPEYLQCVKELYR.G	1																													X																		X												2
	2992,49	21-47 K.VHHANENIFYCPIAIMSALAMVYLGAK.D+ Oxi (M)	1																																																			X								1
	3073,64	160-187 R.NVLQPSSVDSQTAMVLVNAIVFKGLWEK.A	4		X			X											X-D		X-D		X-D		X-D		X-D-2D		X-D-2D	X	X-D								X								X		X		D						2D					15
	3087,53	201-227 R.VTEQESKPVQMMYQIGLFRVASMASEK.M	3																													X																			X		X		X		X					6
	3089,65	160-187 R.NVLQPSSVDSQTAMVLVNAIVFKGLWEK.A + Oxi (M)	4																												D																						X		2D					3
	3346,67	201-229 R.VTEQESKPVQMMYQIGLFRVASMASEKMK.I	3																																																X		X		X		D					4
	3527,79	128-159 R.GGLEPINFQTAADQARELINSWVESQTNGIIR.N	5																						X		X-D-2D		X-D			X																			X		X-2D		3D							7
	3820,85	127-160 K.AFKDEDTQAMPFRVTEQESKPVQMMYQIGLFR.V	4																																																X				X		X-D					2
	3863,01	230-264 K.ILELPFASGTMSMLVLLPDEVSGLEQLESIINFEK.L	1																						X-2D				X-D		X-D																															3

Protein	MH$^+$	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Ovotransferrin (P02789)	601,38	620-624 K.ANKIR.D	1																													X																				X											2
	635,38	316-320 or 653-657 K.DLLFK.D																														X																				X						X					3
	645,36	625-629 R.DLLER.Q																														X																								X		X		X			4
	645,36	195-200 K.GDPKTK.C																																																		X											1
	699,44	310-315 K.KDPVLK.D																									X		X		X																																3
	708,36	674-679 K.EFLGDK.F		X		X	X	X							X				X		X		X				X		X	X	X																															12
	735,40	222-228 K.GDVAFVK.H		X	X	X	X	X							X		X		X		X		X		X		X			X	X	X																											X			15
	747,39	38-43 K.KCNNLR.D	2		X	X	X																																																								3
	750,34	260-265 K.TCNWAR.V																														X																															1
	777,42	321-327 K.DSAIMLK.R		X		X	X	X																																	X		X								X				X		X		X			10
	793,42	321-327 K.DSAIMLK.R + Oxi (M)													X		X		X		X				X																										X		X		X		X		X			10
	830,39	472-479 K.SCHTAVGR.T																																																							X						1
	832,43	572-578 K.NKADWAK.N	1																													X																															1
	837,41	634-641 R.FGVNGSEK.S	1	X	X	X	X	X							X		X-D		X-D		X-D		X-D		X-D		X-D		X-D	X	X-D	X-D																			X-D				X-D		X-D		X-D			20
	846,40	253-259 R.QPVDNYK.T + Gln->pyro-Glu (N-term Q)	1																X																																					X						2
	863,42	253-259 R.QPVDNYK.T	2	X	X	X	X	X							X		X		X		X		X		X		X		X	X	X																															15
	868,40	595-601 R.ANVMDYR.E	1	X	X	X	X	X							X		X		X		X		X		X		X		X-D	X-D	X	X									X		X								X				X		X	X				X-D	23
	884,39	595-601 R.ANVMDYR.E + Oxi (M)	1												X		X		X		X				X				X		X	X																															8
	889,44	44-50 R.DLTQQER.I	2	X	X	X	X	X							X		X		X		X		X		X		X		X	X	X	X																															16
	893,53	266-274 R.VAAHAVVAR.D		X		X	X																																		X																						4
	913,47	353-360 K.DQLTPSPR.E	1	X	X	X	X	X							X		D		D		D		X		X		X		X	X	X	X									X									X-D				X		X		X		X	22
	914,54	623-629 K.IRDLLER.Q		X		X		X																								X																	X	X	X	X	X	X	X	X	X	X		X	15
	926,47	59-66 K.ATYLDCIK.A																														X																															1
	931,44	132-140 K.TSCHTGLGR.S																														X																															1
	953,50	666-673 K.VREGTTYK.E				X																																																									1
	980,44	372-379 K.DEKSKCDR.W		X		X																																																									2
	993,51	633-641 K.RFGVNGSEK.S	1																						X							X-D																														D	4
	1024,50	594-601 R.RANVMDYR.E	1																								X																																				1
	1030,55	625-632 R.DLLERQEK.R	1																													X																															1

Protein	MH^+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Ovotransferrin (P02789)	1035,55	219-228 K.DGKGDVAFVK.H		X		X											X								X		X			X																					X						X		X			9
	1041,57	289-298 R.KDQLTPSPR.E	1																																																X		X		X		X		X			4
	1047,53	540-548 K.YFGYTGALR.C		X		X	X	X							X		X		X				X		X		X		X	X													X						X	X		X		X		X		X	X			20
	1052,54	634-643 R.FGVNGSEKSK.F	1																													X-D																															1
	1057,59	680-688 K.FYTVISSLK.T		X	X	X	X	X							X		X		X		X		X		X		X		X	X	X												X																				16
	1067,50	289-298 K.AQSDFGVDTK.S	1			X	X	X							X		X		X		X		X		X		X		X	X	X	X											X	X	X	X	X				X						X		X		X	23
	1092,63	299-309 K.DLLFKDLTK.C		X																																															X						X		X			4
	1161,51	644-652 K.FMMFESQNK.D	2	X	X	X	X	X							X		X		X		X		X		X		X		X	X	X	X																									X			X			18
	1177,50	644-652 K.FMMFESQNK.D + Oxi (M)	2												X		X		X		X		X		X		X		X	X	X																																10
	1179,54	28-37 R.WCTISSPEEK.K																														X																															1
	1186,65	625-633 R.DLLERQEKR.F	1																																																				X							1
	1193,5	644-652 K.FMMFESQNK.D + 2 Oxi (M)	2																		X				X																																					2
	1201,60	299-309 K.SDFHLFGPPGK.K		X	X	X	X	X							X		X		X				X		X		X		X	X	X																				X						X		X			17
	1203,57	459-468 K.DSNVNWNNLK.G	4		X		X	X							X		X		X		X		X							X		X																						X			X		X			13
	1208,64	633-643 K.RFGVNGSEKSK.F	1																													X-D																															1
	1223,63	279-288 K.VEDIWSFLSK.A			X	X	X	X									X		X		X		X		X		X		X	X	X												X			X					X					X	X		X			19
	1227,72	620-629 K.ANKIRDLLER.Q	1		X																																																										1
	1276,63	364-374 R.IQWCAVGKDEK.S	1																													X																															1
	1299,74	623-632 K.IRDLLERQEK.R	1																													X																															1
	1308,64	120-131 K.GTEFTVNDLQGK.T	2	X		X		X							X		X-D		X		X		X		X-D		X-D		X-D	X-D	X-D	X									X		X		X		X				X	X			X		X	X	X	X			25
	1312,66	353-363 K.GTEFTVNDLQGK.T	2																													X																															1
	1329,69	299-310 K.SDFHLFGPPGKK.D		X	X			X											X		X		X		X		X			X	X																												X			11
	1331,67	458-468 R.KDSNVNWNNLK.G	4	X	X	X	X	X									X		X				D		X		X		X	X	X	X																			X				X		X				X	18
	1353,64	229-240 K.HTTVNENAPDQK.D	3	X	X	X	X	X							X		X		X				X		X		X		X		X										X		X																				15
	1365,72	266-278 R.VAAHAVVARDDNK.D	1		X	X																																																									2
	1376,71	189-200 K.LCRQCKGDPKTK.C	1	X																																																											1
	1376,63	642-652 K.SKFMMFESQNK.D	2			X																										X																															1
	1388,69	459-470 K.DSNVNWNNLKGK.K	4	X																												X																						X							3
	1392,63	642-652 K.SKFMMFESQNK.D + Oxi (M)	2																													X																										X			2

Protein	MH+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Ovotransferrin (P02789)	1393,78	316-327 K.DLLFKDSAIMLK.R		X				X																	X		X		X	X	X	X																			X				X		X		X		X	13
	1409,77	316-328 K.DLLFKDSAIMLK.R + Oxi (M)																							X	X				X		X																			X		X		X		X		X			9
	1436,74	119-131 K.KGTEFTVNDLQGK.T	2	X	X	X		X							X		X		X-D		X		X		X		X		X	X-D	X	X															X				X				X		X		X			20
	1455,84	623-633 K.IRDLLERQEKR.F	1																																																X		X				X				X	4
	1516,79	458-470 R.KDSNVNWNNLKGK.K	4	X																												X-D																														2
	1516,79	459-471 K.DSNVNWNNLKGKK.S	4																													X																														1
	1534,84	141-154 R.SAGWNIPIGTLLHR.G	1	X	X	X	X	X									X		X		X		X		X		X		X	X	X-D	X							X		X			X		X	X		X	X	X			X		X	X	X	X		X	29
	1564,84	480-493 R.TAGWVIPMGLIHNR.T	1	X		X		X																	X					X		X-D																			X		X		X		X		X		X	12
	1565,87	316-328 K.DLLFKDSAIMLKR.V + Oxi (M)			X																																																									1
	1593,83	630-643 R.QEKRFGVNGSEKSK.F	2	X																																																										1
	1614,74	204-218 R.NAPYSGYSGAFHCLK.D	1																													X																														1
	1644,88	458-471 R.KDSNVNWNNLKGKK.S	4																													X																														1
	1652,84	443-457 K.TDERPASYFAVAVAR.K		X		X		X									X		X				X		X		X		X	X	X	X														X	X												X		X	16
	1659,78	361-374 R.ENRIQWCAVGKDEK.S + Gln->pyro-Glu (N-term Q)	2																													X																														1
	1695,82	275-288 R.DDNKVEDIWSFLSK.A	1	X	X		X	X							X		X		X		X		X		X		X		X	X	X	X									X			X	X	X	X		X		X		X				X		X		X	26
	1746,93	674-688 K.EFLGDKFYTVISSLK.T			X			X							X		X		X		X		X		X		X		X	X	X	X													X															X			15
	1769,78	579-593 K.NLQMDDFELLCTDGR.R	2																													X																							X							2
	1777,86	644-657 K.FMMFESQNKDLLFK.D	2					X																	X-D		X-D		X-D	X		X																			X				X		X		X			10
	1780,93	443-458 K.TDERPASYFAVAVARK.D																														X																														1
	1874,91	361-376 R.ENRIQWCAVGKDEKSK.C + Gln->pyro-Glu (N-term Q)	2	X		X																																																								2
	1882,02	299-315 K.SDFHLFGPPGKKDPVLK.D		X				X																								X																			X				X		X		X			7
	1888,90	554-571 K.GDVAFIQHSTVEENTGGK.N	2	X	X			X							X-D		X		X-D		X		X-D		X-D		X		X	X-D	X	X																		X									X			16
	1890,95	361-376 R.ENRIQWCAVGKDEKSK.C	2																													X																				X		X				X				4
	1901,84	144-159 K.TCNPSDILQMCSFLEGK.- + Oxi(M)	1																																														X		X		X				X				4
	1925,88	579-594 K.NLQMDDFELLCTDGRR.A	2																													X																														1
	1959,96	155-173 R.GAIEWEGIESGSVEQAVAK.F	1		X	X	X	X							X-D		X		X		X		X		X		X		X	X-D	X	X							X				X			X	X		X		X		X		X		X		X-D		X	26

Protein	MH+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Ovotransferrin (P02789)	1973,03	602-619 R.ECNLAEVPTHAVVVRPEK.A + Gln->pyro-Glu (N-term Q)	1																			X																								1
	2005,05	625-641 R.DLLERQEKRFGVNGSEK.S	2																			X																		X							2
	2221,07	120-140 K.GTEFTVNDLQGKTSCHTGLGR.S	2																			X																		X							2
	2235,12	64-661 K.FMMFESQNKDLLFKDLTK.C	2																																						X			X				2
	2341,06	574-593 K.ADWAKNLQMDDFELLCTDGR.R	1																																					X							1
	2426,25	299-320 R.EGTTYKEFLGDKFYTVISSLK.T	5																			X																			X			X		X		5
	2450,25	642-661 K.SKFMMFESQNKDLLFK.D	2																			X																									1
	2474,17	199-221 K.TKCARNAPYSGYSGAFHCLKDGK.G	1	X																																				X							2
	2497,16	574-594 K.ADWAKNLQMDDFELLCTDGRR.A	2																			X																									1
	2498,38	328-351 K.SDFHLFGPPGKKDPVLKDLLFK.D	2																																						X			X				2
	2570,28	554-571 K.DQLTPSPRENRIQWCAVGKDEK.S	3																																					X		X			X				3
	2570,34	266-288 R.VAAHADVARDDNQVEDIWSFLSK.A	2																			X																									1
	2606,27	67-92 K.AIANNEADAISLDGGQVFEAGLAPYK.L	3																																	X	X											2
	2631,27	204-228 R.NAPYSGYSGAFHCLKDGKGDVAFVK.H	1																			X																									1
	2698,28	329-351 R.VPSLMDSQLYLGFEYYSAIQSMR.K	2																		X-D																											1
	2714,28	329-351 R.VPSLMDSQLYLGFEYYSAIQSMR.K + Oxi (M)	2																D																													1
	2744,31	275-298 R.DDNKVEDIWSFLSKAQSDFGVDTK.S	2			X																	X																						X			3
	2826,46	93-118 K.LKPIAAEVYEHTEGSTTSYYAVAVVK.K	3																		X	X																								X			3
	2854,39	93-119 K.RVPSLMDSQLYLGFEYYSAIQSMR.K	4																																					X	D			X			X			4
	2954,55	328-352 K.LKPIAAEVYEHTEGSTTSYYAVAVVKK.G	3																																						X			X		X				3
	2982,48	353-374 K.RVPSLMDSQLYLGFEYYSAIQSMRK.D	4																																					X	X			X		X				4
	3153,59	595-622 R.ANVMDYRECNLAEVPTHAVVVRPEKANK.I	1																			X																										1
	3618,82	266-298 R.VAAHAVVARDDNKVEDIWSFLSKAQSDFGVDTK.S	2			X																	X																									2
	3475,79	141-173 R.SAGWNIPIGTLLHRGAIEWEGIESGSVEQAVAK.F	2			X																	X																		X			X			X	2

Protein	MH^+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Vitellogenin-1 (P87498)	750,45	1679-1684 K.FIITTR.K																																																								X		1
	859,49	174-181 K.ANLVDVTK.S	1																																																							X		1
	864,49	271-278 K.QVAEVPPK.E	1																																																							X		1
	913,57	263-270 K.LALIEVQK.Q	1																																	X																						X		2
	914,57	783-790 K.ALLLSEIR.C																																																								X		1
	949,49	1767-1774 K.VTVASWMR.G																																		X																								1
	960,50	919-927 R.NIEDLAASK.M	1																																									X													X		2
	965,49	1767-1774 K.VTVASWMR.G + Oxi (M)																																		X																								1
	987,58	173-181 K.KANLVDVTK.S	1																																																							X		1
	1115,59	56-66 R.SEVEISGIGPK.L																																										X															1
	1144,67	768-778 R.LISSLQTGIGR.Q	1																																X	X																								2
	1148,59	1757-1766 K.NVNFDGEILK.V	2																																	X							X														X		3
	1158,65	831-840 R.LTELLNSNVR.L	2																																X	X								X													X		4
	1163,65	483-494 K.ALGNVGHPASIK.H	1																																																							X		1
	1197,61	733-742 K.MFGQELLFGR.L	1																																								X																1
	1203,65	1513-1524 R.IAALATTGQMAR.K	1																																	X																								1
	1212,76	135-145 R.GILNILELSLK.K	1																																X								X																2
	1228,71	472-482 K.SNIEEVLLALK.A	1																																	X							X	X													X		4
	1261,60	1081-1091 K.ILDDTDNQATR.N	2																																									X													X		1
	1270,62	1455-1465 R.NQGYQATAYVR.S	3																																								X																1
	1273,69	1874-1884 K.ATAVSLLEWQR.S	1																																	X										X													2

Protein	MH+	Peptide + Modification	Dea sites	1 LC-MS/MS	2 LC-MS/MS	3 LC-MS/MS	4 LC-MS/MS	5 LC-MS/MS	6 LC-MS/MS	7 MALDI-TOF	7 LC-MS/MS	8 MALDI-TOF	8 LC-MS/MS	9 MALDI-TOF	9 LC-MS/MS	10 MALDI-TOF	10 LC-MS/MS	11 MALDI-TOF	11 LC-MS/MS	12 MALDI-TOF	12 LC-MS/MS	13 MALDI-TOF	13 LC-MS/MS	14 MALDI-TOF	14 LC-MS/MS	15 MALDI-TOF	15 LC-MS/MS	16 MALDI-TOF	16 LC-MS/MS	17 LC-MS/MS	18 LC-MS/MS	19 LC-MS/MS	20 LC-MS/MS	21 LC-MS/MS	22 LC-MS/MS	23 LC-MS/MS	30 LC-MS/MS	31 LC-MS/MS	32 LC-MS/MS	33 LC-MS/MS	36 LC-MS/MS	37 LC-MS/MS	38 LC-MS/MS	39 LC-MS/MS	40 LC-MS/MS	41 LC-MS/MS	42 LC-MS/MS	43 LC-MS/MS	44 LC-MS/MS	45 MALDI-TOF	45 LC-MS/MS	46 MALDI-TOF	46 LC-MS/MS	47 MALDI-TOF	47 LC-MS/MS	48 MALDI-TOF	48 LC-MS/MS	49 MALDI-TOF	49 LC-MS/MS	50 LC-MS/MS	51 LC-MS/MS	Frequency
Vitellogenin-1 (P87498)	1331,75	1513-1525 R.IAALATTGQMAR.K + Oxi (M)	1																																				X																					X		2
	1365,75	1591-1602 K.VPGVTLYYQGLR.V	1																																											X	X															2
	1384,78	54-66 R.IRSEVEISGIGPK.L																																				X																								1
	1437,78	813-826 K.VAGNVQAQITPSPR.S	3																																			X-D	X							X		X	X											X		6
	1439,64	343-354 R.AANEENYESVWK.Q	2																																				X								X													X		3
	1439,80	365-377 R.YLLDLLPAAASHR.S																																				X	X								X													X		4
	1469,79	1041-1054 K.IQVTIQAGDQAPTK.M	3																																			X	X							X	X															4
	1566,84	1013-1026 R.NVPLYNAIGEHALR.M	2																																				X																							1
	1639,84	1027-1040 R.MSFKPVYSDVPIEK.I																																																										X		1
	1653,89	928-942 K.MTPVLLPEAVPDIMK.M																																												X														X		2
	1655,84	1027-1040 R.MSFKPVYSDVPIEK.I + Oxi (M)																																					X																							1
	1669,89	928-942 K.MTPVLLPEAVPDIMK.M + Oxi (M)																																												X														X		2
	1685,89	928-942 K.MTPVLLPEAVPDIMK.M + 2 Oxi (M)																																																										X		1
	1742,82	1451-1465 R.SDNRNQGYQATAYVR.S	4																																				X								X		X													3
	1842,99	1551-1567 R.YVPGVALVLGFSEAHQR.N	1																																			X																								1
	1892,81	694-708 R.GYSPDKDWETNYDFR.E	1																																																										X		1
	1980,13	1432-1450 R.NFLGDVIPPGITIVAQAVR.S	2																																			X	X							X																3
	2118,06	315-333 R.LQDLVETTYEQLPSDAPAK.A	2																																			X	X							X	X															4
	2122,19	1584-1602 R.SIDTVIKVPGVTLYYQGLR.V	1																																			X	X								X															3
	2144,10	674-693 R.VADPIEVGIAAEGLQEMFVR.G	1																																											X			X													2
	2160,10	674-693 R.VADPIEVGIAAEGLQEMFVR.G + Oxi (M)	1																																			X-D	X							X														D		4
	2265,09	72-91 R.IHSIEAAEYNGIWPTSSFSR.S	1																																				X																							1
	2304,24	284-303 R.GSLQYQFGSELLQLPVHLFK.I	3																																				X-D								X															2
	2430,28	315-336 R.LQDLVETTYEQLPSDAPAKALK.L	2																																												X															1

Protein	MH+	Peptide + Modification	Dea sites		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	Frequency
					LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	
Vitellogenin-2 (P02845)	702,39	713-718 K.ALQGWK.E	1																												X	X	X	X										X		6
	714,42	768-773 R.IANQIR.N	2																												X	X	X	X										X		5
	737,38	1590-1595 K.TFNEVK.F	1																												X	X	X	X										X		6
	760,42	813-820 R.AAVSVEGK.M																													X	X	X	X										X		5
	804,44	1337-1342 R.QEFPKR.K	1																														X													1
	845,38	109-115 K.FEYSSGR.I																															X													1
	871,40	620-626 K.LDSMSYR.Y																													X	X	X	X										X		4
	874,43	216-223 K.GTTAFSYK.L																														X	X	X										X		3
	886,46	754-761 K.TVVEPADR.N																													X	X	X	X										X		6
	890,46	1624-1632 K.SAGEATNLK.A	1																												X	X	X	X										X		4
	893,45	1732-1739 R.MPNGYLAK.N	1																												X		X	X										X		3
	896,48	1020-1027 K.LIGEHEAK.I																														X													2
	915,52	911-918 R.HKAFAVSR.N																																X													1
	922,46	166-172 R.YVIQEDR.K	1																												X	X	X	X										X		5
	958,52	919-927 R.NIGELGVEK.R	1																											X		X		X										X		6
	961,53	1453-1461 K.VSTELVTGR.F																												X	X	X	X			X									6

