Marina Creydt, Markus Fischer

Mass Spectrometry-Based Proteomics and Metaproteomics Analysis of Ancient Manuscripts

Abstract: Ancient manuscripts may contain a large amount of additional information besides the written words. This information is not readily visible to the naked eye and is also analytically very challenging to uncover. In the last ten years, mass spectrometric techniques in particular have emerged in this context to provide additional information about the history of an ancient manuscript and the living conditions at that time through proteomics and metaproteomics analyses. This additional information includes not only the methods of book production and the raw materials used, but also the historical usages of the manuscripts. This review is intended to provide insights into the scientific questions that can be addressed with mass spectrometric proteome analyses, as well as an overview of the possible methods and procedures. In addition to the correct handling of the valuable samples, the various possibilities of sampling as non-invasively as possible as well as technical aspects and data evaluation that need to be considered will be discussed.

1 Introduction

In the last twenty years, the so-called 'omics-disciplines' have increasingly moved into the focus of many research areas. This includes not only the classical disciplines of the natural sciences, but increasingly also the humanities to address archaeometric issues. The suffix '-omics' describes the hypothesis-free, non-targeted analysis of complete or almost complete cellular levels (genome, transcriptome, proteome, metabolome), although even these techniques often fail to detect all substances or sequences. Depending on the cellular level analysed, the terms genomics, transcriptomics, proteomics and metabolomics have been introduced, which together describe the flow of information from genotype to phenotype (Fig. 1).[1] In addition, many other neologisms have since emerged, including the research areas paleogenomics and paleoproteomics. According to

1 Creydt and Fischer 2018; Dettmer et al. 2007.

https://doi.org/10.1515/9783110753301-010

Hendy's recently published definition, the focus of both disciplines is the study of 'archaeological, historical, and paleontological remains and materials' using ancient DNA (aDNA) or ancient proteins or ancient peptides.[2] This is because it can be assumed that, depending on the history, state of preservation, the age of a sample or microbial infestation, some of the proteins have since been degraded into peptides. However, little is currently known about the degree of protein conservation in ancient samples, as the detection methods that have been predominantly used to date are based on the detection of short amino acid sequences (see Section 3.4 top-down approaches).[3] In addition, proteins have a low solubility in water and are comparatively difficult to extract using standard methods, especially since proteins agglomerate and show crosslinking reactions over the time, further complicating their analysis, which is why the analytical focus is usually on peptides.[4]

Palaeoproteomics is a relatively young discipline that emerged in the early 2000s and which has become increasingly relevant due to the advancing development in mass spectrometry (MS).[5] Although the discovery that amino acids can also be detected in fossils that are hundreds of millions of years old is not new and can be traced back to the 1950s.[6] Compared to aDNA, it is assumed that ancient peptide sequences are more stable and can therefore also be used for an analysis if aDNA can no longer be detected. According to current knowledge, the oldest DNA residues that could still be sequenced, were found in a tooth from *Homo antecessor*, which are estimated to be 800,000 years old, and in a horse bone, which is approximately 700,000 years old.[7] In contrast to these results, large proteome residues could also be detected in the molar tooth of a rhinoceros (*Stephanorhinus*), which lived 1.8 million years ago after the attempt to obtain DNA initially failed.[8] Furthermore, performing proteome analyses is quicker and cheaper compared to genome studies.[9]

Nevertheless, the analysis of the proteome cannot completely replace the DNA analysis, so that these two techniques complement each other optimally. For example, a comparatively quick species identification can be made with proteome analyses, but the sex or relatedness cannot be determined, which in

2 Hendy 2021.
3 Cleland and Schroeter 2018; Hendy 2021.
4 Orsini et al. 2017.
5 Ostrom et al. 2000.
6 Abelson 1954a; Abelson 1954b.
7 Orlando et al. 2013; Welker et al. 2020.
8 Cappellini et al. 2019.
9 Fiddyment et al. 2019.

turn is possible using DNA studies.[10] For this reason, it has recently been proposed to carry out proteome analyses first and then proceed with DNA analyses based on the results obtained.[11]

2 Which scientific questions can be answered in the context of manuscript science?

The word 'manuscript' is derived from the Latin 'manu scriptum' which means 'written by hand'. With the development of the first scripts and alphabets in different parts of the world, a wide variety of raw materials were used as writing surfaces, depending on availability. These included, for example, metal, fired and unfired clay, wax tablets, palm leaves, wood, bark, bamboo, cloths made of cotton, linen or silk, bones and ivory, animal skins that haven been converted into leather, parchment, or vellum as well as papyrus and paper. As diverse as the writing surfaces are, so are the writing instruments.[12] Accordingly, a wide variety of analytical platforms have to be used for the analysis of manuscripts and numerous of complementary techniques and sampling procedures are often required in order to obtain correct interpretations and assumptions, as has recently been demonstrated by some forgeries.[13] In addition to the genome and proteome already mentioned, small molecules (<1,000 Da), elements or isotope ratios are also analysed.[14]