Protein	MH⁺	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16	
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS
Vitellogenin-2 (P02845)	967,47	878-885 R.MPMKFDAK.I																											
	973,55	1515-1523 R.MVVALTSPR.T																											
	988,55	746-753 K.ELLQQVMK.T	2																										
	989,54	1515-1523 R.MVVALTSPR.T + Oxi (M)																											
	1008,48	1324-1330 R.HEIYQYR.F	1																										
	1017,56	1771-1780 K.TVQLAGVDSK.C	1																										
	1023,55	1083-1091 R.ILGIDSMFK.V																											
	1025,55	1462-1471 R.FAGHPAAQVK.L	1																										
	1031,54	902-910 R.EETEIVVGR.H																											
	1048,65	1543-1552 R.LPLSLPVGPR.I																											
	1050,62	1092-1100 K.VANKTRHPK.N	1																										
	1069,60	260-268 R.QQLTLVEVR.S + Gln->pyro-Glu (N-term Q)	2																										
	1074,59	684-693 R.VEGLADVIMK.R																											
	1085,63	260-268 R.QQLTLVEVR.S	2																										
	1094,59	656-665 R.TMFPSAIISK.L																											
	1102,65	1531-1539 K.LPDIILYQK.A	1																										
	1110,59	656-665 R.TMFPSAIISK.L + Oxi (M)																											
	1114,62	919-928 R.NIGELGVEKR.T	1																										
	1114,62	1042-1051 K.IQLEIQAGSR.A	2																										
	1134,56	821-830 K.MTPPLTGDFR.L																											
	1150,56	821-830 K.MTPPLTGDFR.L + Oxi (M)																											

MH⁺	Peptide + Modification	17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
		LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
967,47	878-885 R.MPMKFDAK.I																X																	1
973,55	1515-1523 R.MVVALTSPR.T										X	X	X	X	X	X		X	X		X											X		10
988,55	746-753 K.ELLQQVMK.T													X	X	X	X																	4
989,54	1515-1523 R.MVVALTSPR.T + Oxi (M)							X			X								X															3
1008,48	1324-1330 R.HEIYQYR.F														X																			1
1017,56	1771-1780 K.TVQLAGVDSK.C												X	X	X	X			X															5
1023,55	1083-1091 R.ILGIDSMFK.V													X	X		X																	3
1025,55	1462-1471 R.FAGHPAAQVK.L														X	X	X																	3
1031,54	902-910 R.EETEIVVGR.H												X	X	X	X																		4
1048,65	1543-1552 R.LPLSLPVGPR.I									X	X		X	X	X	X	X				X													8
1050,62	1092-1100 K.VANKTRHPK.N																																	2
1069,60	260-268 R.QQLTLVEVR.S + Gln->pyro-Glu (N-term Q)									X	X								X															3
1074,59	684-693 R.VEGLADVIMK.R													X	X	X	X	X																5
1085,63	260-268 R.QQLTLVEVR.S							X	X	X-D	X		X	X	X		X	X		X	X											X		12
1094,59	656-665 R.TMFPSAIISK.L												X	X	X	X	X		X	X	X													8
1102,65	1531-1539 K.LPDIILYQK.A													X	X																			2
1110,59	656-665 R.TMFPSAIISK.L + Oxi (M)										X								X		X													3
1114,62	919-928 R.NIGELGVEKR.T										X	X	X		X	X	X															X		7
1114,62	1042-1051 K.IQLEIQAGSR.A									X			X		X	X	X	X		X														7
1134,56	821-830 K.MTPPLTGDFR.L												X	X	X	X	X			X	X													7
1150,56	821-830 K.MTPPLTGDFR.L + Oxi (M)										X								X		X													3

Protein	MH^+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Vitellogenin-2 (P02845)	1164,49	633-641 R.ADTYFDNYR.V	1																																						X	X	X	X	X																		5
	1167,54	18-27 FDIDPGFNSR	1																																							X	X	X	X			X														5
	1216,66	213-223 R.LTKGTTAFSYK.L																																													X																1
	1230,69	684-694 R.VEGLADVIMKR.N																																								X	X	X	X																		4
	1233,63	106-115 R.LFKFEYSSGR.I																																						X					X	X	X																4
	1246,69	684-694 R.VEGLADVIMKR.N + Oxi (M)																																															X														1
	1295,68	135-145 R.GILNMFQMTIK.K	2																																							X	X	X	X																		4
	1323,64	18-28 K.FDIDPGFNSRR.S	1																																									X																				1
	1324,74	1472-1482 K.LEWPKVPSNVR.S	1																																					X					X			X																3
	1327,67	135-145 R.GILNMFQMTIK.K + Oxi (M)	2																																											X																	2
	1342,67	695-705 R.NIPFAEYPTYK.Q	1																																					X		X	X		X	X		X													X		7
	1374,86	831-842 R.LSQLLESTMQIR.S	2																																					X		X	X	X	X	X																	6
	1383,76	754-766 K.TVVEPADRNAAIK.R	1																																									X																				1
	1401,71	342-353 R.IANADNLESIWR.Q	2																																				X-D	X		X		X	X	X	X	X		X											X		11
	1409,75	615-626 K.LLSPKLDSMSYR.Y																																													X																1
	1418,77	1008-1019 R.DASFIQNTYLHK.L	2																																					X			X	X	X	X																	5
	1434,76	831-842 R.LSQLLESTMQIR.S + Oxi (M)	2																																				X	X								X		X													4

Protein	MH+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Vitellogenin-2 (P02845)	1445,77	642-655 R.VGATGEIFVVNSPR.T	1																																			X	X		X	X	X	X	X	X	X	X	X											X		12
	1437,76	1411-1423 R.VQVFVTNLTDSSK.W	2																																									X	X																	2
	1457,82	1702-1714 K.TITIQVPLWMAGK.T	1																																						X	X	X	X	X	X	X															7
	1448,69	226-239 K.QSDSGTLITDVSSR.Q + Gln->pyro-Glu (N-term Q)	1																																			X																								1
	1457,81	260-271 R.QQLTLVEVRSER.G	2																																				X																							1
	1458,83	733-745 K.ILGQEVAFININK.E	3																																			X				X	X	X	X	X																6
	1461,65	1148-1159 K.KPMDEEENDQVK.Q	2																																							X	X	X	X																	4
	1465,71	226-239 K.QSDSGTLITDVSSR.Q	1																																			X	X		X	X	X	X	X	X	X		X											X		11
	1502,65	954-966 R.EHFAMQGPDSMPR.K	1																																							X	X	X																		3
	1473,82	1702-1714 K.TITIQVPLWMAGK.T + Oxi (M)	1																																			X																								1
	1502,65	58-71 K.LEISGLPENAYLLK.V	1																																			X			X	X	X	X	X	X																7
	1532,75	630-641 K.VIRADTYFDNYR.V	1																																												X															1
	1539,86	754-767 K.TVVEPADRNAAIKR.I	1																																									X																		1
	1560,74	74-86 R.SPQVEEYNGVWPR.D	2																																			X	X-D		X	X	X	X	X		X		X											X		10
	1566,82	1028-1041 K.IVLMPVHTDADIDK.I																																								X	X	X	X																	4
	1560,85	719-732 K.ELPTETPLVSAYLK.I																																								X	X	X	X	X	X													X		7
	1566,82	272-286 R.GSAPDVPMQNYGSLR.Y	2																																			X	X		X	X	X	X	X		X		X													9
	1581,87	1377-1391 K.TPVLAAFLHGISNNK.K	2																																							X	X	X	X														X			5
	1589,75	1147-1159 R.KKPMDEEENDQVK.Q	2																																								X	X	X		X															4
	1591,75	289-302 R.FPAVLPQMPLQLIK.T	2																																				X		X	X	X	X	X	X																7
	1663,89	364-379 R.WLLSAVSASGTTETLK.F																																				X	X		X	X	X	X	X		X		X											X		10
	1607,75	272-286 R.GSAPDVPMQNYGSLR.Y + Oxi (M)	2																																				X								X		X													3
	1706,89	224-239 K.LKQSDSGTLITDVSSR.Q	1																																				X		X			X	X	X	X															6

Protein	MH^+	Peptide + Modification	Dea sites	1	2	3	4	5	6	7		8		9		10		11		12		13		14		15		16		17	18	19	20	21	22	23	30	31	32	33	36	37	38	39	40	41	42	43	44	45		46		47		48		49		50	51	Frequency
				LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	MALDI-TOF	LC-MS/MS	LC-MS/MS	LC-MS/MS	
Vitellogenin-2 (P02845)	1679,85	1006-1019 R.SRDASFIQNTYLHK.L	2																																						X																		1
	1760,87	1638-1653 K.IGSHEIDMHPVNGQVK.L	2																																				X		X	X	X	X	X																	6
	1709,97	1377-1392 K.TPVLAAFLHGISNNKK.T	2																																									X																		1
	1711,90	695-708 R.NIPFAEYPTYKQIK.E	2																																				X																								1
	1757,80	578-592 K.ESNMQVASFVYSHMK.S	2																																									X		X																2
	1777,91	456-471 K.EALQPIHDLADEAISR.G	1																																		X	X		X	X	X	X	X	X	X													X		10
	1801,95	666-683 K.LMANSAGSVADLVEVGIR.V	1																																						X	X	X	X	X																5
	1815,91	72-86 K.VRSPQVEEYNGVWPR.D	2																																				X		X			X			X	X														5
	1914,10	287-302 R.YRFPAVLPQMPLQLIK.T	1																																				X-D		X		X	X	X	X	X															7
	1817,94	666-683 K.LMANAGSVADLVEVGIR.V + Oxi (M)	1																																		X																									1
	1819,99	363-379 R.RWLLSAVSASGTTETLK.F																																										X																		1
	1947,03	1393-1410 K.TGGLQLVVYADTDSVRPR.V	1																																		X				X	X	X	X	X	X															7
	1930,10	287-302 R.YRFPAVLPQMPLQLIK.T + Oxi (M)	2																																		X	X							X	X															4
	1944,94	1147-1162 R.KKPMDEEENDQVKQAR.N	3																																						X																						1
	1972,97	774-790 R.NSIAGQWTQPVWMGELR.Y	3																																				X			X	X	X	X		X		X													7

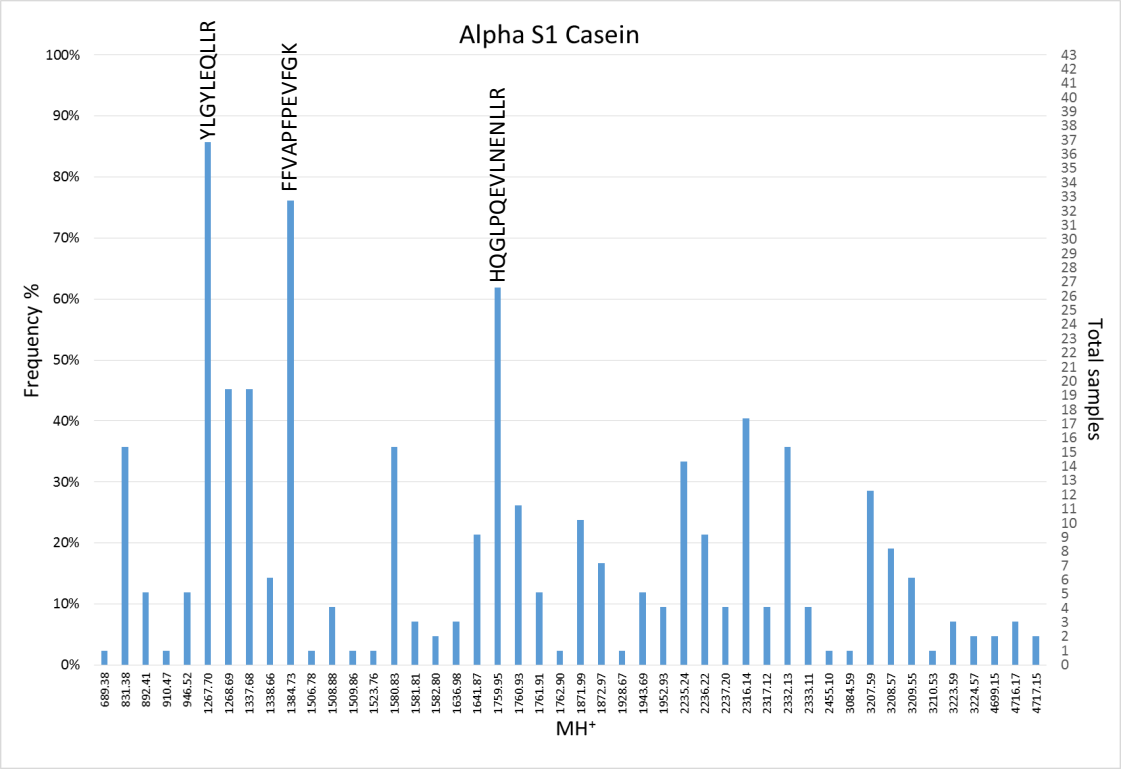


Figure VI.5. Observed frequency percentage for peptides of alpha S1 casein in the milk containing samples.

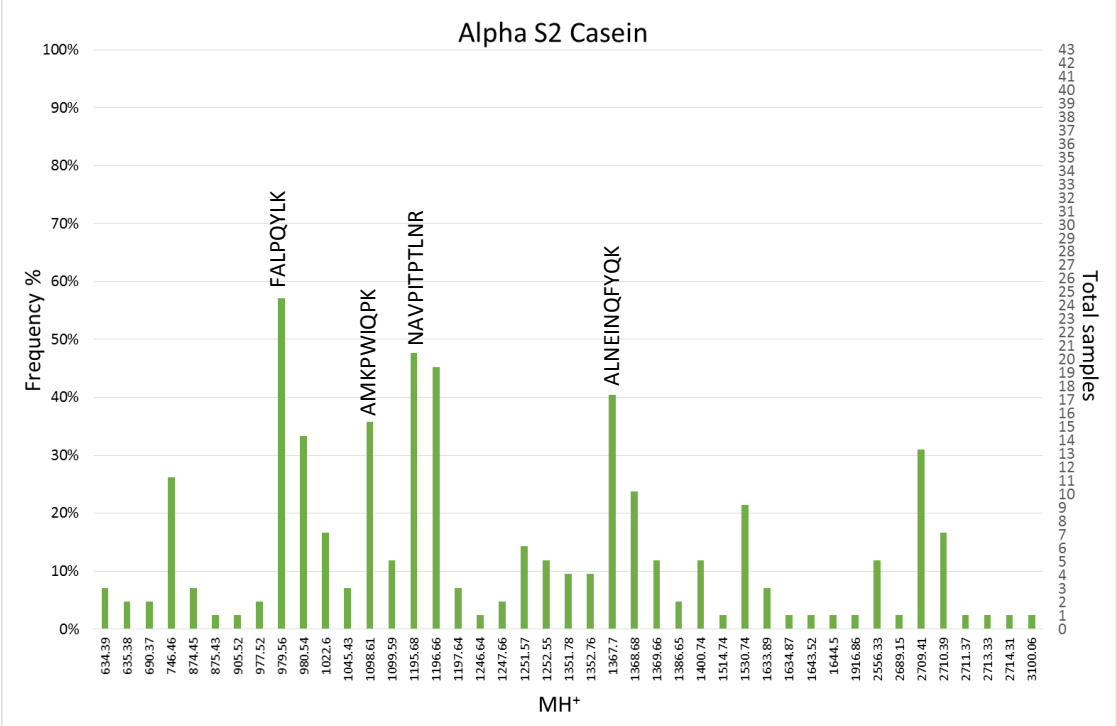


Figure VI.6. Observed frequency percentage for peptides of alpha S2 casein in the milk containing samples.

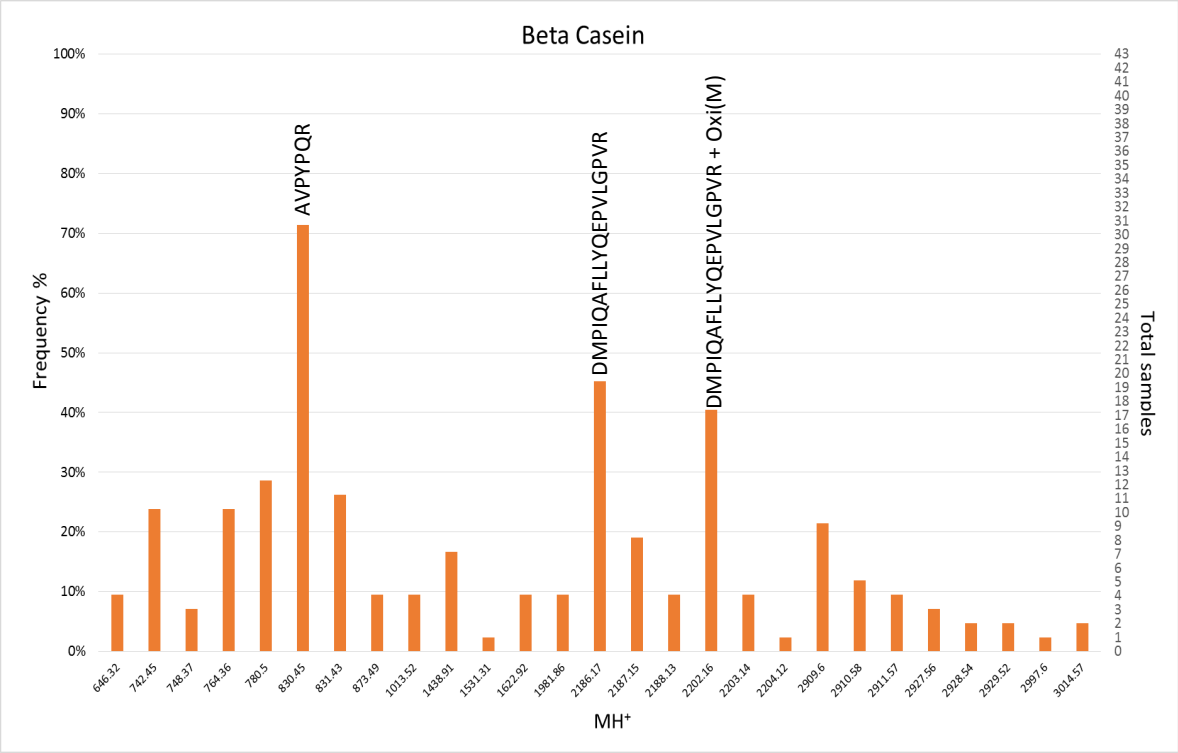


Figure VI.7. Observed frequency percentage for peptides of beta casein in the milk containing samples.

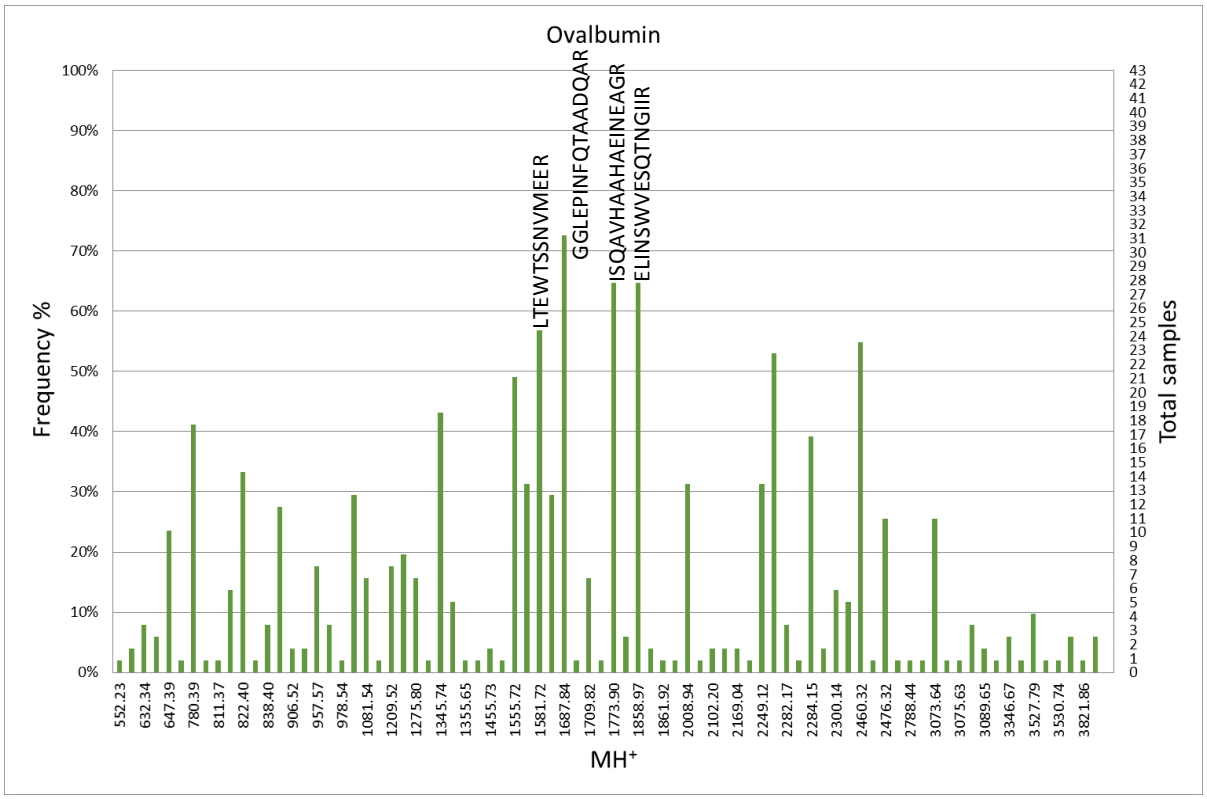


Figure VI.8. Observed frequency percentage for peptides of ovalbumin in the eggs containing samples.

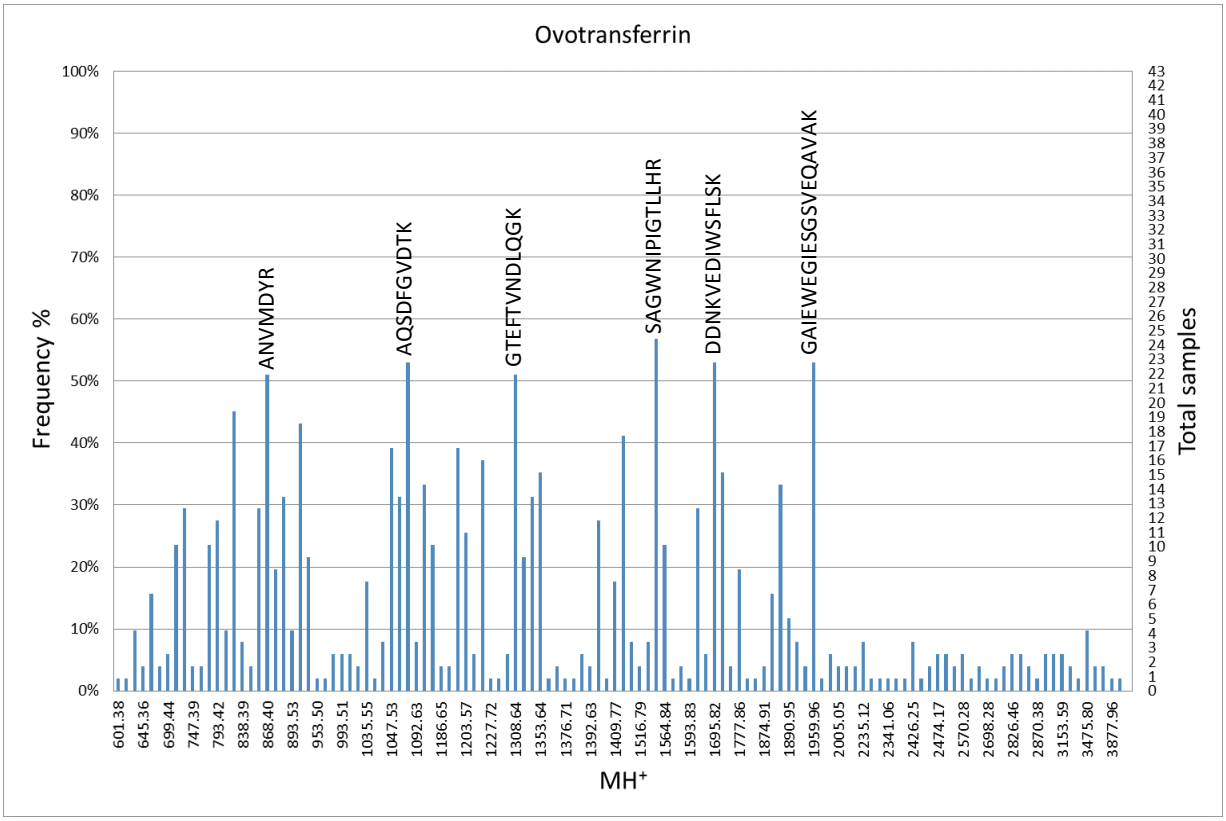


Figure VI.9. Observed frequency percentage for peptides of ovotransferrin in the eggs containing samples.

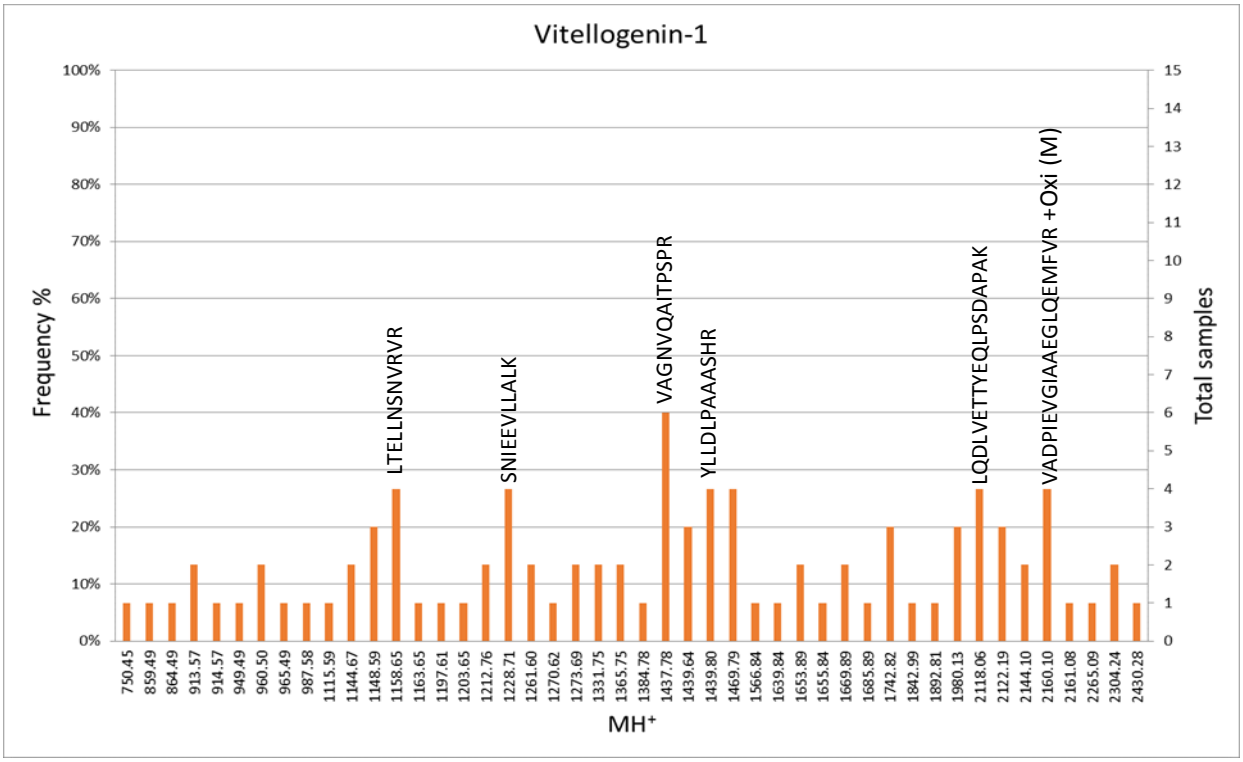


Figure VI.10. Observed frequency percentage for peptides of vitellogenin-1 in the eggs containing samples.

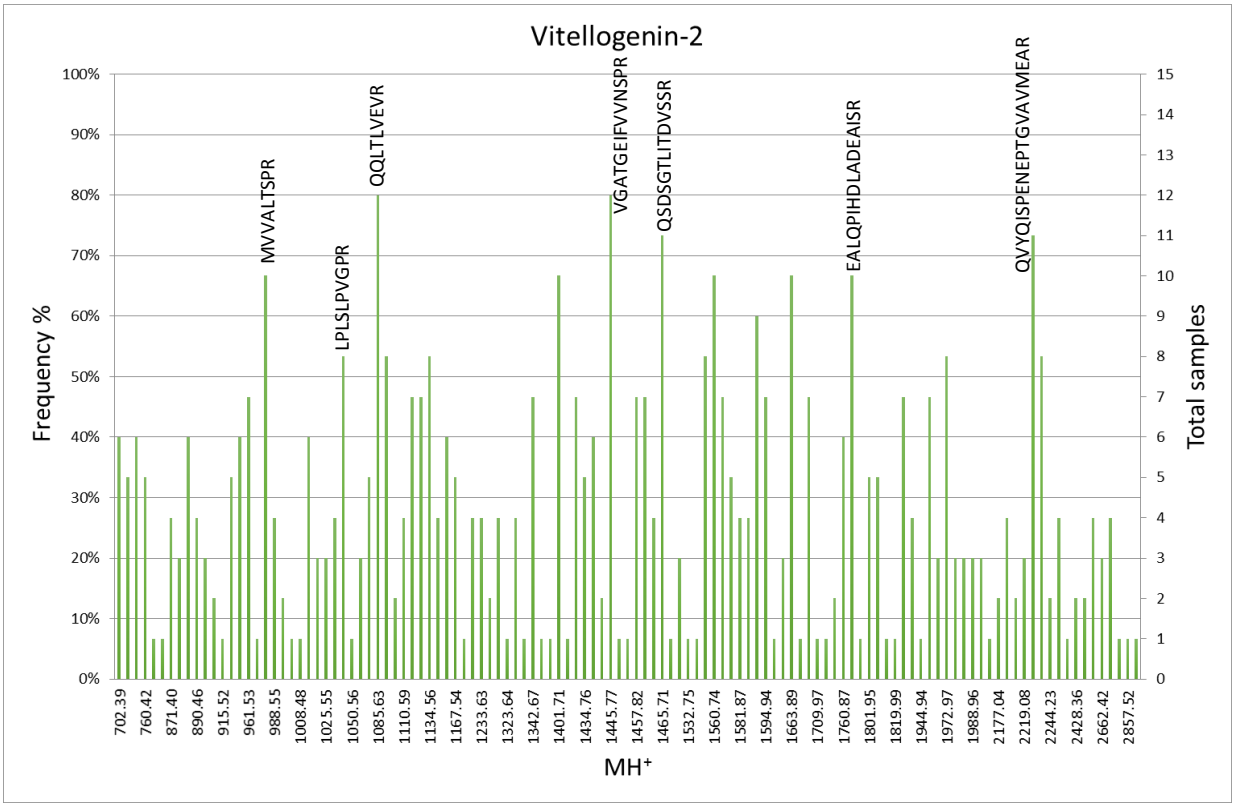


Figure VI.11. Observed frequency percentage for peptides of vitellogenin-2 in the eggs containing samples.

TableVI.5. Proteotypic peptides and selected transitions for each protein as obtained by Skyline software (http://sciex.com/products/software/skyline-software)

Protein	Peptide	Precursor (*m/z*)	Product (*m/z*)	Collision energy (eV)
Alpha S1 Casein	^{106}YLGYLEQLLR^{115}	634.36	1104,64 991,56 934,54 771,47	22
	^{38}FFVAPFPEVFGK^{49}	692.87	1237,66 1090,59 991,52 920,49 823,43	25
	^{23}HQGLPQEVLNENLLR^{37}	880,48	1437,81 1324,72 1227,67 1099,61 970,57	32
	^{38}FVVAPFPEVFR^{48}	654.36	1160.6463 1061.5778 962.5094 891.4723 794.4196	23
Alpha S2 Casein	^{189}FALPQYLK^{196}	490.28	832,49 761,46 648,37 551,32	17
	^{130}NAVPITPTLNR^{140}	598.34	1059.5582 988.5211 931.4996 834.4468 687.3784	21
	^{204}AMKPWIQPK^{212}	549.81	1027,58 896,53 768,44 671,39	19
	^{96}ALNEINQFYQK^{106}	684.35	1221,44 1107,40 978,36 827,40 713,36	25
Beta Casein	^{192}AVPYPQR^{198}	415.73	759,41 660,35 563,29	14
	^{199}DMPIQAFLLYQEPVLGPVR^{217}	1093.59	1383,80 1270,72 1157,63	39
	^{199}DMPIQAFLLYQEPVLGPVR^{217} + Oxi (M)	1101.59	1383,80 1270,72 1157,63	39

Ovalbumin	^{265}LTEWTSSNVMEER^{277}	579.85	1045.56 944.52 815.47 702.39 589.31	20
	^{128}GGLEPINFQTAADQAR^{143}	844.42	860.42 1331.67 1234.62 1121.53 1007.49	30
	^{324}ISQAVHAAHAEINEAGR^{340}	887.45	1374.69 1275.62 1138.56 1067.52 996.46	32
	^{144}ELINSWVESQTNGIIR^{159}	929,99	1389.71 1302.68 1116.60 1017.53	33
Ovotransferrin	^{595}ANVMDYR^{601}	434.70	797.36 683.32 584.25 453.21	15
	^{289}AQSDFGVDTK^{298}	534.25	996.46 868.40 781.37 666.35	19
	^{120}GTEFTVNDLQGK^{131}	654.83	1251.62 1150.57 1021.53 874.46 773.42 674.35	23
	^{141}SAGWNIPIGTLLHR^{154}	767.93	1319.76 1133.68 1019.64 906.55 809.50	27
	^{275}DDNKVEDIWSFLSK^{288}	848.41	1351.73 1223.63 1124.56 995.52 880.49	30
	^{155}GAIEWEGIESGSVEQAVAK^{173}	980.49	1403.70 1274.66 1217.64 1104.55	35
	^{831}LTELLNSNVR^{840}	579.83	1045.56 944.52 815.47 702.39	28

			589.31	
Vitellogenin-1	^{472}SNIEEVLLALK^{482}	614.86	1141.68 1027.64 914.56 785.51 656.47	22
	^{813}VAGNVQAQITPSPR^{826}	719.39	1210.65 1096.61 997.54 869.48 798.45	26
	^{365}YLLDLLPAAASHR^{377}	720.40	1276.74 1163.65 1050.57 935.54 822.46	26
	^{315}LQDLVETTYEQLPSDAPAK^{333}	1059.53	1449.67 1350.61 1221.56 1120.52 1019.46	34
	^{674}VADPIEVGIAAEGLQEMFVR^{693} + Oxi (M)	1080.56	1436.72 1379.70 1266.61 1195.58 1124.54	39
Vitellogenin-2	^{1515}MVVALTSPR^{1523}	487.28	842.51 743.44 644.37 573.34	17
	^{1543}LPLSLPVGPR^{1552}	524.83	935.57 838.51 725.43 638.40 525.31	18
	^{260}QQLTLVEVR^{268}	543.32	957.57 829.51 716.43 615.38	19
	^{919}NIGELGVEKR^{928}	557.81	1000.58 887.49 830.47 701.43 588.35	20
	^{642}VGATGEIFVVNSPR^{655}	723.39	1289.68 1218.65 1117.60 1060.58 931.54 818.45	26

	^{226}QSDSGTLITDVSSR^{239}	733.36	1250.62 1135.59 1048.56 991.54 890.49 777.41	26
	^{456}EALQPIHDLADEAISR^{471}	889.46	1464.74 1336.69 1239.63 1126.55 989.49	32
	^{240}QVYQISPFNEPTGVAVMEAR^{259}	1118.56	1420.69 1273.62 1159.58 1030.54	40

Publications

Microchemical Journal 126 (2016) 341–348

Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc

Proteomic strategies for cultural heritage: From bones to paintings☆

Roberto Vinciguerra [a], Addolorata De Chiaro [a], Piero Pucci [a], Gennaro Marino [b], Leila Birolo [a,*]

[a] Dipartimento di Scienze Chimiche, Università di Napoli Federico II, Complesso Universitario Monte S.Angelo, Via Cinthia, Napoli, Italy
[b] Università degli Studi 'Suor Orsola Benincasa', 80132 Napoli, Italy

ARTICLE INFO

Article history:
Received 30 July 2015
Received in revised form 18 December 2015
Accepted 18 December 2015
Available online 29 December 2015

Keywords:
Proteomics
Ancient proteins
LC-MS/MS

ABSTRACT

In recent years, proteomics procedures have become increasingly popular for the characterization of proteinaceous materials in ancient samples of several cultural heritage objects. The knowledge of the materials used in a work of art is crucial, not only to give an insight in the historical context of objects and artists, but also to analyse degradation processes taking place in aged objects and to develop appropriate conservation and/or restoration treatments. However, protocols routinely applied for typical modern samples still need to be fully adapted to take into account the low amount of proteinaceous material, the heterogeneity and the unusual physical state of the samples, as well as the high levels of damage found in ancient samples. This paper deals with some examples of the adaptation of classical proteomic strategies in the analysis of ancient samples to meet the different aims in the cultural heritage field.

1. Introduction

Proteomics is typically considered to be associated with the study of living organisms; however, its inherently multidisciplinary nature has recently led to the application of proteomic methods to oddly assorted areas ranging from forensics, food analysis, clinical medicine and even for studying the origins of life on earth; it has been proven to be an effective tool also for the scientific analysis of artworks [1].

Proteomics for cultural heritage, i.e. the identification of proteinaceous material used by artists in their masterpieces and found in archaeological remains, is still in its infancy, with the first paper dating back to the early 2000 [2]. In particular, protocols routinely applied for typical modern samples still need to be fully adapted to take into account the low amount of proteinaceous material, the heterogeneity and the unusual physical state of the samples, as well as the high levels of damage found in ancient samples.

Although analyses in proteomics are *per se* invasive, modern mass spectrometry instrumentations enable the characterization of proteins with extremely high sensitivity even in crude mixtures in which the dynamic range of components abundance exceeds 1000-fold, and on very limited amount of sample, typically less than 10 μg. Modern instruments are therefore perfectly adequate to afford the minimal quantities of ancient samples. However, all the steps of the proteomic procedure need to be thoughtfully adapted, from the optimization of specific protocols for sample preparation to the development of data analysis tools that can cope with ancient, damaged samples.

Moreover, although for merely identifying purposes, the detection of as few as two peptides is sufficient to properly pinpoint the protein, the characterization of the modifications induced by ageing and deterioration processes requires a deeper examination of most of the protein primary structure.

This paper deals with the reasoning behind the choice of the steps to be carried out in the analysis of ancient samples and with the adaptation of classical protocols to meet the different aims. It is not intended to cover the whole panel of possibilities and tricks but, rather to illustrate the logic that might address the choice of sample treatment and data analysis to get the most of the experiment.

2. Material and methods

Ammonium hydrogen carbonate (Ambic), Ethylenediaminetetraacetic acid (EDTA); Tri(hydroxymethyl)aminomethane (TRIS), Urea, GuHCl and TPCK-treated trypsin were from Sigma; Formic acid and Acetonitrile (ACN) were purchased from Baker. Deionized water was obtained from Millipore cartridge equipment. Hydrochloric acid was purchased from Carlo Erba.

Models of paint layers were prepared with milk as binders and azurite ($Cu_3(CO_3)_2(OH)_2$), minium (Pb_3O_4), calcite ($CaCO_3$), and vermilion (HgS) as pigments on glass slides and skimmed milk as control without

☆ Selected papers presented at TECHNART 2015 Conference, Catania (Italy), April 27–30, 2015.

* Corresponding author at: Dipartimento di Scienze Chimiche, Università di Napoli, Federico II, Napoli, Italy.
E-mail addresses: roberto.vinciguerra@unina.it (R. Vinciguerra), addolorata.dechiaro@unina.it (A. De Chiaro), pucci@unina.it (P. Pucci), gmarino@unina.it (G. Marino), birolo@unina.it (L. Birolo).

http://dx.doi.org/10.1016/j.microc.2015.12.024
0026-265X/

pigments. Paint replicas were left to dry at RT on the bench for one month. Bone sample was a fragment from human bone of the first century A.D.

2.1 Sample treatment

Urea pre-treatment: 10 μL of a solution of 6 M Urea was added to micro-samples (ca 300–800 μg) and incubated for 10 min at RT, followed by sonication for 20 min. Urea was then 6-fold diluted with water.

EDTA pre-treatment: about 100 μL of a solution of 0.5 M EDTA was added to the bone fragment 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 Urea protocol described above was applied.

HCl pre-treatment: 50 μL of 0.6 M HCl was added to the bone fragment and incubated at 4 °C for 4 h. After centrifugation for 2 min at 10,000 rpm in a benchtop microfuge the supernatant was removed and washed with 20 μL of 10 mM Ambic. Washes were repeated for four times. 100 μL of 50 mM Ambic was added and sample was left at 65 °C for 3 h. After centrifugation at 10,000 rpm for 15 min the supernatant was removed.

GuHCl pre-treatment: 200 μL of 0.6 M HCl was added and incubated at 4 °C for 18 h. After centrifugation for 1 min at 14,000 rpm in a benchtop microfuge, and the acid-insoluble pellet washed three times with 200 μL of distilled water. The pellet was incubated at 4 °C for 72 h in a buffer containing 100 mM Tris and 6 M GuHCl at pH 7.4. The sample was then centrifuged for 1 min at 14,000 rpm in a benchtop microfuge. The supernatant was buffer-exchanged into 10 mM Ambic using 3 K molecular weight cut-off Amicon Ultra, centrifugal filter unit.

2.2 Protein digestion and LC-MS/MS analysis

After any pre-treatment of the sample, enzymatic digestion was carried out as in the minimally invasive proteomic analytical procedure described by Leo *et al.* [3]. Briefly, trypsin was added to a final concentration of 10 ng/μL to micro-samples (ca 300–800 μg) as directly suspended in 50 μL 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 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 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 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.

2.3 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 Chordata as taxonomy restriction for protein identification in paint reconstruction samples and with *Homo sapiens* for bone samples.

A licenced 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, formation of pyroglutamic acid from glutamine residues at the N-terminal position of peptides, and deamidation at asparagines and glutamines were considered as variable modifications [4]. 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 [4]. 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 generally 31 for paintings and 43 for human samples. Spectra with Mascot score <10 were rejected.

3. Results and discussion

3.1 The minimal protocol

The commonly used “bottom-up” approach to identify proteins is based on the enzymatic digestion of the proteins and can be directly performed in heterogeneous phase on a sample fragment followed by mass spectrometric analyses of the released peptides (Fig. 1). Even when the protein is embedded in a complex mixture such as that of a painting layer, few peptides released by the protease without any pre-treatment of the sample are enough to identify the protein [3], without significantly affecting the sample itself. This avoid the extraction of the whole protein from the sample using harsh methods, while digestion can rather be carried out by depositing on the surface of a small sample an aqueous neutral solution containing the enzyme that directly trims protruding peptides. The solution will then be gently removed and, once released, peptides can be analysed by mass spectrometric methods such as MALDI-TOF and LC-MS/MS, and database searches by bioinformatics tools such as Mascot (www.matrixscience.com) allow protein identification. This micro-invasive protocol (i.e. the intervention is intrinsically invasive but requires only a minimal quantity of material to work with) proved to be successful when applied to fragments of paintings from the collapsed vault of the Basilica di S. Francesco in Assisi [3], and samples collected from the Camposanto Monumentale in Pisa [5], and it was not significantly affected by the pigment that was present in the sample, i.e. the different metals do not actually affect the quality of the results [3]. It is worth mentioning that this protocol is closely similar to the procedure adopted when bio-cleaning of works of art is carried out [6], and therefore the meaning of “destructive or invasive” approach should be resized and this can be safely considered as a minimally invasive or rather micro-invasive procedure.