Using proteomics-based methods, it is possible to more deeply characterise organic writing materials that still contain protein residues and to identify the species (Fig. 1). This works especially well with writing materials that are made from animal skins. In this context, the analysis of parchment is particularly relevant because a large part of the historically preserved documents were made on parchment. The identification of the animal species used allows conclusions to be drawn about the geographical origin, since the parchment was made with animals (e.g. sheep, cattle, goat) that were available locally. Sometimes the intended use is related to the type of animal skin. For example, a special animal

10 Teasdale et al. 2021.
11 Collins et al. 2006; Fiddyment et al. 2019.
12 Colomo 2012.
13 Newton et al. 2018; Rabin and Hahn 2020.
14 Dallongeville et al. 2016; Vilanova and Porcar 2020.

parchment was sometimes used for illustrations.[15] Toniolo et al. were the first to obtain a sufficient amount of proteins or peptides from a very thin parchment (vellum) to perform species identification on a thirteenth-century pocket Bible.[16]

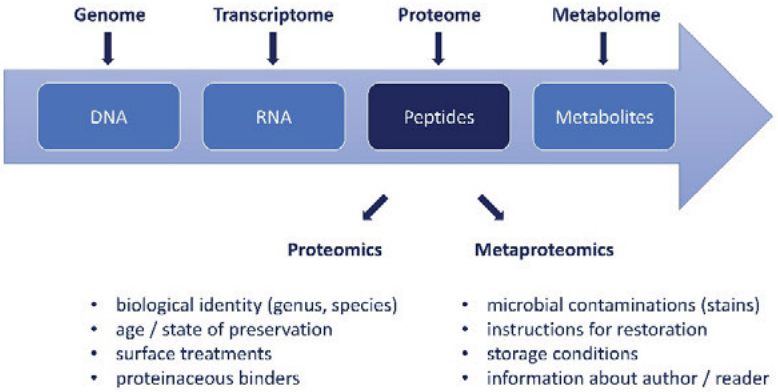

Fig. 1: The cellular cascade and its hidden information that can be obtained with proteomics and metaproteomics analyses.

While this study is based on destructive sampling, Collins and Buckley used a less invasive sampling method that came to be known as 'zooarchaeology by mass spectrometry' (ZooMS) and 'electrostatic zooarchaeology by mass spectrometry' (eZooMS). ZooMS describes the almost non-destructive identification of animal species on the basis of characteristic peptide sequences (peptide mass fingerprinting, PMF), which are obtained from collagen I by tryptic digestion (see Section 3.4, bottom-up approach). eZooMs includes non-invasive sampling with the help of a commercially available plastic eraser (see Section 3.3.1). In the research of parchment, this approach is therefore particularly suitable for a species identification of the animals used, since animal skins consist mainly of the protein collagen I, so that in the meantime several studies have been published in this context.[17]

While proteomics studies on the analysis of ancient manuscripts made from animal skins are comparatively advanced, research on other writing surfaces is still in its infancy. For example, manuscripts made from silk, also an animal

15 Fiddyment et al. 2015; Teasdale et al. 2021.
16 Toniolo et al. 2012.
17 Collins et al. 2010; Fiddyment et al. 2015; Fiddyment et al. 2021; Teasdale et al. 2021.

product that was primarily used in China, could be suitable for proteomics studies. Even if no studies have yet been published in the context of manuscripts. Nevertheless, degradation processes of silk have already been successfully investigated by means of proteomics analyses on cultural relicts made from ancient silk, and species identifications have been carried out on silk spinners in order to be able to trace the origin of ancient artefacts.[18]

In contrast, when using plant-based raw materials, it can be assumed that performing proteome analyses will be even more challenging. This is mainly due to the lower protein content of these materials and possibly the higher degree of processing, such as with paper, which makes it even more difficult to extract sufficient amounts of proteins and peptides from ancient manuscripts. Nevertheless, plant raw materials also have a small proportion of proteins and peptides that potentially still be detected as the sensitivity of MS instruments increases and depending on degradation processes. However, there is still too little knowledge in this regard to be able to make a more detailed assessment. Alternatively, genomics and metabolomics (in particular lipidomics) analyses should be more appropriate in this context.[19]