Protein identification can fail because of overwhelming proteins from unavoidable contaminants. Samples coming from artwork have, indeed, an intrinsic contamination problem that originates from i.e. environmental exposure, restoration interventions, and so on, that cannot be overcome by just operating all the chemical manipulations in controlled conditions as in ordinary proteomic analysis. To circumvent unavoidable problems arising from “historical” contaminations, an exclusion list of the peptides which, in a first LC-MS/MS run allowed for identification of keratins or other protein contaminants and that in a further subsequent run have to be ignored by the mass spectrometer for fragmentation, can be adopted. It is a sort of instrumental trick to avoid “waste” of the mass spectrometer's time in fragmenting “useless” peptides derived from trypsin autodigestion or common protein contaminants [3]. The list is created *ad hoc* from the raw data of the LC-MS/MS analysis and a second analysis is carried out with the same LC-MS/MS method, but for the addition of the exclusion list.

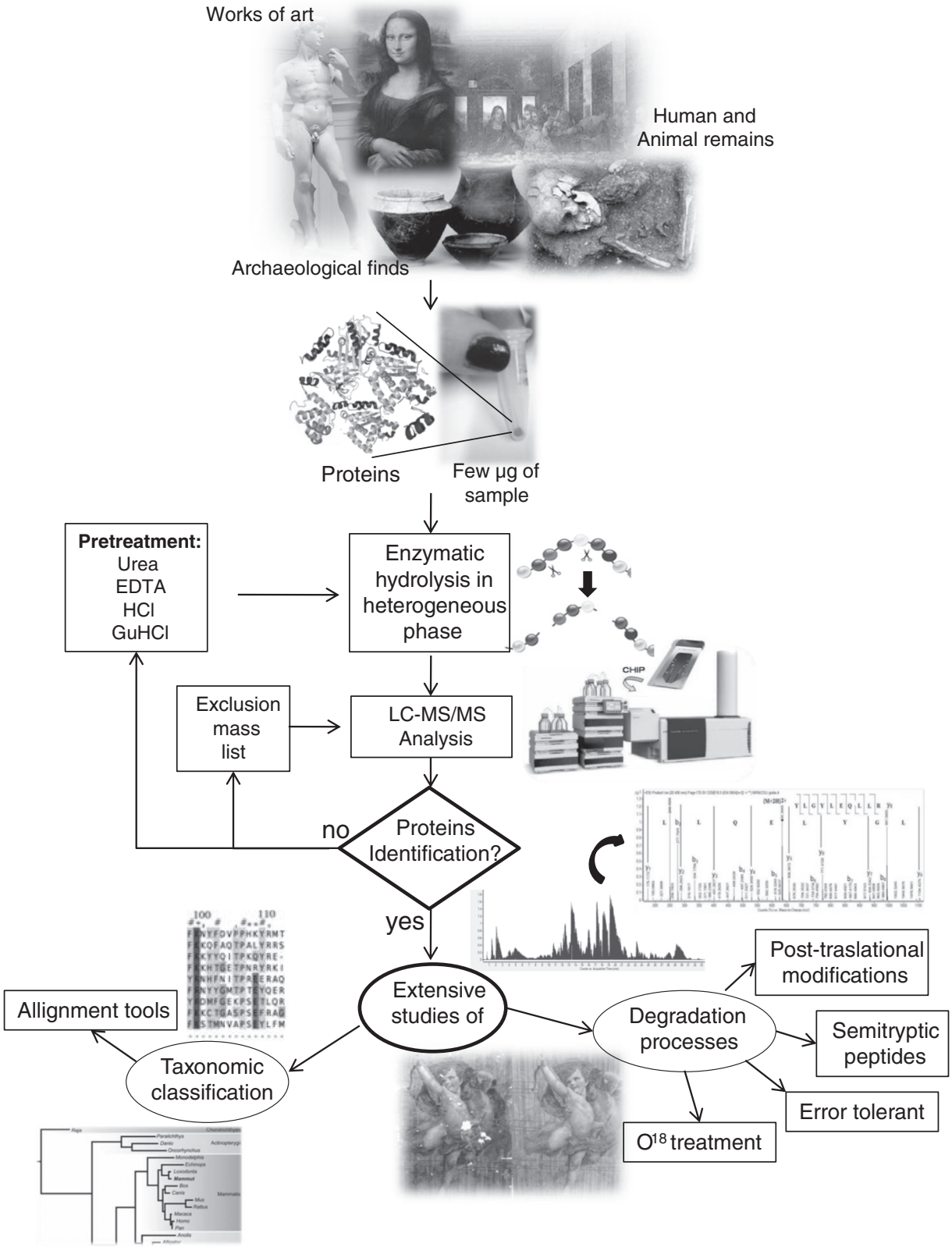


Fig. 1. General flow chart of proteomic investigations.

3.2 Sample pre-treatments

It is evident from what outlined above that accessibility of proteases to the protein in the sample is likely the most important aspect for a successful proteomic experiment in cultural heritage application (obviously beside the actual presence in the sample of proteinaceous material). Whatever reduces the proteolysis yield and, therefore, determines an inefficient and poor production of peptides, would sensibly affect

protein identification. For instance, the persistence of stabilizing interactions such as those occurring in structured proteins can greatly impair the efficacy of the enzymatic digestion. By reasoning as in classical biochemical experiments, where denatured, unfolded proteins are digested much more efficiently than structured, folded proteins, we introduced a denaturing step before trypsin digestion to "open" residual structural elements in proteins in the perspective that flexibility of the polypeptide substrate is an absolute requirement for the protease to properly hydrolyze peptide bonds. In the hypothesis that treatment with traditional protein denaturing agents such as urea or guanidinium chloride will make digestion sites more amenable to protease attack even in an unusual "dehydrated, non-soluble" physical state experienced by proteins in works of art or archaeological remains, we tested the denaturing pre-treatment in some cases where the minimally invasive approach described above failed. As an example, Table 1 reports a comparison of the results obtained by the minimally invasive approach, and the analysis carried out on the same pictorial sample after pre-treatment with 6 M urea, followed by dilution and tryptic digestion. The results clearly show that the pre-treatment with urea improves the quality of identification. We successfully applied the urea pre-treatment protocol to some gilding samples [7], to ink samples from the Qumran archaeological site [8], and to a small stone flake from Sibudu Cave dated 49,000 years ago [4]. This urea pre-treatment protocol is slightly more aggressive and therefore more invasive than that described above, and it should be used as a second attempt, when the above protocol with only protease in bicarbonate buffer fails or when more peptides or higher quality spectra are required. It is worth noting that it can be also directly applied on the same sample after the first attempt with the protease in the simple ammonium bicarbonate buffer has failed, thus not reducing the sample size. In principle, after the removal of the bicarbonate buffer containing the protease from the solid sample, the urea pretreatment can be carried out without any further step in between.

Moreover, it was very recently demonstrated [9] that the introduction of a deglycosylating step with PNGaseF before the digestion with the protease, greatly improves proteins identification when egg containing samples are analysed. This further demonstrates that accessibility of peptidic linkages to proteases is the absolute requirement for a reliable identification.

While two or very few peptides can be enough to merely identify materials, more extended sequence coverage could be needed for more detailed analyses, such as species discrimination, and/or conservation state evaluation. In fact, if the goal is to discriminate the organism of origin of proteins that are highly conserved throughout the evolution and among species, it is important to identify proteotypic peptides, i.e. peptides that are unique to the protein sequence specific for an individual organism and not in common to other species.

This can be easily explained with the example of milk proteins: among the peptides that have been identified in the analysis of an ancient food residue by Hong *et al.* [10], some are shared by goat/sheep and cattle sequences while others can be used to discern the origin of milk in the ancient residue. Similarly, detection of β-lactoglobulin in dental calculus is *per se* a direct evidence of milk consumption and can constitute a signature for adoption of dairying habits. Most interestingly, identification of specie-specific peptides allowed discrimination of the origin of dairy product whether they are cattle, sheep or goat dairy product [11]. Specie-specific information that are peculiar of proteomics are therefore extremely useful for delineating domestication timelines and paleodietary habits. To meet aims like these, however, the mild protocols described above might be not always adequate and alternative procedures might be required, to obtain a higher number of peptides and consequently higher sequence coverage, thus increasing the probability of detecting proteotypic peptides.

This is definitively the case of collagen from bones, where an in depth analysis might be needed to gain extremely important information that can be used for a wide range of purposes, from taxonomic analyses [4,12–14], to the characterization of degradation processes [15]. Proteins and collagen in particular, are indeed gaining momentum and are now supporting DNA in evolutionary studies, since they are more stable than nucleic acids, can persist much longer, and, moreover, the survived molecules bear the signature of time.

As commented by the authors of a tremendous paper in Nature this year [14,16] "ancient proteins could now prove as revolutionary as DNA for studying the tree of life", since we could be able to find proteins that are orders of magnitude older than the oldest DNA discovered so far. Intrinsic collagen protein stability and its entrapment within the hydroxylapatite protective cage make the collagen the longest surviving protein in ancient bone [17]. Consequently and adversely, collagen is highly resistant to extraction and it can be quite challenging to obtain a good sequence coverage that would allow highly reliable sequence comparisons for taphonomic studies (i.e. studies of decaying organisms over time). Procedures need to be optimized for protein extraction from ancient bones, and several procedures have been proposed [16–18 and references therein]. Table 2 reports the sequence coverage of human COL1α1 (collagen alpha-1(I)) and COL1α2 (collagen alpha-2(I)), the two chains constituting type I collagen, as obtained in the analysis of a human bone dated the first century A.D., using selected different extraction procedures. It can be noted that even with the minimally invasive protocol, where the bone fragment is simply deepened in a trypsin solution, the number of peptides detected can be high enough to obtain good sequence coverage (24% and 23% for COL1α1 and COL1α2, respectively. However, when decalcification is carried out with EDTA 0.5 M, nearly the whole sequences can be verified, reaching 75% and 72% of sequence coverage for mature COL1α1 and COL1α2, respectively.

However, while decalcification with EDTA seems to be the ultimate choice at this stage for collagen protein sequence coverage, it might not be the best choice when non-collagenic proteins (NCPs) are searched [17,19]. NCPs have potentially higher phylogenetic value

Table 1
Proteins identified in the paint replica containing minium and milk by LC-MS/MS. Aliquots were treated in heterogeneous phase with trypsin with the minimally invasive protocol either without any sample pre-treatment or with Urea pre-treatment and the resulting peptide mixtures were analysed by LC-MS/MS. Proteins were identified in the Uniprot database with Mascot MS/MS Ion Search software, with Chordata as taxonomic restriction, with methionine oxidation, formation of pyroGlu at the N-terminus of Gln, and deamidation (N, Q) of peptides as variable modifications. Individual ion scores >31 indicate identity or extensive homology. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits (http://www.matrixscience.com/help/interpretation_help.html).

Sample	Protein (UniProt accession number)	Minimally invasive protocol			Urea pre-treatment		
		Sequence coverage (%)	Protein score	n° of peptides	Sequence coverage (%)	Proteins core	n° of peptides
Milk and minium	Alpha-S1 casein (P02662)	68	548	11	75	828	22
	Alpha-S2 casein (P02663)	33	259	8	54	753	13
	Beta-casein (P02666)	49	298	6	79	536	11
	Kappa-casein (P02668)	22	166	5	74	319	8
	Beta-lactoglobulin (P02754)	38	308	6	36	309	7

Table 2
Sequence coverage (%) of Collagen alpha-1(I) and Collagen alpha-2(I) from the digestion of an ancient human bone with different pretreatment protocols followed by LC-MS/MS analysis. Sequence coverage was calculated on the sequence of the mature form of the protein. Proteins were identified in the UniProt database with Mascot MS/MS Ion Search software, with *Homo sapiens* as taxonomic restriction, with methionine oxidation, formation of pyroGlu at the N-terminus of Gln, deamidation (N, Q), and hydroxylation (K, P) of peptides as variable modifications.

Protein (UniProt accession number)	Minimally invasive protocol (%)	Urea (%)	EDTA (%)	HCl (%)	GuHCl (%)
Collagen alpha-1(I) (P02452)	24.0	60.0	75.0	37.5	63.4
Collagen alpha-2(I) (P08123)	22.8	59.4	72.0	27.5	52.2

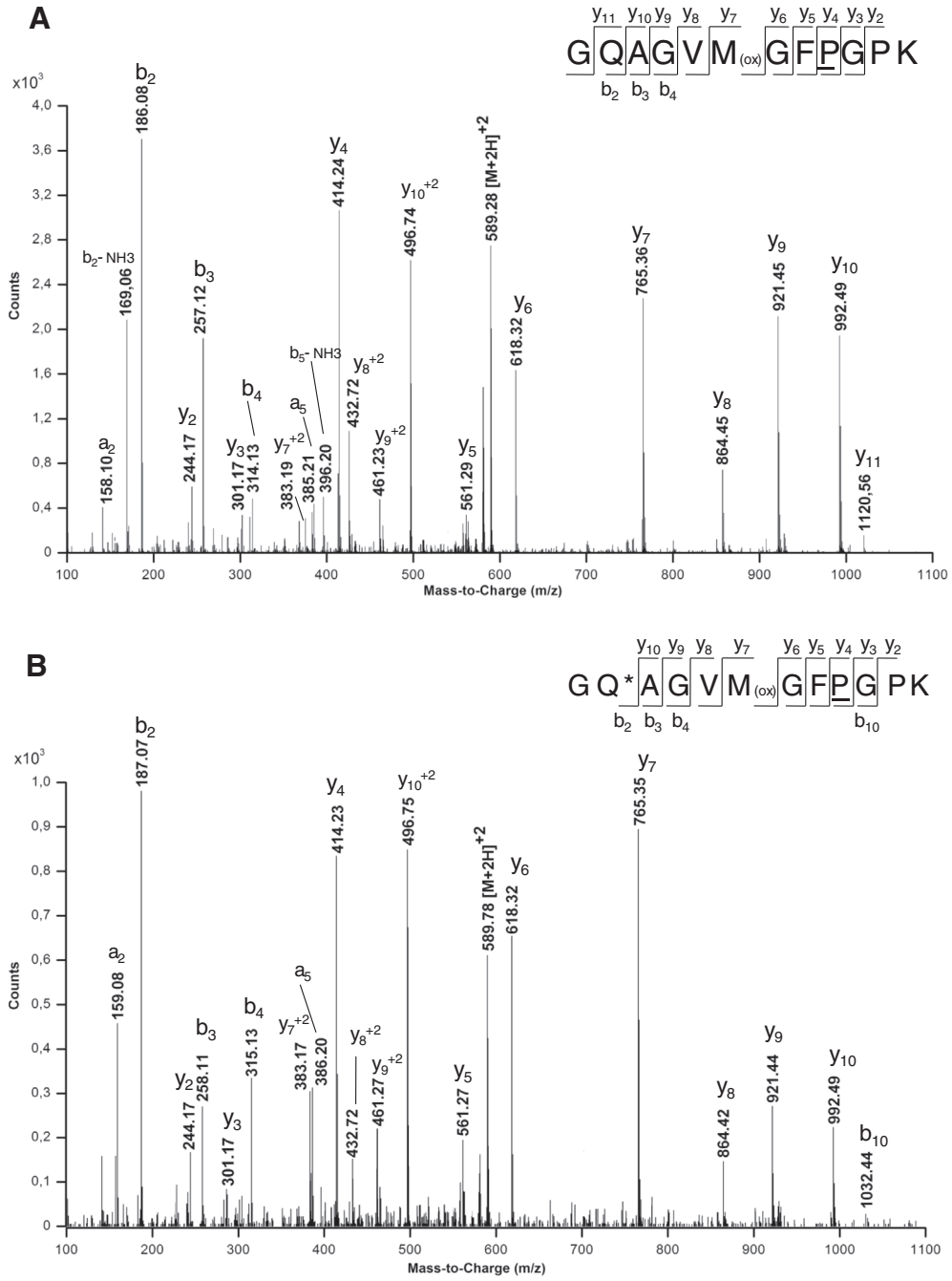


Fig. 2. MS/MS spectra of the doubly charged ions at *m*/*z* 589.28 (A) of the peptide GQAGVMGFPGPK of human collagen alpha-1(I) (P02452) and its deamidated form at *m*/*z* 589.78 (B) identified in the analysis of the sample of human bone dated 1st A.D. The product ions are indicated with the observed mass. * Indicates deamidation site, and hydroxylation site is underlined.

than collagen I because of their greater sequence variation, but they are less resistant than collagen I, constitute a relatively small fraction of the total protein content of bones, and are more soluble than collagen and might then be lost in demineralization steps. The choice of the protocol to be used for bone treatment will ultimately depend on the protein/s of interest but also on the analytical application. In fact, the extraction protocol used might also influence the results in the following characterization of the proteins recovered from bones.

Obtaining as many peptides as possible and the choice of the appropriate protocol can be also extremely important to study the molecular modifications occurring post-mortem. It is still an open question whether deamidation of glutamine (Q) and asparagine (N) could be used as a dating technique in ancient sample, but it is widely accepted that deamidation can be considered as a biomolecular marker of deterioration and natural ageing of proteins in artistic and archaeological materials [5,15,20,21]. Detection of deamidation can offer interesting prospects in the evaluation of the conservation state of work of arts and archaeological remains. However, deamidation is a delicate modification, since it is strongly influenced by several parameters such as pH and temperature. Deamidation is also a delicate modification from a purely technical point of view, since it induces a mass shift of only 0.98 Da. Fragmentation spectra, however, not only clearly rule out any doubt about the modification but also localize the deamidation site within the peptide sequence (Fig. 2) since only fragments ions containing the deamidation site will differ between deamidated and non-deamidated peptide.

Deamidation can occur also as a by-product of sample preparation. To consider deamidation as a signature of ageing, its genuine pre-extraction origin must be verified. $H_2^{18}O$ labelling can be used to assess any deamidation occurring during digestion process [5,20,22], and Fig. 3 illustrates the different mass shifts that would be observed upon hydrolysis in $H_2^{18}O$ in a glutamine containing peptide as a function of deamidation occurring before or during the treatment. However, simple rules can also be used in evaluating the significance of deamidation: conditions that could favour deamidation reaction such as extreme of pH or high temperature should be avoided, for instance, in collagen extraction, or, since glutamine deamidation is much slower than asparagines conversion to aspartic, glutamine containing peptides should be preferred over asparagines containing ones in the seek for markers of deamidation.

The use of extraction protocol as mild as possible is also important when searching for unpredicted modifications induced by ageing processes, since our knowledge of alterations occurring to proteins because of diagenetic events as well as environmental factors is not yet complete, and we need to be sure that the detected modifications are not induced during sample processing as it could for instance occur in arsh sample treatment.

3.3 Data handling

Development of bioinformatics tools that can address the specific issues, such as the identification of ageing signatures in proteins of ancient samples or to handle extinct species when genomic data are available for the extant ones, is another challenge in the field of proteomics for cultural heritage.

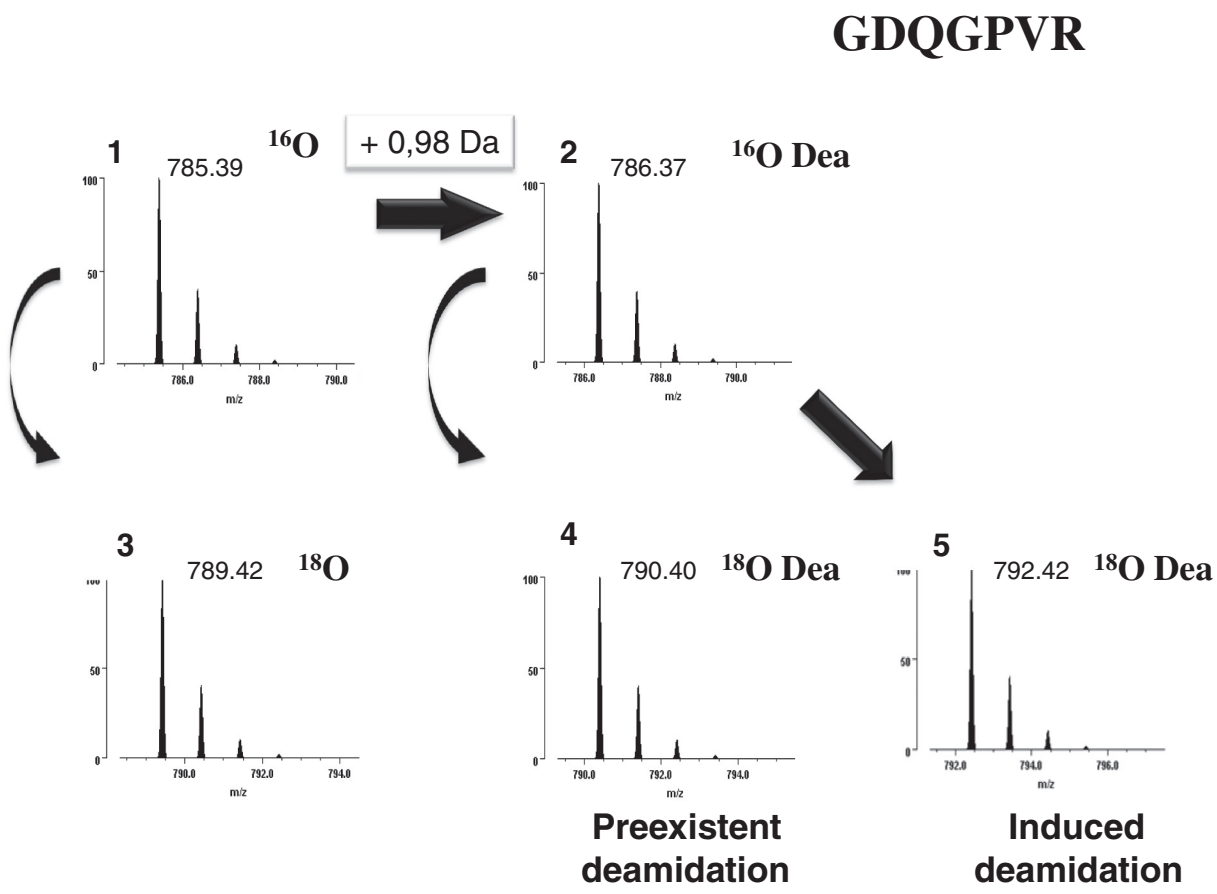


Fig. 3. Scheme of expected mass shifts upon enzymatic digestion in $H_2^{18}O$ of the peptide GDQGPVGR (MH^+ 735.39) of human collagen alpha-1(I) (P02452). Upon digestion with trypsin in $H_2^{18}O$ up to 2 ^{18}O are incorporated at the C-terminus of newly generated peptides with a mass shift of about 4 Da (3) in respect to digestion in $H_2^{16}O$ (1). Different mass shifts are observed depending on deamidation occurring either during the trypsin digestion (5) or being a preexistent modification of the peptide (4) in respect to enzymatic hydrolysis in $H_2^{16}O$ (2).