In addition, proteome-based analyses can also be used to estimate the age or the state of preservation of the manuscripts, whereby exogenous influences must also be taken into account (see Section 3.2). The age of ancient artefacts is usually estimated by means of radiocarbon dating, by comparing the decay of the radioactive carbon isotopes ^{14}C to ^{12}C. However, this method is comparatively laborious and requires a sample volume of 3-10 mg, i.e. it is a destructive sampling, which is why it cannot always be performed.[20] However, research efforts are underway to reduce the sample size required for isotope measurements to less than 50 µg.[21] In addition, it may also be useful to use several complementary approaches to verify the results obtained or to expose potential falsifications.[22]

Furthermore, proteomics analyses can provide information about surface treatments and, in the case of illuminated manuscripts, information about the colour binders and glues used, in order to better place the ancient manuscripts in their historical and also technical context or, if necessary, to be able to take

18 Chen et al. 2020; Gu et al. 2019; Li et al. 2019; Solazzo 2019.

19 Kostyukevich et al. 2018; Schulz et al. 2018.

20 Brock 2013.

21 Kasso et al. 2020.

22 Mazza 2019; Rabin and Hahn 2020.

appropriate conservation and restoration steps.[23] The protein-containing binders, which are suitable for the production of colours, include, e.g. eggs (egg yolk/egg white) from different birds, gelatin or collagen from various animals, and casein of milk. Using proteomics, it is not only possible to differentiate between individual protein binders as such, but also to characterise the raw materials used in more detail. For example, whether whole eggs or only the yolk or white were used, since the proteome is composed differently depending on the raw material used. Such sub-differentiation is not possible by means of DNA analysis.[24]

While proteomics refers to the analysis of the proteome of an individual organelle, cells, tissues or an entire organism,[25] metaproteome analyses include proteome studies of other organisms. Originally, when the term metaproteomics was introduced in 2004, it primarily referred to microbial communities.[26] However, when analysing ancient artefacts, in addition to microbial proteome residues, the remains of many other organisms can often be detected, too. These foreign protein and peptide profiles can be used to obtain information not only about the manuscript itself, but also, for example, about a possible pest infestation, biodeterioration, and even about the author or the reader.[27] For instance, on the original manuscript of *Master and Margarita* by Mikhail Bulgakov, peptide biomarkers for the author's nephrotic syndrome could be found.[28] Other applications include the detection of *Mycobacterium tuberculosis* in the notebooks of the Russian writer Anton Chekov and the detection of plague pathogens (*Yersinia pestis*) in death registries of Milan from 1630. In the last-mentioned study, residues of vegetable proteins could also be proved, which allow conclusions to be drawn about the meals of the secretaries from this time.[29] Furthermore, a recent study using a medieval parchment document on childbirth detected human body fluids in addition to numerous nonhuman residues of egg, milk, legumes, cereals and honey, presumably used for medical reasons, suggesting that this manuscript was indeed used in childbirth.[30]

23 Van der Werf et al. 2017.
24 Calvano et al. 2020a; Dallongeville et al. 2013; Kuckova et al. 2004; van der Werf et al. 2017.
25 Lim and Elenitoba-Johnson 2004.
26 Rodriguez-Valera 2004.
27 Marvasi et al. 2019.
28 Zilberstein et al. 2016.
29 D'Amato et al. 2018a.
30 D'Amato et al. 2018b.

3 Basics of proteomics analyses of ancient manuscripts

3.1 Handling and laboratory work with ancient manuscripts for proteome analyses

Ancient manuscripts have a high cultural heritage and are unique, rare as well as irreplaceable. Therefore, they should be handled with care and only by appropriately trained experts. Manuscripts are best stored for longer periods in special archives that ensure optimal storage parameters and protection against fire and water damage and take precautions against theft. By monitoring temperature and humidity as well as ensuring adequate ventilation, pest and fungal infestation can also be avoided, and chemical degradation processes can be reduced. Ancient objects should also not be exposed to short-wave UV light, particularly rich in energy, because UV light accelerates the decay of organic materials, so that they turn yellow and brittle. In addition, inks, and colours fade. Dust and dirt should also be kept out of the archives, or the manuscripts should be protected in appropriate acid-free packaging. In the best-case scenario, laboratories that regularly work with manuscripts meet these requirements mentioned and have appropriately equipped rooms for storing valuable artefacts. Further details in this context can be found, among others, in ISO 11799,[31] which defines the requirements for archive rooms and packaging.