A main difficulty in identifying ancient proteins is indeed the paucity of ancient genomic data. In shotgun proteomics, proteins are routinely identified by matching experimental MSMS spectra of enzymatic digests of protein samples to simulated spectra from protein databases, usually derived from genomic sequences. Identification in database search procedures rely on how well experimental spectra fit to theoretical spectra obtained from sequences that are present in databases. While this procedure works for most of the identifying purposes such as whether egg, milk or animal glue has been used as binder in a painting [3,23] or organic component in food residues [10,24], the issue might be more complex when trying to assess the species of origin and/or in the case of extinct organisms. Standard database search algorithms fail to identify peptides that are not exactly contained in a protein database, such as those arising from unreported mutations occurred throughout evolution. Good quality fragmentation spectra are in principle discarded because of even point mutation that makes them non-ascribable to known sequences. Identification of the specific species will then depend on the detection of two or few more peptides that are conserved between the ancient protein and its modern counterpart, despite possible differences in other peptides, which will be ignored in a standard database searches. This is an intrinsic limit of the otherwise powerful proteomic approach that will mask novel sequences or points of divergence with reported sequences.

The tremendous attractive possibility of proteomics of reaching much further back in time to gain information on more ancient samples than genomics, thanks to the higher stability of proteins in respect to nucleic acids, and to characterize in depth diagenetic alterations directly on surviving molecules, can be therefore limited in principle by the lacks of corresponding DNA information.

However, many software tools have been developed for the automated identification of peptides by *de novo* sequencing directly from the MS/MS spectrum of peptides. Representative *de novo* sequencing software packages include PEAKS [25], PepNovo [26], NovoHMM [27], and Lutefisk [28], and very recently, Welker *et al.* [14] extensively and successfully used PEAKS to overcome the absence of corresponding genomic data to resolve the evolutionary history of Darwin's south American ungulates.

Moreover, the lacks of corresponding DNA information can also be partially overcome with bioinformatic tools such as the error tolerant search utility in Mascot, which allows for single substitution in peptides in respect to sequences which are present in databases, taking for granted that any obtained match has to be manually confirmed afterwards. These approaches allow taxonomic attribution of extinct species, in specimens that no longer can yield DNA [13] and can be combined to a wide panel of other bioinformatics tools such as generation *ad hoc* of peptide databases as experimented by Waters and collaborators [29] that used a collagen sequences database to analyse the mass spectrometric data obtained from a mastodon bone, and classical homology search tools like BLAST [14].

Error tolerant searches can also be used to detect diagenetic modifications [30], among which the observation of semi-tryptic peptides accounting for partial hydrolysis of the polypeptidic chain within the samples [11,15]. High occurrence of partial hydrolysis is indeed expected as degradation effect. Table 3 reports a comparison of the results obtained on several pictorial models and on a sample from the Camposanto Monumentale in Pisa when identification is carried out, on the same set of experimental data with Mascot search programme allowing for semi-tryptic cleavages and only for specific tryptic cleavages: the abundance of non-tryptic termini suggests hydrolysis as a likely and expected effect of protein degradation.

Moreover, a clever use of the "open mass" modification search utility in the Protein Prospector database search, as recently suggested by Hill *et al.* [15], in the analysis of fossil bones, allowed to identify extensive surviving galactosylation and glucosyl-galactosylation of hydroxylysine residues in collagen. This bioinformatics strategy of analysis offers the potentiality to detect unexpected modifications and certainly will hold interesting prospects in the characterization of molecular details of degradation processes.

Table 3
Proteins identified in the paint replica containing different inorganic pigments mixed with milk as binder, and in the sample from the Camposanto Monumentale of Pisa, by trypsin digestion in heterogeneous phase and LC-MS/MS analysis. Proteins were identified in the UniProt database with Mascot MS/MS Ion Search software, with Chordata as taxonomic restriction, with methionine oxidation, formation of pyroGlu at the N-terminus of Gln, and deamidation (N, Q) of peptides as variable modifications. Individual ion scores >31 indicate identity or extensive homology. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits (http://www.matrixscience.com/help/interpretation_help.html).

Sample	Protein (UniProt Accession number)	Tryptic identification			Semitryptic identification		
		Sequence coverage (%)	Protein score	n° of peptides	Sequence Coverage (%)	Protein score	n° of peptides (specific tryptic peptides)
Casein and minium	Alpha-S1 casein (P02662)	76	641	16	77	1496	27 (16)
	Alpha-S2 casein (P02663)	56	512	17	56	603	21 (17)
	Beta-casein (P02666)	53	426	9	69	1104	14 (9)
	Kappa-casein (P02668)	24	242	5	50	520	8 (5)
Casein and cinnabar	Alpha-S1 casein (P02662)	69	666	15	69	1718	31 (15)
	Alpha-S2 casein (P02663)	52	530	12	53	792	14 (12)
	Beta-casein (P02666)	79	465	12	79	1070	21 (12)
	Kappa-casein (P02668)	46	345	6	53	728	8 (6)
Casein and $CaCO_3$	Alpha-S1 casein (P02662)	64	483	12	64	1072	21 (12)
	Alpha-S2 casein (P02663)	52	356	6	52	453	13 (6)
	Beta-casein (P02666)	47	266	7	61	732	21 (7)
	Kappa-casein (P02668)	26	214	3	47	363	6 (3)
Casein	Alpha-S1 casein (P02662)	71	589	14	72	1339	23 (14)
	Alpha-S2 casein (P02663)	53	512	14	53	560	18 (14)
	Beta-casein (P02666)	79	445	12	79	1058	19 (12)
	Kappa-casein (P02668)	44	272	7	53	567	9 (7)
Sample from Camposanto Monumentale	Alpha-S1 casein (P02662)	50	565	11	56	1377	31 (11)
	Alpha-S2 casein (P02663)	27	252	7	29	374	10 (7)
	Beta-casein (P02666)	71	432	8	72	1601	32 (8)
	Kappa-casein (P02668)	22	141	3	46	391	10 (3)
	Collagen alpha-1(I) chain (P02453)	36	854	25	42	879	28 (25)
	Collagen alpha-2(I) chain (P02465)	23	522	21	25	558	22 (21)

The modifications that occur on proteins in natural environment over time are still under investigation and delineating both *in vivo* and diagenetically derived alterations will provide important information on the physiology and/or phylogenies of organisms, as well as on the ageing mechanisms. These can be function of specific environmental factors, and their knowledge will lead to a more conscious preservation of ancient samples.

Moreover, the intrinsic damaged nature of the ancient proteins can intuitively be expected to be the first evidence of authenticity of a sample, that ruling out possible contaminations from modern materials [31].

While much effort has been already devoted to the development of identification tools, a lot of work has still to be done to understand and characterize the whole range of modifications occurring upon ageing on deteriorating proteins in samples that have been exposed to a wide spectrum of different environmental conditions, thus contributing to what can be called the field of paleoproteomics.

4. Conclusions

In 2010 *Science* magazine [32] ranked molecular palaeontology as one of the top 10 scientific fields that made a break in the first decade of the new millennium [33]. Ongoing developments in the analysis of ancient proteins, in terms of adaptation of well established protocols to the peculiar requirements of studying ancient life, are rapidly putting paleoproteomics in a worthy position in the wide panorama of studies that include molecular evolution, archaeology, palaeontology and any kind of cultural heritage we may think of.

It can be expected that in the very next future ancient proteins and paleoproteomics will fully recover the lag in time behind ancient DNA, contributing to shed light on ancient life and masterpieces from a scientific perspective that nicely integrate humanistic points of view.

Acknowledgements

The authors are grateful to Distretto ad Alta Tecnologia dei Beni Culturali (DATABENC scarl) (PON03PE_00163_1) for financial support.

References

[1] S.T. en Have, Y. Ahmad, A.I. Lamond, Proteomics — current novelties and future directions, J. Anal. Bioanal. Tech. 01 (2011) 1–6, http://dx.doi.org/10.4172/2155-9872.S3-001.
[2] P.H. Ostrom, M. Schall, H. Gandhi, T.-L. Shen, P.V. Hauschka, J.R. Strahler, et al., New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry, Geochim. Cosmochim. Acta 64 (2000) 1043–1050, http://dx.doi.org/10.1016/S0016-7037(99)00381-6.
[3] 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. (2009) 2269–2280.
[4] P. Villa, L. Pollarolo, I. Degano, L. Birolo, M. Pasero, C. Biagioni, et al., A milk and ochre paint mixture used 49,000 years ago at sibudu, South Africa, PLoS One 10 (2015), e0131273http://dx.doi.org/10.1371/journal.pone.0131273.
[5] G. Leo, I. Bonaduce, A. Andreotti, G. Marino, P. Pucci, M.P. Colombini, et al., Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings, Anal. Chem. 83 (2011) 2056–2064, http://dx.doi.org/10.1021/ac1027275.
[6] G. Ranalli, G. Alfano, C. Belli, G. Lustrato, M.P. Colombini, I. Bonaduce, et al., Biotechnology applied to cultural heritage: biorestoration of frescoes using viable bacterial cells and enzymes, J. Appl. Microbiol. 98 (2005) 73–83, http://dx.doi.org/10.1111/j.1365-2672.2004.02429.x.
[7] M. Gambino, F. Cappitelli, C. Cattò, A. Carpen, P. Principi, L. Ghezzi, et al., A simple and reliable methodology to detect egg white in art samples, J. Biosci. 38 (2013) 397–408, http://dx.doi.org/10.1007/s12038-013-9321-z.
[8] K.L. Rasmussen, A.L. Tenorio, I. Bonaduce, M.P. Colombini, L. Birolo, E. Galano, et al., Reply to Ira Rabin's comment on our paper Rasmussen etal. J. Archaeol. Sci. 43 (2014) (2012) 155–158, http://dx.doi.org/10.1016/j.jas.2013.11.030.
[9] R. Vinciguerra, E. Galano, F. Vallone, G. Greco, A. Vergara, I. Bonaduce, et al., Deglycosylation step to improve the identification of egg proteins in art samples, Anal. Chem. (2015) 150930065206005 doi:10.1021/acs.analchem.5b02423.
[10] C. Hong, H. Jiang, E. Lü, Y. Wu, L. Guo, Y. Xie, et al., Identification of milk component in ancient food residue by proteomics, PLoS One 7 (2012) 1–7, http://dx.doi.org/10.1371/journal.pone.0037053.
[11] C. Warinner, J. Hendy, C. Speller, E. Cappellini, R. Fischer, C. Trachsel, et al., Direct evidence of milk consumption from ancient human dental calculus, Sci. Rep. 4 (2014) 7104, http://dx.doi.org/10.1038/srep07104.
[12] N. Rybczynski, J.C. Gosse, C.R. 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) 1550, http://dx.doi.org/10.1038/ncomms2516.
[13] M. Buckley, A molecular phylogeny of plesiorycteropus reassigns the extinct mammalian order "*Bibymalagasia*,", PLoS One 8 (2013)http://dx.doi.org/10.1371/journal.pone.0059614.
[14] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, et al., Ancient proteins resolve the evolutionary history of Darwin/'s south American ungulates, Nature 522 (2015) 81–84, http://dx.doi.org/10.1038/nature14249.
[15] R.C. Hill, M.J. Wither, T. Nemkov, A. Barrett, A. D'Alessandro, M. Dzieciatkowska, 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 14 (2015) 1946–1958, http://dx.doi.org/10.1074/mcp.M114.047787.
[16] J. Marchant, Mystery of Darwin's "strange animals" solved, Nature 1–3 (2015)http://dx.doi.org/10.1038/nature.2015.17138.
[17] 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, http://dx.doi.org/10.1002/rcm.6821.
[18] X. Jiang, M. Ye, X. Jiang, G. Liu, S. Feng, L. Cui, et al., Method development of efficient protein extraction in bone tissue for proteome analysis, J. Proteome Res. 6 (2007) 2287–2294, http://dx.doi.org/10.1021/pr070056t.
[19] T.P. Cleland, K. Voegele, M.H. Schweitzer, Empirical evaluation of bone extraction protocols, PLoS One 7 (2012) 1–9, http://dx.doi.org/10.1371/journal.pone.0031443.
[20] 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, http://dx.doi.org/10.1021/ac301333t.
[21] C. Solazzo, J. Wilson, J.M. Dyer, S. Clerens, E. Plowman, I. Von Holstein, Modeling Deamidation in Sheep α-Keratin Peptides and Application to Archeological Wool Textiles, Anal. Chem. 86 (2014) 567–575, http://dx.doi.org/10.1021/ac4026362.
[22] 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, http://dx.doi.org/10.1002/rcm.6351.
[23] C. Tokarski, E. Martin, C. Rolando, C. Cren-Olivé, Identification of proteins in renaissance paintings by proteomics, Anal. Chem. 78 (2006) 1494–1502, http://dx.doi.org/10.1021/ac051181w.
[24] M. Buckley, N.D. Melton, J. Montgomery, Proteomics analysis of ancient food vessel stitching reveals >4000-year-old milk protein, Rapid Commun. Mass Spectrom. 27 (2013) 531–538, http://dx.doi.org/10.1002/rcm.6481.
[25] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, et al., PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry, Rapid Commun. Mass Spectrom. 17 (2003) 2337–2342, http://dx.doi.org/10.1002/rcm.1196.
[26] A. Frank, P. Pevzner, PepNovo: de novo peptide sequencing via probabilistic network modeling, Anal. Chem. 77 (2005) 964–973, http://dx.doi.org/10.1021/ac048788h.
[27] B. Fischer, V. Roth, J. Buhmann, J. Grossmann, S. Baginsky, W. Gruissem, et al., A hidden Markov model for de novo peptide sequencing, Anal. Chem. 77 (2005) 7265–7273.
[28] J.A. Taylor, R.S. Johnson, Sequence database searches via de novo peptide sequencing by tandem mass spectrometry, Rapid Commun. Mass Spectrom. 11 (1997) 1067–1075, http://dx.doi.org/10.1002/(SICI)1097-0231(19970615)11:9<1067::AID-RCM953>3.0.CO;2-L.
[29] M.R. Waters, T.W. Stafford, H.G. McDonald, C. Gustafson, M. Rasmussen, E. Cappellini, et al., Pre-clovis mastodon hunting 13,800 years ago at the Manis site, Washington, Science 334 (2011) 351–353, http://dx.doi.org/10.1126/science.1207663.
[30] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.a.R. Da Fonseca, T.W. Stafford, et al., Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, J. Proteome Res. 11 (2012) 917–926, http://dx.doi.org/10.1021/pr200721u.
[31] M. Buckley, A. Walker, S.Y.W. Ho, Y. Yang, C. Smith, P. Ashton, et al., Comment on "protein sequences from mastodon and tyrannosaurus rex revealed by mass spectrometry,", Science 319 (2008) 33 (80-.). doi:10.1126/science.1147046.
[32] T.N. Staff, Stepping away from the trees for a look at the forest, Science 330 (2010) 1612–1613, http://dx.doi.org/10.1126/science.330.6011.1612.
[33] A. Gibbons, Tiny time machines revisit ancient life, Science 330 (2010) 1616, http://dx.doi.org/10.1126/science.330.6011.1616.

analytical chemistry

Technical Note

pubs.acs.org/ac

Deglycosylation Step to Improve the Identification of Egg Proteins in Art Samples

Roberto Vinciguerra,[†] Eugenio Galano,[†] Fabiana Vallone,[†] Giovanna Greco,[‡] Alessandro Vergara,[†,∥] Ilaria Bonaduce,[§] Gennaro Marino,[†] Pietro Pucci,[†,∥] Angela Amoresano,[†] and Leila Birolo*[,†,∥]

[†]Dipartimento di Scienze Chimiche, Università degli Studi di Napoli Federico II, 80126 Napoli, Italy

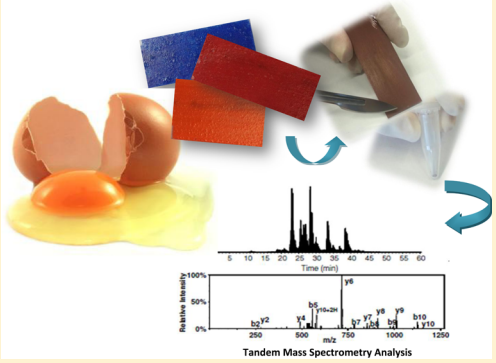
[‡]Dipartimento di Studi Umanistici, Università degli Studi di Napoli Federico II, 80133 Napoli, Italy

[§]Dipartimento di Chimica e Chimica Industriale, Università di Pisa, 56124 Pisa, Italy

[∥]Distretto ad Alta Tecnologia dei Beni Culturali (DATABENC) Scarl, Napoli, Italy

Supporting Information

ABSTRACT: A deglycosylation step using Peptide-N-Glycosidase F (PNGaseF) has been introduced in a standard proteomic protocol to more confidently identify egg based binders. The ingenuity of introducing a PNGaseF digestion was aimed at removing the molecular hindrance, made up by the heavily glycosylated egg proteins, before the protease(s) hydrolysis. This novelty in the protocol resulted in obtaining a significant increase of proteolytic egg peptides thus improving the quality and reliability of egg identification in artwork samples. The protocol has been set up on paint replicas and successfully tested on two historical samples of different origin.



Proteinaceous materials, such as animal glue, egg (both yolk and albumen), and milk have long been used in paintings as binders, coatings, and adhesives. Nowadays the identification of the materials used by artists is of greatest significance in revealing working practices, defining conservation protocols, and occasionally for authenticating and dating the artworks.[1]

Following the seminal papers by Hynek et al.[2] and by Tokarsky et al.,[3] several proteomic strategies, aimed at the unequivocal identification of proteins present in art objects or in archeological remains, have been designed by extensively exploiting modern mass spectrometry. The "bottom-up" approach that has been generally used so far is based on the direct enzymatic digestion of the sample followed by either MALDI-TOF or LC−MS/MS mass spectrometric analysis of peptides released from the object under study.[4−12] The very critical step in this approach is related to the efficiency of either protein extraction from solid matrixes or proteolytic digestion of substrates incorporated within the matrix itself. In our experience[4] the detection of egg proteins in paint samples remains often rather unsatisfactory and is certainly less confident than other most common binders, i.e., milk and animal glue.

It is well-known that the most abundant proteins in hen egg, either in the albumen (i.e., ovalbumin, ovotransferrin, ovomucoid, ovoglobulins, and beta-ovomucin) and in the yolk (i.e., vitellogenin) are heavily glycosylated.[13,14] We reasoned that the extensive glycosylation of egg proteins might create a significant molecular hindrance which hampers proteases to efficiently interact with the proteinaceous substrates, thus greatly decreasing proteolytic efficiency. This, in turn, would result in inefficient and poor production of peptides, impairing a confident protein identification. A possible approach to avoid this molecular hindrance would be to trim out the glycosidic decoration by preceding the proteolytic digestion with a deglycosylating step.[15] This working hypothesis was investigated and results are herein reported. The study demonstrated that the ingenuity of introducing a N-Glycosidase F digestion before the protease(s) treatment greatly increases the release of digested peptides, substantially improving the quality and reliability of egg identification in samples from artworks.

EXPERIMENTAL SECTION

Reagents. Ammonium hydrogen carbonate (AMBIC), ethylenediaminetetraacetic acid (EDTA); tri(hydroxymethyl)-aminomethane (TRIS), TPCK-treated trypsin were from Sigma; recombinant Peptide N-Glycosidase F (PNGaseF) was from Roche. Formic acid and acetonitrile (ACN) were purchased from Baker, respectively. Deionized water was obtained from the Millipore cartridge equipment.

Painting Samples. Paint reconstructions were prepared using egg white, milk, and animal glue as binders and azurite

Received: June 29, 2015
Accepted: September 23, 2015
Published: September 23, 2015

DOI: 10.1021/acs.analchem.5b02423
Anal. Chem. 2015, 87, 10178−10182

($Cu_3(CO_3)_2(OH)_2$), red ochre (Fe_2O_3), minium (Pb_3O_4) and vermilion (HgS) and were applied on glass slides. Paint replicas were prepared in 2010, left to dry at room temperature for 1 month, then artificially aged for 720 h at 25 °C, 50% relative humidity in indoor light conditions, and stored since then in the darkness at room temperature. Details are reported elsewhere.[16,17] Sample DSFL6 was collected from the gilded aureole of the angel in "Holey Conversation", mural painting by Amico Aspertini, 1506−1510, San Frediano Church, Lucca, Italy. Sample Purple 12M080 was collected from a mural painting of the urban district of Cuma archeological site (first century AD)

Protein Deglycosylation. Treatment with PNGaseF was carried out by adding to microsamples (~300−800 μg) 50 μL of AMBIC 50 mM containing 60m U/μL of PNGaseF solution and incubating at 37 °C for 2 h. The reaction was stopped by incubation of the sample in boiling water for 2 min.

Protein Digestion and LC−MS/MS Analysis. Protein samples were enzymatically digested on the basis of the minimally invasive proteomic analytical procedure described by Leo et al.[4] Briefly, trypsin was added to a final concentration of 10 ng/μL both to the samples from PNGaseF pretreatment and, in the case of trypsin alone protocol, to microsamples (~300−800 μg) directly suspended in 50 μL of AMBIC 50 mM. After incubation at 37 °C for 16 h, the supernatants were recovered by centrifugation, and the peptide mixture was filtered on 0.22 μm PVDF membrane (Millipore), concentrated, and purified using a reverse-phase C18 Zip Tip pipet 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 analyzed 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) 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 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. Each LC−MS/MS analysis was preceded and followed by blank runs to avoid carryover contamination. Double and triple charged ions were preferably isolated and fragmented. 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 the taxonomy restriction for protein identification. A licensed version of MASCOT software (www.matrixscience.com) version 2.4.0. was used with trypsin as the 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 residues, formation of pyroglutamic acid from glutamine residues at the N-terminal position of peptides, and deamidation at asparagines and glutamines were considered as variable modifications.[6] 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.[4] 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 generally 33. Spectra with a MASCOT score of <10, having low quality were rejected.

RESULTS AND DISCUSSION

The experimental observations that egg proteins are seldom identified in paint samples and, even when analyzing model samples, their identification is based on a small number of peptides, thus impairing a highly confident identification, prompted us to develop a novel analytical protocol to investigate and to circumvent this problem. A deglycosylation step was introduced before trypsin hydrolysis in the hypothesis that the high content of glycosylated proteins of egg might constitute a molecular barrier hampering proteolytic digestion. Several enzymes have been successfully applied to the release of N-linked glycans, such as peptide-N-glycosidase F (PNGaseF), endoglycosidase F and H.[19,20] Among these, PNGase F has emerged as a widely used glycoamidase.[21,22] As already shown in other circumstances,[23] this enzyme is able to remove the glycosidic moieties leaving the protein substrate more amenable to tryptic digestion. A simple protocol was thus set up with a deglycosylating step before the tryptic digestion as described in the Experimental Section.

Paint reconstructions containing red ochre as pigment and alternatively albumen or yolk or whole egg were used. Aliquots of each sample were analyzed using both the classical proteomic approach, based on simple treatment with trypsin (herein indicated as "Trypsin alone") and the protocol with PNGaseF pretreatment (herein indicated as "PNGaseF + Trypsin"). It is worth noting that both steps of PNGaseF and trypsin digestions were carried out in the heterogeneous phase, that means direct enzymatic digestion on solid samples without protein extraction.[4] This can be extremely useful when dealing with paint samples in which proteins are aged and bonded with pigments and fillers[16,17] and are thus difficult to solubilize. After digestion with trypsin, samples were analyzed by LC−MS/MS and proteins were identified by database search with the MS/MS ion search mode within a licensed version of MASCOT, with Chordata as the taxonomic restriction in the SwissProt protein database.

Table 1 reports a comparison of the results obtained on a replica made of albumen and red ochre. It is evident that pretreating the sample with PNGaseF results in a much higher number of peptides obtained, greatly improving the reliability of the identification. In the aliquot treated with trypsin alone, only ovalbumin and ovotransferrin were identified, with a few detected peptides.

PNGaseF pretreatment led to a definitively more confident and reliable identification of eight proteins from chicken egg white, showing also a much larger number of detected peptides for ovalbumin and ovotransferrin. Details on protein identifications are reported in the Supporting Information, Table S1, where the N-glycosylation signature of deamidation in the Asn-X-Ser consensus sequence can be easily spotted in some of the observed peptides demonstrating the efficacy of the

DOI: 10.1021/acs.analchem.5b02423
Anal. Chem. 2015, 87, 10178−10182

Table 1. Proteins Identified in the Paint Replica Containing Red Ochre and Albumen by LC−MS/MS[a]

identified protein. (accession number)	protocol					
	Trypsin		PNGaseF + Trypsin		Trypsin + PNGaseF	
	score[b]	no. of peptides	score[b]	no. of peptides	score[b]	no. of peptides
Ovalbumin (P01012)	159	4	960	23	299	11
Ovotransferrin (P02789)	90	2	2147	63	196	7
Ovostatin (P20740)			137	5		
Ovalbumin releted protein Y (P01014)			452	11	85	3
Mucin 5B (Q98UI9)			161	2		
Lysozyme C (P00698)			340	7		
Ovalbumin related protein X (P01013)			82	3		
Protein Tenp (O42273)			82	2		

[a]Aliquots of paint replica were treated in heterogeneous phase with trypsin with or without treatment with PNGaseF (before and after tryptic digestion) and analysed by LC−MS/MS. Details of the identification are given in Table S-1 [b]Protein scores are derived from ions scores as a nonprobabilistic basis for ranking protein hits (http://www.matrixscience.com/help/interpretation_help.html). Ions score is −10 Log(P), where P is the probability that the observed peptide match is a random event.

deglycosylation procedure. As control, an aliquot of sample was digested with trypsin and then with PNGaseF, in order to check whether the complexity of the sample mixture rather than the steric hindrance created by glycosylated peptides was responsible for the lack of results in PNGaseF untreated samples.

Very similar results were obtained on paint replicas containing either whole egg or egg yolk as binder (Table S-2). In the absence of PNGaseF pretreatment, 7 proteins were identified in the whole egg samples while following PNGaseF incubation a total of 14 proteins could be confidently identified, mainly ovalbumin and related protein and vitellogenins ones. Similar results were obtained in the analysis of the egg yolk paint replica, where mainly vitellogenins were identified, demonstrating that the usefulness of PNGaseF treatment is not limited to sample containing proteins from albumen only, but it is extended to any sample where egg proteins are present (Table S-2). Moreover, in all cases a higher number of peptides were detected compared to the trypsin alone protocol (from 2 to 19 times more).

In order to investigate if the PNGaseF treatment affects the identification of milk and animal glue, the other two common proteinaceous paint binders we analyzed with the newly developed protocol paint replicas containing the same pigment, red ochre, mixed with milk and animal glue, as well as different combinations of binders. Table S-3 shows that PNGaseF treatment had no substantial effects on the results when the sample does not contain egg. This result is not surprising, since milk proteins and collagen, the most abundant protein in animal glue, are much less glycosylated than egg proteins and are thus good substrates for tryptic hydrolysis even in the absence of the deglycosylation step. On the contrary, whenever egg is present in mixture with other binders, PNGaseF pretreatment improved identifications.

As in paint samples, inorganic materials are simultaneously present, mostly as pigments and dryers, and it is necessary to investigate whether they can inhibit the activity of PNGaseF. To this aim, paint replicas containing a range of metals commonly occurring in paintings were selected. In particular the replicas contained albumen and cinnabar (HgS), minium (2 $PbO{\cdot}PbO_2$), and azurite ($Cu_3(CO_3)_2(OH)_2$), alongside the previously tested red ochre (Fe_2O_3). Experiments were carried out as above on aliquots of each sample using the trypsin alone and the PNGaseF + trypsin protocols in parallel. Data clearly indicate that PNGaseF treatment always improved the egg identifications, regardless of the pigment used (Table S-4), by increasing the number of peptides that are released from the proteins after tryptic digestion. These results indicate that PNGaseF is effective also in the presence of some widespread metal containing pigments, including copper. Moreover the data clearly show that the newly developed protocol is fundamental in order to maximize the possibility of identifying egg in an aged paint sample. It is well-known, in fact, that pigments can affect the protein identification by amino acid analysis if suitable purification steps are not adopted.[24,25] To fully understand the influence of pigments in the protein identification through proteomics procedures, the effect of pigments still needs be systematically investigated. Despite this, the data presented here clearly indicate that pigments do have an influence in the number of proteins that can be identified by proteomics analysis in a paint sample, as a consequence of the strong interactions taking place between pigments and proteins,[17,18] and the developed protocol is a suitable analytical tool to help us to improve our success rate.

Historical Paint Samples. Finally, our improved protocol was tested on two samples collected from historical objects: sample Purple 12M080 from a mural painting from the urban district of Cuma archeological site and DSFL6 from a mural painting of San Frediano Church in Lucca, Italy (16th century). It is worth noting that sample DSFL6 was previously analyzed by GC/MS and the presence of egg was accordingly inferred.[26] Both samples provided no result when analyzed by the Trypsin alone protocol.

Aliquots of the two historical samples were analyzed following the newly developed procedure including PNGaseF pretreatment before trypsin digestion and LC−MS/MS analysis of the resulting peptide mixtures.

Table 2 reports the identifications of the proteins obtained in the two samples with details of the identifications reported in Table S-5. A number of egg proteins were identified in both samples confirming the presence of an egg-containing binder and showing very clearly that deglycosylation of the samples before tryptic digestion led to the identification of egg proteins that could not be detected before. Moreover, it should be underlined that no proteins from albumen were identified in the sample from Cuma thus allowing us to confidently assess that only yolk was used as paint binder.

■ CONCLUSIONS

Diagnostic methods represent a crucial aspect of the scientific investigation of artworks and proteomics procedures are increasingly applied for the identification of protein binders in samples from archeological objects and works of art. Proteomics strategies applied to artworks has begun to be

DOI: 10.1021/acs.analchem.5b02423
Anal. Chem. 2015, 87, 10178−10182

Table 2. Proteins Identified in Historical Paint Samples by LC−MS/MS[a]

sample	identified protein (accession number)	PNGaseF + Trypsin	
		score	no. of peptides
Purple 12M080	Vitellogenin-2 (P02845)	1163	35
	Vitellogenin-1 (P87498)	626	20
	Apolipoprotein B (P11682)	107	4
	Apovitellenin-1 (P02659)	90	2
DSFL6	Ovalbumin (P01012)	20	2

[a]The mural painting fragment "Purple 12M080" from the urban district of Cuma archeological site and a fragment of the mural painting "Sacra Conversazione" (DSFL6) by Amico Aspertini 1506−1510 from the S. Frediano Basilica (Lucca) were subjected to trypsin digestion after PNGaseF pretreatment and the resulting peptide mixtures analyzed by LC−MS/MS. Details of the identification are given in Table S-5

accepted as the gold standard analytical technique when proteins have to be identified and characterized because of the sensitivity, the capability of identifying species-specific proteins, and detecting degradation processes. However, methodological adaptations to specific analytical problems of samples from cultural heritage still need to be developed to tailor appropriate approaches. The physical state of the samples, enormously different from the natural environment of proteins, and the degradation processes undergone during aging provide unusual problems that require to be counteracted by specific adaptations of the classically adopted protocols used in the analysis of biological samples. A typical example of these situations is reported in this paper. Egg proteins can be easily identified when occurring in solution with glycosylation not hampering the procedure. However, when egg is used as binder in paintings, in a solid state mixed with pigments, the oligosaccharide moieties seem to create a molecular hindrance that prevent the accessibility of proteases, greatly impairing the identification process. The implementation of a deglycosylation step in the analysis of paint samples prior to the tryptic digestion has proven to significantly improve the number of identified peptides from egg proteins in several different paint reconstructions as well as on two historical samples of completely different origin. Moreover it was shown that the same glycosylation step does not affect the capability of correctly identifying other proteinaceous paint binders, such as milk and animal glue.

The protocol to identify egg based binders described in this work, based on a sample pretreatment with PNGaseF, has revealed to be useful, reliable, cost-effective, and sensitive enough to cope with the small amounts of degraded proteins that can be found in samples from artworks.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b02423.

Additional tables reporting details of protein identifications in paint replicas and historical samples of Table 1 and 2 and the proteins identificated in several combinations of pigment and protein binders (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: birolo@unina.it.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors wish to thank Sandro Baroni for kindly providing the samples, Dr. Lisa Ghezzi for preparing the paint reconstructions, Dr. Anna Lluveras-Tenorio for helpful discussion and critical reading of the manuscript, Dr. Paolo Caputo for kindly providing the samples of Cuma, and Dr. Antonella Tomeo for discussion of the mural painting of Cuma. The authors acknowledge the Ministero Italiano dell'Istruzione, dell'Università e della Ricerca for the financial support (Grant "PRIN 2010-2011": Sustainability in cultural heritage: from diagnosis to the development of innovative systems for consolidation, cleaning and protection" and Grant "PRIN 2008: Metodiche di proteomica e immunochimica per lo studio della componente proteica in patine ad ossalato e manufatti pittorici), and the Distretto ad Alta Tecnologia dei Beni Culturali (DATABENC scarl) for financial support.

■ REFERENCES

(1) Colombini, M. P.; Andreotti, A.; Bonaduce, I.; Modugno, F.; Ribechini, E. *Acc. Chem. Res.* **2010**, *43*, 715−727.
(2) Hynek, R.; Kuckova, S.; Hradilova, J.; Kodicek, M. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1896−1900.
(3) Tokarski, C.; Martin, E.; Rolando, C.; Cren-Olive, C. *Anal. Chem.* **2006**, *78*, 1494−1502.
(4) Leo, G.; Cartechini, L.; Pucci, P.; Sgamellotti, A.; Marino, G.; Birolo, L. *Anal. Bioanal. Chem.* **2009**, *395*, 2269−2280.
(5) Chambery, A.; Di Maro, A.; Sanges, C.; Severino, V.; Tarantino, M.; Lamberti, A.; Parente, A.; Arcari, P. *Anal. Bioanal. Chem.* **2009**, *395*, 2281−2291.
(6) Leo, G.; Bonaduce, I.; Andreotti, A.; Marino, G.; Pucci, P.; Colombini, M. P.; Birolo, L. *Anal. Chem.* **2011**, *83*, 2056−2064.
(7) Dallongeville, S.; Koperska, M.; Garnier, N.; Reille-Taillefert, G.; Rolando, C.; Tokarski, C. *Anal. Chem.* **2011**, *83*, 9431−9437.
(8) Dallongeville, S.; Richter, M.; Schäfer, S.; Kühlenthal, M.; Garnier, N.; Rolando, C.; Tokarski, C. *Analyst* **2013**, *138*, 5357−5364.
(9) Fremout, W.; Dhaenens, M.; Saverwyns, S.; Sanyova, J.; Vandenabeele, P.; Deforce, D.; Moens, L. *Anal. Chim. Acta* **2010**, *658*, 156−162.
(10) Fremout, W.; Kuckova, S.; Crhova, M.; Sanyova, J.; Saverwyns, S.; Hynek, R.; Kodicek, M.; Vandenabeele, P.; Moens, L. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1631−1640.
(11) Kuckova, S.; Anca Sandu, I. C.; Crhova, M.; Hynek, R.; Fogas, I.; Muralha, V. S.; Sandu, A. V. *Microchem. J.* **2013**, *110*, 538−544.
(12) Calvano, C. D.; van der Werf, I. D.; Palmisano, F.; Sabbatini, L. *Anal. Bioanal. Chem.* **2011**, *400*, 2229−2240.
(13) Guérin-Dubiard, C.; Pasco, M.; Mollé, D.; Désert, C.; Croguennec, T.; Nau, F. *J. Agric. Food Chem.* **2006**, *54*, 3901−3910.
(14) Jiang, K.; Wang, C.; Sun, Y.; Liu, Y.; Zhang, Y.; Huang, L.; Wang, Z. *J. Agric. Food Chem.* **2014**, *62*, 7245−7254.
(15) Bailey, U. M.; Schulz, B. L. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2013**, *923−924*, 16−21.
(16) Amoresano, A.; Minchiotti, L.; Cosulich, M. E.; Campagnoli, M.; Pucci, P.; Andolfo, A.; Gianazza, E.; Galliano, M. *Eur. J. Biochem.* **2000**, *267*, 2105−2112.

DOI: 10.1021/acs.analchem.5b02423
Anal. Chem. 2015, 87, 10178−10182

(17) Duce, C.; Ghezzi, L.; Onor, M.; Bonaduce, I.; Colombini, M. P.; Tine', M. R.; Bramanti, E. *Anal. Bioanal. Chem.* **2012**, *402*, 2183−2193.
(18) Duce, C.; Bramanti, E.; Ghezzi, L.; Bernazzani, L.; Bonaduce, I.; Colombini, M. P.; Spepi, A.; Biagi, S.; Tiné, M. R. *Dalton Transactions* **2013**, *42*, 5975−5984.
(19) O'Neill, R. A. *J. Chromatogr. A* **1996**, *720*, 201−215.
(20) Zhang, W.; Wang, H.; Zhang, L.; Yao, J.; Yang, P. *Talanta* **2011**, *85*, 499−505.
(21) Tarentino, A. L.; Gómez, C. M.; Plummer, T. H., Jr *Biochemistry* **1985**, *24*, 4665−4671.
(22) Chu, F. K. *J. Biol. Chem.* **1986**, *261*, 172−177.
(23) 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*, 487−496.
(24) Bonaduce, I.; Cito, M.; Colombini, M. P. *J. Chromatogr. A* **2009**, *1216*, 5931−5940.
(25) De la Cruz-Cañizares, J.; Doménech-Carbò, M. T.; Gimeno-Adelantado, J. V.; Mateo-Castro, R.; Bosh-Reig, F. *J. Chromatogr. A* **2004**, *1025*, 277−285.
(26) Bonaduce, I.; Colombini, M. P.; Diring, S. *J. Chromatogr. A* **2006**, *1107*, 226−232.

DOI: 10.1021/acs.analchem.5b02423
Anal. Chem. 2015, 87, 10178−10182

PLOS ONE

RESEARCH ARTICLE

A Milk and Ochre Paint Mixture Used 49,000 Years Ago at Sibudu, South Africa

Paola Villa[1,2,3]*, Luca Pollarolo[2,4], Ilaria Degano[5], Leila Birolo[6], Marco Pasero[7], Cristian Biagioni[7], Katerina Douka[8], Roberto Vinciguerra[6], Jeannette J. Lucejko[5], Lyn Wadley[9]

1 University of Colorado Museum, Boulder, Colorado, United States of America, **2** School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa, **3** Istituto Italiano di Paleontologia Umana, Rome, Italy, **4** Laboratoire Archéologie et Peuplement de l'Afrique, Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland, **5** Dipartimento di Chimica, Università di Pisa, Pisa, Italy, **6** Dipartimento di Scienze Chimiche, Complesso Universitario Monte S. Angelo, Napoli, Italy, **7** Dipartimento di Scienze della Terra, Università di Pisa, Pisa, Italy, **8** Research Lab for Archaeology and the History of Art, University of Oxford, Oxford, OX1 3QY, England, **9** Evolutionary Studies Institute, Centre of Excellence, University of the Witwatersrand, Johannesburg, South Africa

* villap@colorado.edu

Citation: Villa P, Pollarolo L, Degano I, Birolo L, Pasero M, Biagioni C, et al. (2015) A Milk and Ochre Paint Mixture Used 49,000 Years Ago at Sibudu, South Africa. PLoS ONE 10(6): e0131273. doi:10.1371/journal.pone.0131273

Editor: Nuno Bicho, Universidade do Algarve, PORTUGAL

Received: February 20, 2015

Accepted: June 1, 2015

Published: June 30, 2015

Data Availability Statement: The Sibudu collections are housed in the Acacia unit of the Evolutionary Studies Institute at the University of the Witwatersrand and are catalogued 2931CA. All research data generated by this paper are within the paper and its Supporting Information files.

Funding: Research by PV and LP was funded by the National Science Foundation (BCS 0613319), the Paleontological Scientific Trust Scatterlings of Africa programme, the School of Geography, Archaeology and Environmental Studies and the Institute for Human Evolution, University of the Witwatersrand.

Abstract

Gas chromatography/mass spectrometry, proteomic and scanning electron microscopy with energy-dispersive X-ray spectroscopy (SEM/EDS) analyses of residue on a stone flake from a 49,000 year-old layer of Sibudu (South Africa) indicate a mixture of ochre and casein from milk, likely obtained by killing a lactating wild bovid. Ochre powder production and use are documented in Middle Stone Age South African sites but until now there has been no evidence of the use of milk as a binder. Our analyses show that this ochre-based mixture was neither a hafting adhesive nor a residue left after treating animal skins, but a liquid mixture consisting of a powdered pigment mixed with milk; in other words, a paint medium that could have been applied to a surface or to human skin. The significance of our finds also lies in the fact that it establishes the antiquity of the use of milk as a binder well before the introduction of domestic cattle in South Africa in the first millennium AD.

Introduction

Paint is a mixture of solid pigment and a liquid vehicle that can be applied to a surface or to a body for decorative or protective [1] purposes. We report here an early case of a paint, preserved as a mineral and organic residue on the working edge of a stone flake from a Middle Stone Age (MSA) layer of Sibudu (South Africa), dated to c. 49,000 years ago [2]. Gas chromatography/mass spectrometry (GC/MS), proteomic and SEM/EDS analyses indicate a mixture of ochre and casein from bovid milk [3]. The presence of bovid milk is surprising at this time since domestic cattle are documented in South Africa only in the Iron Age, about 300 AD [4]. However, hunting practices in the MSA suggest that milk could be obtained by killing a

Competing Interests: The authors have declared that no competing interests exist.

lactating female wild bovid [5]. The novelty of our discovery is highlighted by what is known about ochre use from previous research.

Ochre use is ancient, predating anatomically modern humans in both Europe and Africa. Ochre, possibly in a liquid solution, was used by Neanderthals at Maastricht-Belvédère, Netherlands, 200–250 ka BP [6]. Use of ochre in the African Middle Pleistocene is clearly documented, as shown for example at site GnJh-15 in the Kapthurin Formation, Kenya, 285 ka BP [7], in the 250 ka BP Lupemban Industry of Twin Rivers, Zambia [8], and in the early MSA at Sai Island, Sudan, 200 ka BP [9]. There is an increase in the collection and use of ochre showing grinding, scraping or scoring marks at final Middle Pleistocene sites such as Border Cave layers 5BS and 5WA [10–11] and Pinnacle Point Cave 13B, South Africa (164 ka BP) [12–13]. In the Late Pleistocene use of ochre is a common feature of the South African MSA. Discoveries of ground ochre pieces, ochre powder or traces of ochre on grindstones or stone tools have been made at sites such as Apollo 11 [10] Die Kelders [14], Blombos Cave [15–16], Diepkloof [17], Hoedjiespunt 1 [18], Hollow Rock Shelter [19], Klasies River [20–22], Klipdrift [23], Rose Cottage [24] and Sibudu [25–26]. Ochre pieces are also common in all of the South African Cape west coast MSA shell middens [27]. Deliberately engraved and incised ochre pieces have been found at Blombos and Klasies River Cave 1 [22, 28–29].