Concerning the practical implementation of proteome-based studies on ancient artefacts, there are currently no generally applicable guidelines, as this is a very young research discipline. Due to this gap, some helpful information have recently been published in this context, which should promote the process of establishing generally recognised standards.[32] In addition, recommendations for dealing with aDNA can serve as orientation, which also include the construction and installation of suitable laboratories.[33] As when working with aDNA, contamination of the samples must be avoided. Consequently, practical work with ancient materials should be carried out spatially separated from other

31 International organization for standardization (ISO) 11799 (2017), Information and documentation — Document storage requirements for archive and library materials, <https://www.iso.org/standard/63810.html> (accessed on 24 Jan. 2021).

32 Fiddyment et al. 2019.

33 Fulton and Shapiro 2019; Gilbert et al. 2005.

work, especially with modern samples. In the best case, a separate laminar flow workbench and/or a clean room is available for this purpose.[34]

To avoid contaminations brought from other laboratories, the work should be carried out in clean rooms firstly in the morning. Furthermore, appropriate protective clothing should also be worn (full bodysuit including shoes, hairnet, safety glasses, face mask and powder-free gloves). Gloves made of nitrile are suitable, but latex gloves should be avoided as they can contain high levels of protein. The same applies to clothing or jewellery made of wood, wool, rubber, silk, or leather. In contrast, synthetic materials are particularly recommended. It is advisable to wear two pairs of gloves on top of each other, of which only the upper pair is changed regularly. Chemicals should be used in the highest purity available. In addition, stock solutions must be portioned, and the aliquots regularly exchanged. The reuse of consumables should also be avoided. Likewise, only pipette tips with filters should be used. Bleach solution and subsequently 70% ethanol are suitable for regular cleaning of all work surfaces and equipment. As with DNA studies, irradiation of surfaces and materials can help denature proteins to ensure the cleanest possible work environment.[35]

3.2 Dealing with contaminations

In addition to proteins and peptides from the material of the writing medium, signals can also come from bacteria, viruses, fungi, insects, leftovers of the readers or authors, or manual handling. They do not always have to be contaminations in the negative sense but can be used for metaproteomics analyses to obtain further information about the manuscript (see Section 2). Nevertheless, dealing with unwanted contaminant is not easy, e. g. if it originates from recent conservation and restoration measures, was created during the recovery of historical documents, or an incorrect handling, e. g. if gloves were not worn. In the best case, such manipulations are known and can be taken into account accordingly in order to avoid misinterpretations.

A comparatively reliable method to distinguish ancient from modern age proteins and peptides allows the degree of the post-translational modifications (PTMs) e. g. deamidation from glutamine to glutamic acid and asparagine to aspartic acid with the loss of the ammonium group. Glutamine deamidation is preferred for age estimation because deamination of asparagine is about 10

34 Fiddyment et al. 2019.

35 Fiddyment et al. 2019; Fulton and Shapiro 2019; Gilbert et al. 2005; Llamas et al. 2017.

times faster compared to glutamine. Nevertheless, in addition to age, other exogenous factors such as the storage parameters have an impact on the degree of deamination, which may have to be taken into account, so direct correlations are not always possible.[36] Another way to estimate age and protein degradation is to study amino acid racemisation (AAR), which is the conversion of the L-configuration of amino acids commonly found in living organisms to D-enantiomers. There are a number of ways to detect D-amino acids, for example by chromatographic separation on chiral stationary phases or by converting the amino acids into diasteromers, which can then also be separated on achiral phases. In addition, enzymatic detection methods, for example, are also suitable. Nevertheless, even with these methods, the influences of the environment during the storage of the documents must be taken into account.[37] In addition to the age-dependent modifications explained, other reactions can also provide information about the age of proteins, such as hydroxylation in collagen-containing matrices.[38] Furthermore, additional information can be obtained from the fragmentation pattern of certain proteins and peptides. For this purpose, the systematic analysis of samples with known ageing grades is suitable, which can be used as references and in the determination of further molecular markers for a comparison.[39] Furthermore, common contaminants, e.g., such as human creatine or human serum albumin, can be identified with the help of databases such as the 'common Repository of Adventitious Proteins' (cRAP, https://www.thegpm.org/crap/)[40] or the MaxQuant database.[41]

During the laboratory work, extraction blanks should also be processed and then measured, in order to be able to take into account the purity of the chemicals and contamination if necessary. In addition, samples and protein standards should also be stored separately from one another in order not to risk cross contamination.