Ochre powder is excellent for tanning hides because it reverses the process of decay [30–31] but only circumstantial evidence implies the use of ochre for hide preparation in the MSA. Ochre-stained bone awls from Blombos Cave are thought to have pierced hides, based on experimental replications [16]. Residues and use wear on three Sibudu scrapers tentatively imply their use with ochre on hides [32] but a larger sample size is needed. Ochre had other potential roles in the MSA: ochre nodules seem to have been used as soft stone hammers [33] and at MSA sites like Rose Cottage, Sibudu and Umhlatuzana ochre powder was used in compound adhesives for hafting stone tools [24, 34–35].

Watts [13] suggests that the Pinnacle Point red ochre powders were intended as ingredients for body paints used during ritual performance. The inference of early paint manufacture has some support because an ochre-rich compound that may have been blended with marrow or fat was found stored in two Blombos abalone shells with ages of 101 ± 4 ka BP [36]. The purpose of such paint remains speculative, but Henshilwood and colleagues propose that possible uses include decoration of various surfaces and skin protection. Some types of ochre are effective sun screens [1]. Ochre traces on the surface or inside perforated marine shells from Blombos (more than 40 *Nassarius kraussianus* shell beads from the Still Bay layers) Sibudu (three perforated *Afrolittorina africana* from the Sill Bay layers) Border Cave (the *Conus* shell associated with the burial of BC3 an infant with an estimated date of c. 76 ka corresponding to the Howiesons Poort at the site) and North African sites suggest that they were worn against painted bodies, but it is also possible that either the shells or strings for threading them had been deliberately covered with red ochre [37–42].

Materials and Methods

Sibudu is a large rock shelter, approximately 40 km north of Durban and 15 km inland from the Indian Ocean. It has a 2.7 m deep MSA sequence spanning from ca. 77,000 to 38,000 years ago excavated from 1998 to 2011 under the direction of Lyn Wadley [43] (S1 Text. Site setting and excavation). The stone specimen, an unretouched flake of dolerite (length = 26.7 mm; width = 17.5 mm; thickness = 5.1 mm) was recovered from layer MOD dated by single grain OSL to 49.4 ± 2.1 ka [2]. The residue is along the thin, unretouched, right lateral edge (Fig 1A).

Dolerite is an igneous rock that occurs in the vicinity of the site [44]. The flake was found in square D6, quadrant d, where 98 flakes of dolerite were also found. Square D6 yielded 19 pieces

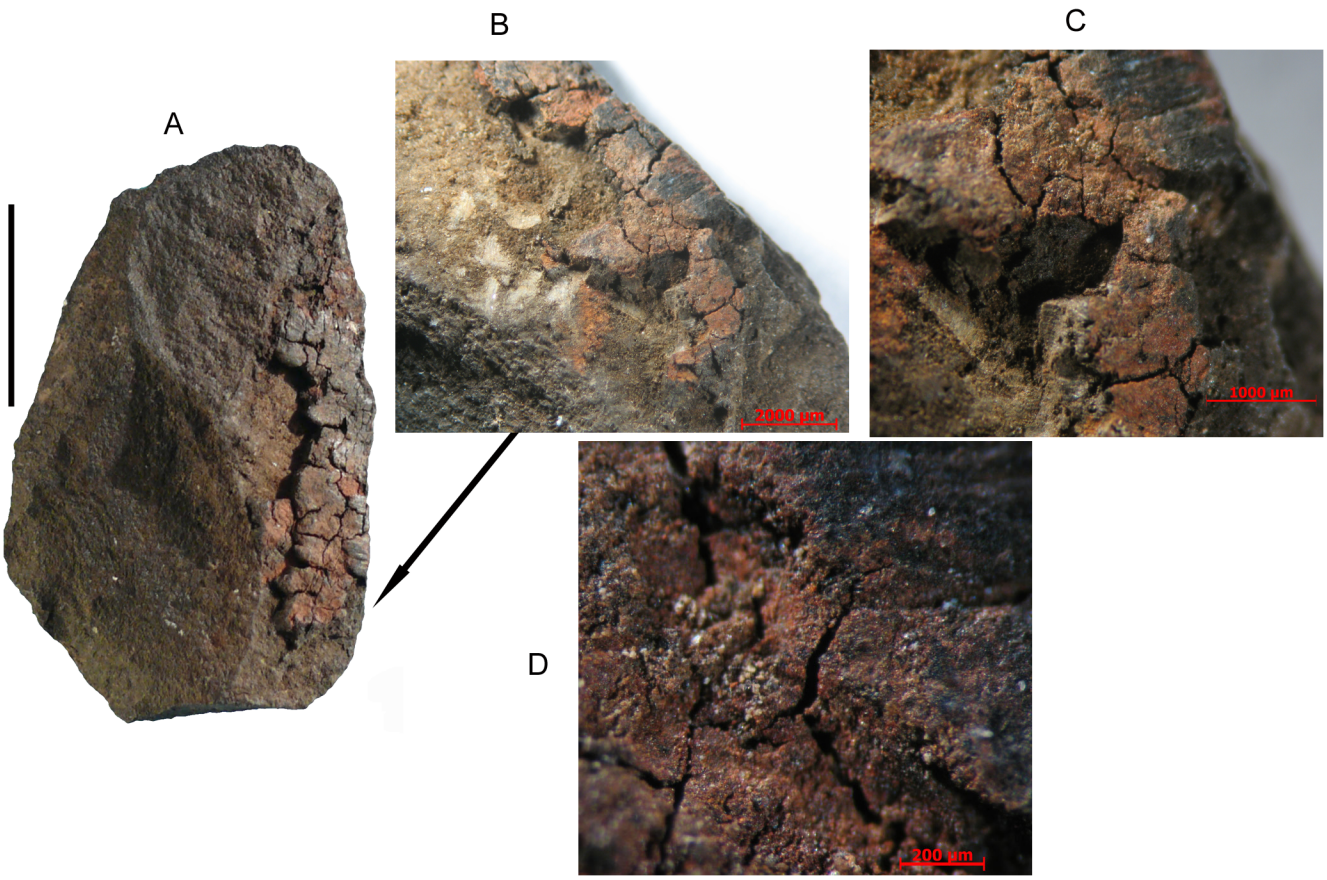


Fig 1. (A) MOD flake before sampling, scale bar = 1 cm. (B) Detail of residue left after sampling for chemical and proteomic analyses. (C) View at 40 x. (D) View at 128 x.

doi:10.1371/journal.pone.0131273.g001

of unutilized and 8 pieces of utilized ochre [45] 405 blades and flakes, 3 cores and 32 retouched pieces. In layer MOD there is a total of 192 ochre pieces.

The excavation permit (number 007/09) was issued by Amafa KwaZulu-Natal Heritage Agency in accordance with KwaZulu-Natal Heritage Act 4 of 2008. The permit holder is Lyn Wadley. The Sibudu collections are housed in the Acacia unit of the Evolutionary Studies Institute at the University of the Witwatersrand and are catalogued 2931CA. The export permit of the flake and other pieces for analyses was issued by Amafa KwaZulu-Natal no. 0011/02. Analysis of the MOD lithic assemblage is provided in S2 Text.

Analysis of the pigment

A small fragment of the residue (dimensions ca. 0.05 x 0.05 x 0.01 mm^3) was carefully hand-picked and mounted on a Gandolfi camera (R = 57.3 mm) for collecting the powder X-ray diffraction pattern using Ni-filtered Cu *Kα* radiation. Owing to the small dimensions of the sample, a 7-day exposure time was set. However, even after 7 days we did not register any sign of diffraction. We tried to capture some weak diffracted spots by means of an air-cooled CCD detector (more sensitive with respect to the photographic film used with the Gandolfi camera) mounted on a Bruker Smart Breeze diffractometer, by rotating the same crystals for several minutes along the φ axis with monochromatized Mo *Kα* radiation. However we still did not get

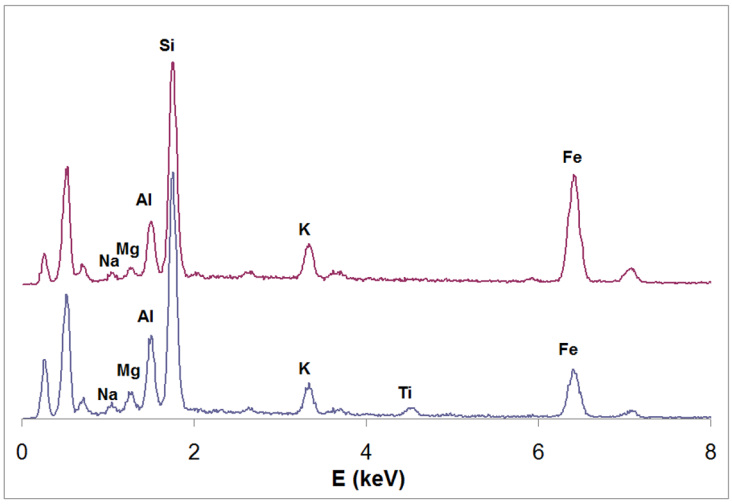


Fig 2. EDS spectra of the residue.

doi:10.1371/journal.pone.0131273.g002

any diffraction. This means that the pigment (Fig 1B–1D) consists of non-crystalline or poorly crystalline material.

The same fragment was studied with a Philips XL30 Scanning Electron Microscope (SEM) equipped with an EDAX DX4 Spectrometer for chemical (elemental) analysis. A total of 5 spot analyses were performed. All spot analyses showed the same elements (in order of decreasing relative abundance): Si, Fe, Al, K, Mg, Na, and minor Ti. The analyses can be grouped in two subsets, with a different ratio between iron, on one side, and the other major elements, on the other (Fig 2).

We conclude that the pigment consists of a mixture, in variable proportions, of ochre which gives rise to the peak of Fe, and a phyllosilicate possibly belonging to clay minerals, e.g. illite/montmorillonite, giving rise to the other peaks. A shale quarry, likely source of clayey ochre, is 1 km from the site [45].

Gas chromatography/mass spectrometry

A micro-sample (3.3 mg) of the amorphous residue on the MOD flake was analyzed by gas chromatography/mass spectrometry (GC/MS) using a combined procedure for the identification of lipids, waxes, proteins, resinous materials and polysaccharides possibly present in the same sample (S4 Text, Chemical analyses). To rule out the possibility of contamination due to burial conditions, some milligrams of sediment were subjected to the same combined analytical procedure.

Protein fraction (analysis of amino acids). The amount of amino acids (0.77 μg) in the sample was significantly above the quantitation limit of the procedure (0.33 μg). The amino acidic profile (Table 1 below and Fig A in S4 Text) is consistent with the presence of casein, given the high percentages of leucine, proline and glutammic acid.

This was confirmed by submitting the amino acidic profile of the sample to Principal Component Analysis (PCA) together with a dataset of more than 100 reference samples of animal glue, casein, and egg, analyzed as raw materials and paint layers mixed with pigments, both naturally and artificially aged. The resulting score plot is reported in Fig 3, and shows that the sample is located in the cluster of casein. The absence of animal glue is confirmed by the absence of hydroxyproline. The amount of amino acids in the soil sample was below detection limit.

Table 1. Amino acidic profile of the MOD flake residue.

AA	%
Alanine	6.6
Glycine	8.6
Valine	11.1
Leucine	18.5
Isoleucine	8.4
Serine	6.4
Proline	16.8
Phenylalanine	6.6
Aspartic acid	5.0
Glutammic acid	12.1
Hydroxyproline	0.0

doi:10.1371/journal.pone.0131273.t001

Polysaccharide fraction (analysis of sugar and uronic acids). The analysis of the sugar and uronic acids fraction did not yield significant results: the amount of analytes was below the quantitation limit of the procedure (0.6 μg).

Lipid and resinous fraction. Several fatty acids were identified (Table A and Fig B in S4 Text). In particular, linear monocarboxylic acids C12-C24 with a minor amount of azelaic acid and glycerol were detected, suggesting the presence of a lipid of plant origin. Sediment samples were used as analytical blanks. The amount of fatty acids in the sediment samples was several orders of magnitude lower than that in the flake sample. Therefore, we can exclude soil contamination. However, the flake fatty acid profile may have been altered by bacteria present in the sediment. Thus, we cannot be confident that the lipid material was actually of plant origin.

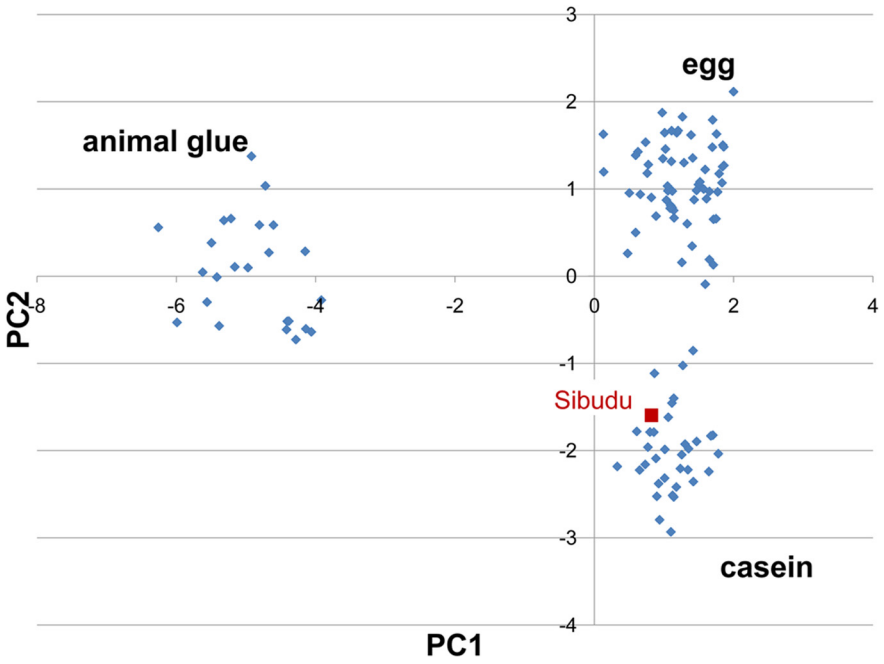


Fig 3. PCA score plot of the database samples and Sibudu sample marked in red.

doi:10.1371/journal.pone.0131273.g003

In conclusion, the occurrence of casein was inferred from the PCA score plot. This interpretation was tested and confirmed by proteomic analyses.

Proteomic analyses

Proteins in the sample were identified following a minimally invasive proteomic procedure using a 4 mg sample of the residue (S4 Text, Chemical analyses). Raw data from nano liquid chromatography-tandem mass spectrometry (LC-MS/MS) were used to query the Swiss Prot database which contains protein sequences of modern species. A database search provided a good score for four bovine milk caseins: Alpha-S1-casein, (P02662), Alpha-S2-casein, (P02663), Beta-casein, (P02666), and Kappa-casein, (P02668). Two collagens [Collagen alpha-1(I) chain, (P02453), and Collagen alpha-2(I) chain, (P02465)] found in bovine species were identified in the sample (Table C in S4 Text).

Proteomics identification is straightforward but it relies on the presence of the sequence to be matched in the database or, at least, that a highly homologous protein from another species is available. Some peptides identical to the ones present in the known protein homologue are then expected to be produced in the enzymatic digestion of the protein of interest and can be matched. Although casein proteins are highly divergent across mammalian milks [46], they are conserved enough to allow identification, with some peptides identical to the ones present in the old protein homologue. Divergence, moreover, allowed genus-specific sequence information, since several of the matched peptides are specific enough to allow us to confidently identify the bovid origin of the caseins. In Fig F of S4 Text the alpha-S1 casein sequences from *Bos taurus* were aligned to those from some other potential sources of milk, i.e. human, sheep and goat. The matched peptides in the MOD flake (highlighted in grey and bold) clearly indicate that the casein belongs to an animal in the bovid family. Likewise, alignments of the other identified caseins with homologous proteins are similarly reported by [47] showing that several of the identified peptides are proteotipic of bovine caseins sequences.

Collagen was not identified in the fraction of the sample that was analyzed by GC/MS given the absence in the amino-acidic profile of hydroxyproline which is a marker of animal glue. This indicates that collagen was not uniformly mixed in the sample as one would expect if collagen was used as animal glue and that the collagen source was most likely a minute bone fragment or a fragment of connective or epithelial tissue (see Discussion).

Type I collagen is sufficiently variable between mammal genera to be taxonomically useful. Collagen from several bone samples of equid and bovid fauna from layer MOD was extracted and analyzed. Peptide matches from each sample and from the MOD flake were manually examined and aligned to the collagen sequences from modern species. Significantly comparisons with samples of equid and bovid fauna from layer MOD indicate that the collagens were from a bovid and not from an equid (Figs G and H in S4 Text).

In sum, independent analyses by two laboratories on different fraction of the residue and with different techniques support the identification of casein. Moreover proteomic analysis shows that the casein belongs to an animal in the bovid family.

Direct dating

Following the GC/MS and proteomic analyses which suggested the presence of organic material still preserved in the residue, the sample was considered suitable for direct AMS dating. The analysis was performed at the Oxford Radiocarbon Accelerator Unit (ORAU), University of Oxford, UK.

The direct AMS dating of the residue is 23.3 ± 0.4 ka BP (uncalibrated radiocarbon years before present). Lack of chemical cleaning, a decision imposed by the very small size of the

sample (13 mg), meant that any contaminants would remain intact contributing to the overall ^{14}C measurement. The determination therefore is considered minimum age only. The larger than normal standard error is a factor of the small amount of C measured in the sample (390μg), about 5 times smaller than the normal-sized graphites routinely measured at the ORAU.

The significance of this direct measurement lies in the fact that it establishes the antiquity of the use of milk well before the introduction of domestic cattle in South Africa in the 1st millennium AD (S3 Text. Direct dating).

Discussion

Milk is an emulsion of fats within a water-based fluid which also contains carbohydrates, proteins and minerals. Casein is the major protein component of milk and is extracted by heating milk and adding acid [3]. Casein then precipitates into small, solid globules. Use of casein as a glue requires mixing with quicklime or hydrated lime (calcium hydroxide, also called slaked lime) [48]. Lime is obtained by crushing and heating limestone. However, recent geological mapping shows that there is no limestone in the Sibudu area [49]. For use as a binder for pigments, lumps of casein can also be dissolved and thinned with water and wood ashes (mainly composed of calcium carbonate or calcium oxide). However calcium (Ca) is absent in the spectra of Fig 2. Given the absence of lime and wood ashes in the Sibudu residue, we reject an interpretation of the casein as either glue or a complex form of binder. Instead, we conclude that powdered ochre was simply mixed with milk in its liquid form and casein is the residue left after evaporation, since milk is a water-based liquid.

The identification of a milk and ochre mixture on a 49,000 year old flake from a late MSA layer of Sibudu raises a number of questions indicated below.

How was the milk obtained from wild bovids?

Some African bovids separate from the herd when giving birth and hide their calves until they are strong enough to keep up with the herd as a protection against carnivores [50]. This is the case with the eland (*Taurotragus oryx*), the kudu (*Tragelaphus strepsiceros*), and the impala (*Aepicerus melampus*). The bushbuck (*Tragelaphus scriptus*) and the red and blue duiker (*Cephalophus natalensis* and *Philantomba monticola*) are solitary animals; they hide their young and go off to browse alone. All these species and the buffalo (*Syncerus caffer*, a herd animal) are represented in the MOD fauna.

Klein [5] notes that in the MSA deposits of Klasies River Cave I the giant buffalo (*Pelorovis antiquus*) a very large bovid with a horn span of 3 m or more, falls into two age groups: newborn or perhaps fetal individuals and physically mature individuals with worn molars. According to Klein, the explanation for this peculiar distribution may be that the Klasies people, faced with such a formidable prey, focused on females in advanced pregnancy or in the process of giving birth. Cows are easy prey when giving birth and would already have milk. It would not be difficult for hunters to locate lactating cows, particularly among seasonal breeders. This hunting pattern would result in access to milk and might explain the inhomogeneous presence of collagen observed by proteomic analyses. In the process of collecting milk from the mammary glands of a dead animal, a small amount of its epithelial tissue could have been taken as well.

Was the mixture a hafting adhesive?

The use of adhesives made from resin or plant gum combined with ochre has previously been suggested for the hafting of stone points and backed pieces to be used as hunting tools at

Sibudu. Howiesons Poort and later MSA tools were studied by optical microscopy in combination with spatial distribution of residues on the tool [35, 51]. Replication experiments [34, 52] demonstrated the usefulness of ochre as a loading agent with products such as plant gum. Microscopic analysis showed that ground ochre was mixed with a plant exudate; experiments suggested that ochre-loaded adhesives are robust, dry fast and are not hygroscopic. Most archaeological residues occurred at the base of the point or on the back of backed pieces implying that the mixture was used for hafting. However optical microscopy alone cannot securely distinguish between plant gum (polysaccharides) and resin (terpenoid secretions from trees) [53] thus new analyses using GC/MS are appropriate.

The residue on the MOD flake occurs on the thin edge of an unretouched flake, not on the back or proximal part of the piece as would be expected if used for hafting. An alternative explanation is that the flake was used to prepare or apply the ochre-loaded mixture. GC/MS has been used to analyze six tools, dated between 65 and 38 ka BP, from Sibudu (n = 5) and Rose Cottage (n = 1). The residues were preliminarily ascertained by visual inspection. All samples were provided by Lyn Wadley. The purpose was to identify the residues on the MSA tools and to determine whether casein could have been used as glue in a variant of the recipes used at Sibudu.

The results (S4 Text, Chemical analyses) indicate that two of the Howiesons Poort segments, from layer GR in square C6c and from layer PGS in square B5a, contained a relevant amount of diterpenes indicating the use of a conifer resin, possibly from *Afrocarpus* (syn. *Podocarpus*) *falcatus*, and some lipid material. There is neither evidence of plant gums (no polysaccharides) nor proteinaceous material such as casein. No significant molecular markers of organic material were detected on the other samples; neither saccharides nor terpenes were detected. Fatty acids were present in very low quantities suggesting soil contamination. The use of conifer resin for hafting (*Podocarpus elongatus*) is also indicated by GC/MS analysis on a quartz backed flake dated to ca. 56 ± 10 ka BP in the Late Howiesons Poort at the site of Diepkloof [54].

As mentioned before, the use of casein as a glue requires mixing with hydrated lime [48] which is obtained by crushing and heating limestone. Recent geological mapping shows that there is no limestone in the Sibudu area [49] so, in conclusion, there is no evidence that the MOD flake residue was a hafting adhesive.

Was the MOD flake used to treat animal skin?

Three scrapers from layer SS at Sibudu, dated to 59.6 ± 2.3 ka by OSL [2], have ochre and animal products (fat, muscle tissue) on their edges. Ochre and fat were quite probably used for processing hides and some of the animal products were the result of use on animal skin [32]. The presence of milk is however not consistent with use for processing hides.

What was the use of the MOD flake mixture?

Our analyses show that this ochre-based mixture was neither a hafting adhesive nor a residue left after treating animal skins, but a liquid mixture consisting of a solid (powdered) pigment mixed with milk; in other words, a paint medium that could have been applied to a surface or on the skin.

The edge of the flake may have been used as a mixing implement for combining the substances. A mixture of milk and ochre may have been used for body painting, skin protection or for painting on a rock surface like a stone slab. The oldest representational rock art in Africa is on seven small quartzite slabs from the Apollo 11 rock shelter in Namibia. The date of 27,500 years ago is an average of 3 radiocarbon samples of charcoal in hearths from the same layer

[55]. The pigments and binders of the slabs have not yet been analyzed. Traces of red coloring on ostrich eggshell fragments from the older MSA layers in Apollo 11 [56] imply that there may have been simpler precursors to the representational art. Milk or casein has never been documented as media for pigment in San rock art [57]. In fact, little is known about the methods or ingredients that San used for making paint and even less is known about paint binders [58–61]. Blood was identified in rock art paint at Rose Cottage [62]. Some ethnographies and historic sources mention milk and egg as a binder for rock paint but there is as yet no direct evidence for its use [57].

Milk was not used as a binder with pigments for body paint by the San (Bushman) nor by the Hottentots and Nama herders of South Africa, Namibia and Botswana; animal or vegetable fat were the most commonly reported media for body paint [58]. However clarified butter is mixed with ochre for body paint used by the Himba of Namibia [1]. Only further research on pigments and binders of rock art in South Africa will allow us to identify similarities or differences that may support one hypothesis over the other.

Nevertheless, obtaining milk by killing a lactating wild bovid and then mixing it with ochre shows that MSA people experimented with coloring materials in creative ways and may have attributed a special significance and value to that product. In the present state of research we cannot demonstrate conclusively whether paint such as that found at Sibudu was a precursor to rock art painting or used for body painting. In either case this find suggests that toward the end of the Middle Stone Age new techniques of symbolic communication were evolving that became, some millennia later, key elements in the social life of hunter-gatherers in South Africa.

Supporting Information

S1 Text. Site setting and excavation
(PDF)

S2 Text. Lithic analysis of layer MOD
(PDF)

S3 Text. Direct dating
(PDF)

S4 Text. Chemical analyses
(PDF)

Acknowledgments

PV is grateful to Scott Ortman, Payson Sheets and John Hoffecker for constructive comments. LB thanks Gennaro Marino for helpful discussions and critical reading.

Author Contributions

Conceived and designed the experiments: PV. Performed the experiments: PV LP ID LB MP CB KD RV JJL. Analyzed the data: PV LP ID LB MP CB KD RV JJL. Contributed reagents/materials/analysis tools: ID LB MP CB KD RV JJL LW. Wrote the paper: PV LP ID LB MP CB KD LW.

References

1. Rifkin RR, d'Errico F, Boulliot L, Summers B. Assessing the photoprotective effects of red ochre on human skin by in vitro laboratory experiments. S Afr J Sci. 2015; 111(3/4): 1–8.

2. Jacobs Z, Wintle AG, Duller GAT, Roberts RG, Wadley L. New ages for the post-Howiesons Poort, late and final Middle Stone Age at Sibudu, South Africa. J Archaeol Sci. 2008; 35: 1790–1807.
3. Colombini MP, Modugno F, Eds. Organic mass spectrometry in art and archaeology. Chichester: Wiley; 2009.
4. Robbins LH, Campbell AC, Murphy ML, Brook GA, Srivastava P, Badenhorst S. The advent of herding in Southern Africa: Early AMS dates on domestic livestock from the Kalahari desert. Curr Anthropol. 2005; 46: 671–677.
5. Klein RG. Middle Stone Age man-animal relationships in Southern Africa: Evidence from Die Kelders and Klasies River Mouth. Science. 1975; 190: 265–267.
6. Roebroeks W, Sier MJ, Nielsen TK, De Loecker D, Parés JM, Arps CES, et al. Use of red ochre by early Neandertals. Proc Natl Acad Sci. 2012; 109: 1889–1894. doi: 10.1073/pnas.1112261109 PMID: 22308348
7. McBrearty S, Brooks AS. The revolution that wasn't: A new interpretation of the origin of modern human behavior. J Hum Evol. 2000; 39: 453–563. PMID: 11102266
8. Barham LS. Possible early pigment use in South-Central Africa. Curr Anthropol. 1998; 39: 703–710.
9. Van Peer P, Fullagar R, Stokes S, Bailey RM, Moeyersons J, Steenhoudt F, et al. The Early to Middle Stone Age transition and the emergence of modern human behaviour at site 8-B-11, Sai Island, Sudan. J Hum Evol. 2003; 45: 187–193. PMID: 14529653
10. Watts I. Ochre in the Middle Stone Age of southern Africa: ritualized display or hide preservative? S Afr Archaeol Bull. 2002; 57 (175): 64–74.
11. Grün R, Beaumont P, Tobias PV, Eggis S. On the age of the Border Cave 5 human mandible. J Hum Evol. 2003; 45: 155–167. PMID: 14529650
12. Marean CW, Bar-Matthews M, Bernatchez J, Fisher E, Goldberg P, Herries AIR, et al. Early human use of marine resources and pigment in South Africa during the Middle Pleistocene. Nature. 2007; 449: 905–908. PMID: 17943129
13. Watts I. The pigments from Pinnacle Point Cave 13B, Western Cape, South Africa. J Hum Evol. 2010; 59: 392–411. doi: 10.1016/j.jhevol.2010.07.006 PMID: 20934093
14. Thackeray AI. Middle Stone Age artefacts from the 1993 and 1995 excavations of Die Kelders Cave 1, South Africa. J Hum Evol. 2000; 38: 147–168. PMID: 10627401
15. Henshilwood CS, Sealy J, Yates R, Cruz-Uribe K, Goldberg P, Grine FE, et al. Blombos Cave, Southern Cape, South Africa: preliminary report on the 1992–1999 excavations of the Middle Stone Age levels. J Archaeol Sci. 2001; 28: 421–448.
16. Henshilwood CS, d'Errico F, Marean CW, Milo RG, Yates R. An early bone tool industry from the Middle Stone Age at Blombos Cave, South Africa: implications for the origins of modern human behaviour, symbolism and language. J Hum Evol. 2001; 41: 631–678. PMID: 11782112
17. Dayet L, Texier P-J, Daniel F, Porraz G. Ochre resources from the Middle Stone Age sequence of Diepkloof Rock Shelter, Western Cape, South Africa. J Archaeol Sci. 2013; 40: 3492–3505.
18. Will M, Parkington JE, Kandel AW, Conard NJ. Coastal adaptations and the Middle Stone Age lithic assemblages from Hoedjiespunt 1 in the Western Cape, South Africa. J Hum Evol. 2013; 64: 518–537. doi: 10.1016/j.jhevol.2013.02.012 PMID: 23628196
19. Evans U. Hollow Rock Shelter, a Middle Stone Age site in the Cederberg. South Afr Field Archaeol.1994; 3: 63–73.
20. Singer R, Wymer J. The Middle Stone Age at Klasies River Mouth in South Africa. Chicago: Chicago University Press; 1982.
21. Deacon HJ. Two late Pleistocene-Holocene archaeological depositories from the Southern Cape, South Africa. S Afr Archaeol Bull. 1995; 50: 121–131.
22. d'Errico F, García Moreno R, Rifkin RF. Technological, elemental and colorimetric analysis of an engraved ochre fragment from the Middle Stone Age levels of Klasies River Cave 1, South Africa. J Archaeol Sci. 2012; 39: 942–952.
23. Henshilwood CS, van Niekerk K, Wurz S, Delagnes A, Armitage SJ, Rifkin RF, et al. Klipdrift Shelter, southern Cape, South Africa: Preliminary report on the Howiesons Poort layers. J Archaeol Sci. 2014; 45: 284–303.
24. Gibson NE, Wadley L, Williamson BS. Microscopic residues as evidence of hafting on backed tools from the 60 000 to 68 000 year-old Howiesons Poort layers of Rose Cottage Cave, South Africa. South Afr Humanit. 2004; 16: 1–11.
25. Wadley L. Cemented ash as a receptacle or work surface for ochre powder production at Sibudu, South Africa, 58,000 years ago. J Archaeol Sci. 2010; 37: 2397–2406.

26. Hodgskiss T. Ochre use in the Middle Stone Age at Sibudu, South Africa: Grinding, rubbing, scoring and engraving. J Afr Archaeol. 2013; 11: 75–95.
27. Parkington JE. Moderns and middens: Shellfishing and the Middle Stone Age of the Western Cape. S Afr J Sci. 2003; 99: 243–247.
28. Henshilwood CS, d'Errico F, Yates R, Jacobs Z, Tribolo C, Duller GAT, et al. Emergence of modern human behaviour: Middle Stone Age engravings from South Africa. Science. 2002; 295: 1278–1280. PMID: 11786608
29. Henshilwood CS, d'Errico F, Watts I. Engraved ochres from theMiddle Stone Age levels at Blombos Cave, South Africa. J Hum Evol. 2009; 57: 27–47. doi: 10.1016/j.jhevol.2009.01.005 PMID: 19487016
30. Audouin F, Plisson H. Les ochres et leurs témoins au Paléolithique en France: Enquéte et expériences sur leur validité archéologique. Cahiers du Centre de Recherches Préhistoriques 1982; 8: 33–80.
31. Rifkin RF. Assessing the efficacy of red ochre as a prehistoric hide tanning ingredient. J Afr Archaeol. 2011; 9: 131–158.
32. Wadley L, Langejans G. Preliminary study of scrapers around combustion features in layer SS, Sibudu, 58 000 years ago. S Afr Archaeol Bull. 2014; 69: 19–33.
33. Soriano S, Villa P, Wadley L. Ochre for the toolmaker: Shaping the Still Bay points at Sibudu (KwaZulu-Natal, South Africa). J Afr Archaeol. 2009; 7: 41–54.
34. Wadley L, Hodgskiss T, Grant M. Implications for complex cognition from the hafting of tools with compound adhesives in the Middle Stone Age, South Africa. Proc Natl Acad Sci. 2009; 106: 9590–9594. doi: 10.1073/pnas.0900957106 PMID: 19433786
35. Lombard M. The gripping nature of ochre: the association of ochre with Howiesons Poort adhesives and Later Stone Age mastics from South Africa. J Hum Evol. 2007; 53: 406–419. PMID: 17643475
36. Henshilwood CS, d'Errico F, van Niekerk K, Coquinot Y, Jacobs Z, Lauritzen SE, et al. A 100,000-year-old ochre-processing workshop at Blombos Cave, South Africa. Science. 2011; 334: 219–222. doi: 10.1126/science.1211535 PMID: 21998386
37. Henshilwood CS, d'Errico F, Vanhaeren M, van Niekerk K, Jacobs Z. Middle Stone Age shell beads from South Africa. Science. 2004; 304: 404. PMID: 15087540
38. d'Errico F, Henshilwood CS, Vanhaeren M, van Niekerk K. *Nassarius kraussianus* shell beads from Blombos Cave: Evidence for symbolic behaviour in the Middle Stone Age. J Hum Evol. 2005; 48: 2–14.
39. d'Errico F, Vanhaeren M, Wadley L. Possible shell beads from the Middle Stone Age layers of Sibudu Cave, South Africa. J Archaeol Sci. 2008; 35: 2675–2685.
40. d'Errico F, Vanhaeren M, Barton N, Bouzouggar A, Mienis H, Richter D, et al. Additional evidence on the use of personal ornaments in the Middle Paleolithic of North Africa. Proc Natl Acad Sci. 2009; 106: 16051–16056. doi: 10.1073/pnas.0903532106 PMID: 19717433
41. Grün R, Beaumont P. Border Cave revisited: a revised ESR chronology. J Hum Evol. 2001; 40: 467–482. PMID: 11371150
42. Millard AR. Bayesian analysis of ESR dates with application to Border Cave. Quat Geochronol. 2006; 1: 159–166.
43. Wadley L., Jacobs Z Z.. Sibudu Cave: background to the excavations, stratigraphy and dating. South Afr Humanit. 2006; 18: 1–26.
44. Wadley L, Kempson H. A review of rock studies for archaeologists, and an analysis of dolerite and hornfels from the Sibudu area, KwaZulu-Natal. South Afr Humanit.2011; 23: 87–107.
45. Hodgskiss T. An investigation into the properties of the ochre from Sibudu, KwaZulu-Natal, South Africa. South.Afr Humanit. 2012; 24: 99–120.
46. Lemay DG, Lynn DJ, Martin WF, Neville MC, Casey TM, Rincon G, et al. The bovine lactation genome: insights into the evolution of mammalian milk. Genome Biol. 2009; 10 (4): R43. doi: 10.1186/gb-2009-10-4-r43 PMID: 19393040
47. Kawasaki K, Lafont AG, Sire JY. The evolution of milk casein genes from tooth genes before the origin of mammals. Mol Bio Evol. 2011; 28: 2053–61.
48. Forest Products Laboratory. Casein glues: their manufacture, preparation and application. U.S. Forest Service Research Note, FPL-0158; 1967. pp. 1–15.
49. Clarke BM, Schoeman P, Kota MW. The geology of the Verulam area. Pretoria: Council for Geoscience, D&V Premier Print; 2007.
50. Stuart C, Stuart T. Field guide to the mammals of Southern Africa. Cape Town: Struik; 1997.
51. Lombard M. Finding resolution for the Howiesons Poort through the microscope: micro-residue analysis of the segments from Sibudu Cave, South Africa. J Archaeol Sci. 2008; 35, 26–41.

52. Wadley L. Putting ochre to the test: Replication studies of adhesives that may have been used for hafting tools in the Middle Stone Age. J Hum Evol. 2005; 49: 587–601. PMID: 16126249

53. Burger P, Charrié-Duhaut A, Connan J, Albrecht P. Taxonomic characterisation of fresh Dipterocarpaceae resins by gas chromatography–mass spectrometry (GC-MS): Providing clues for identification of unknown archaeological resins. Archaeol Anthropol Sci. 2011; 3; 185–200.

54. Charrié-Duhaut A, Porraz G, Cartwright CR, Igreja M, Connan J, Poggenpoel C, et al. First molecular identification of a hafting adhesive in the Late Howiesons Poort at Diepkloof Rock Shelter (Western Cape, South Africa). J Archaeol Sci. 2013; 40: 3506–3518.

55. Wendt WE. Art mobilier' from the Apollo 11 Cave, South West Africa: Africa's oldest dated works of art. S Afr Archaeol Bull. 1976; 31: 5–11.

56. Vogelsang R. Middle-Stone-Age-Fundstellen im Südwest Namibia. Africa Praehistorica 11. Köln: Heinrich Barth Institute; 1998.

57. Rudner I. Paints of the Khoisan artists. S Afr Archaeol Soc., Goodwin Series 1983; 4: 14–20.

58. Rudner I. Khoisan pigments and paints and their relationship to rock paintings. Ann S Afr Mus. 1982; 87: 1–181.

59. Arocena JM, Hall K, Meiklejohn I. Minerals provide tints and possible binder/extender in pigments in San rock paintings (South Africa). Geoarchaeol. 2008; 23: 293–304.

60. Bonneau A, Pearce DG, Pollard AM. A multi-technique characterization and provenance study of the pigments used in San rock art, South Africa. J Archaeol Sci. 2012; 39: 287–294.

61. Prinsloo L, Tournié A, Colomban P, Paris C, Bassett ST. In search of the optimum Raman/IR signatures of potential ingredients used in San/Bushman rock art paint. J Archaeol Sci. 2013; 40: 2981–2990.

62. Williamson BS. Direct testing of rock painting pigments for traces of haemoglobin at Rose Cottage Cave, South Africa. J Archaeol Sci. 2000; 27: 755.

Other publications

1. Itri, F., Monti, D.M., Della Ventura, B., Vinciguerra, R., Chino, M., Gesuele, F., Lombardi, A., Velotta, R., Altucci, C., Birolo, L., Piccoli, P., Arciello, A. Femtosecond UV-laser pulses to unveil protein–protein interactions in living cells. (2015) Cellular and Molecular Life Sciences CMLS; DOI: 10.1007/s00018-015-2015-y.

2. Iannuzzi, C., Carafa, V., Altucci, L., Irace, G., Borriello, M., Vinciguerra, R., Sirangelo I. Glycation of Wild-Type Apomyoglobin Induces Formation of Highly Cytotoxic Species. (2015) Journal of Cellular Physiology. Apr 3. doi: 10.1002/jcp.25011.

3. Amore, A., Parameswaran, B., Kumar, R., Birolo, L., Vinciguerra, R., Marcolongo, L., Ionata, E., La Cara, F., Pandey, A., and Faraco, V. Application of a new xylanase activity from Bacillus amyloliquefaciensXR44A in Brewer's Spent Grain saccharification. (2014) Journal of Chemical Technology and Biotechnology. doi: 10.1002/jctb.4589.

4. Amore, A., Serpico, A., Amoresano, A., Vinciguerra, R., Faraco, V. Analysis of the role of O-glycosylation in GH51 α-L-arabinofuranosidase from Pleurotusostreatus. (2014) BiotechnolApplBiochem. doi: 10.1002/bab.1325.

5. Montibeller, V. W., Vandenberghe, L. P. S., Amore, A., Soccol, C. R., Birolo, L., Vinciguerra, R., Salmon, D. N. X., Spier, M. R., Faraco, V. Characterization of hemicellulolytic enzymes produced by Aspergillus niger NRRL 328 under solid state fermentation on soybean husks.(2014)BioRes. 9(4), 7128-7140.

6. Giacobbe, S., Pepe, O., Ventorino, V., Birolo, L., Vinciguerra, R., and Faraco, V. Identification and characterisation of a pectinolytic enzyme from Paenibacillus xylanolyticus. (2014) BioRes. 9(3), 4873-4887.

Chapter of ICOMOS Publication (in press)
Anna Lluveras-Tenorio, Leila Birolo, Catharina Blaensdorf, Ilaria Bonaduce, Marine Cotte, Eugenio Galano, Roberto Vinciguerra, Emeline Pouyet, Maria Perla Colombini.
"Analysis of binders to reconstruct the painting technique of the Western and Eastern Buddhas of the Bamiyan valley (Afghanistan)".

Manuscripts in preparation and submitted

1. Lluveras-Tenorio, A., Vinciguerra, R., Galano, E., Blaensdorf, C., Emmerling, E., Colombini, M.P., Birolo, L., Bonaduce, I. The Painting Technique of the Giant Buddhas of Bamiyan (Afghanistan). Submitted.

2. Vergara, A., et al. Different technological solutions of Roman engineers for public outdoor and private indoor wall paintings. Manuscript in preparation.

Oral communication

1. Proteomics and cultural heritage: from bones to paintings.
   Roberto Vinciguerra, Addolorata De Chiaro, Piero Pucci, Gennaro Marino, Leila Birolo. Non-destructive and micro-analytical techniques in art and cultural heritage. TECHNART 2015 - Catania, Italy, April 27- 30, 2015.

2. Multidiagnostic Chemical Analysis of Wall Paintings from Cuman Archaeological Site
   Leila Birolo, Roberto Vinciguerra, Marco Trifuoggi, Luciano Ferrara, Carla De Maio, Giovanna Greco, Antonella Tomeo, Alessandra Lucchini, Luigi Paduano, Alessandro Vergara. 8th International Conference on the Application of Raman spectroscopy in Art and Archaeology, Wroclaw, Poland, September 1-5, 2015.

3. A first insight into the Asian clay sculptures painting techniques and materials: use of proteomics strategies on the Buddhas of the Bamiyan Valley (Afghanistan)
   Roberto Vinciguerra, Eugenio Galano, Addolorata De Chiaro, Ilaria Bonaduce, Anna Lluveras-Tenorio, Maria Perla Colombini, Catharina Blaensdorf, Pietro Pucci, Leila Birolo.
   ITPA 'Italian Proteomics Association' Napoli, June 24-27, 2014.

4. Proteomic strategies for the study of works of art and archaeological finds.
   Roberto Vinciguerra, Eugenio Galano, Angela Amoresano, Piero Pucci, Leila Birolo.
   "I giovani e la spettrometria di massa 2013" 4 MS-J-Day, Società chimica italiana, November, 14, 2013.

Poster section

- Proteomics strategies for the analysis of ancient bones.
  De Chiaro, A., Finicelli, M., Cirillo, A., Vinciguerra, R., Cipollaro, M., Pucci, P., Peluso G., Birolo. L. IX ItPA Congress 2014, Napoli.

- Preliminary Multidiagnostic Chemical analysis of wall paintings from a cuman archaeological site
  Birolo, L., Vinciguerra, R., Ferrara, L., De Maio, C., Greco, G., Tomeo, A., Trifuoggi, M., Vergara, A. IVth Conference "Diagnosis, Conservation and Valorization of Cultural Heritage" 12/13 December 2013.

- Enzyme Activities in the Organic Horizon and Soil Profile of a South Italy Beech Stand
  Antonella Pellegrino, Stefania Papa, Roberto Vinciguerra, Giovanna Ventriglia, Francesco Guadagno, Antonietta Fioretto. XXI Congresso Nazionale S.It.E. I limiti dello sviluppo: beni e servizi ecosistemici, impatti e gestione. Palermo, October 3-6 2011.