3.3 Sampling strategies

Sampling should be as gentle and minimally invasive as possible, at best non-destructive, without contaminating the manuscript. In addition, it must be

36 Leo et al. 2011; Ramsøe et al. 2020; Schroeter and Cleland 2015.
37 Carenzi et al. 2020; Collins et al. 2006; Rosini et al. 2020.
38 Cleland et al. 2015; Hill et al. 2015.
39 Li et al. 2015.
40 Mellacheruvu et al. 2013.
41 Cox and Mann 2008.

ensured that the analytes are not subject to changes that could, in the worst case, lead to false conclusions. The location of the sampling on the manuscript is also crucial, as various factors can have an influence on the preservation of proteins and peptides.

Starting from destructive micro-sampling strategies, in which a few milli- or micrograms of materials are taken from the analysis object using tools such as scalpels and tweezers, the focus in the last five years has been on the development of non-invasive or almost non-invasive methods. The criticism that non-destructive methods are relatively insensitive is increasingly being eclipsed by new mass spectrometric developments with ever-improving sensitivity.

3.3.1 On-Site Sampling Strategies

On-site sampling strategies have the advantage that the historical documents do not have to be transported to a laboratory, i.e. sampling can also be done directly in libraries, museums, archives or private collections, which greatly facilitates access to the samples. In addition, these methods are relatively easy to perform and require only a minimal labour. Furthermore, the documents can be of any size and do not need to be cut up (Fig. 2).

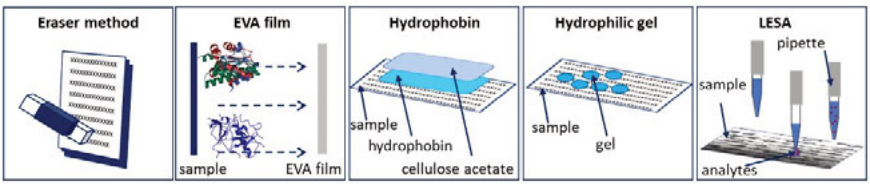

Fig. 2: Overview of different on-site sampling strategies that are suitable for manuscripts.

The first almost non-invasive on-site sampling method was described by the group of Matthew Collins, in which analytes are detached from the surface of different materials by means of a simple plastic eraser made of polyvinyl chloride (PVC) using the triboelectric effect. In this procedure, the analytes are picked up by the eraser through careful rubbing. The resulting polymer residues, with the adhering peptides, can presumably be stored at room temperature for any length of time without changes occurring. At an appropriate time, eraser crumb samples are mixed with a trypsin-containing solution for enzymatic

digestion (see Section 3.4) so that the peptides obtained pass into solution and can be analysed.[42]

Another non-destructive alternative is the use of a special ethyl-vinyl acetate (EVA) film made from C8 and/or C18 resins as well as cations/anion exchangers to which the molecules are bound. The film is produced using an extruder, in a thickness of 150–200 µm. Before use, the film is moistened and then placed on the document for several minutes. In addition to macromolecules such as proteins and DNA, peptides as well as metabolites are also absorbed. Superficially, non-covalent bonds such as polar ion-ion interactions and hydrogen bonds as well as hydrophobic van der Waals forces are effective. Non-covalent bonds such as polar ion-ion interactions and hydrogen bonds as well as hydrophobic van der Waals forces are the primary agents. Depending on the material and preferred analyte class, the EVA technology offers different modifications by adding various chemical additives to the starting material so that an optimal result can be achieved. Compared to an invasive method on wall paintings and canvases for the detection of binders, qualitatively fewer proteins were detected, but the binders used could be identified on the basis of the remaining signals. In addition, studies on the surface of parchment, for example, have shown that no damage occurs, or residues remain.[43]

An alternative option for sampling is the use of a class I hydrophobin (Vmh2) of the fungus *Pleurotus ostreatus*. With hydrophobin Vmh2 as mediator, *in situ* digestion with trypsin can also be performed. Hydrophobin Vmh2, which is initially in an ethanolic solution, is spotted onto small plates made of cellulose acetate. After the alcohol has evaporated, trypsin is added, which is immobilised on hydrophobin Vmh2. Shortly before use, the plates are moistened with an aqueous buffer solution so that the optimal ambient conditions for the activation of trypsin are present. The cellulose acetate plates are placed on the sample material for approx. 10 min and after drying, the peptides can simply be washed off and analysed. This method was originally developed for the detection of binders in paintings. In the meantime, however, several metaproteomics analyses have been carried out on paper and on parchment documents.[44]

Another method, also suitable for *in situ* digestion with trypsin, is based on the use of a hydrophilic gel of poly(2-hydroxyethyl methacrylate)/poly(vinyl-

42 Fiddyment et al. 2015.

43 D'Amato et al. 2018a; D'Amato et al. 2018b; Manfredi et al. 2017; Righetti et al. 2019, Righetti et al. 2020; Zilberstein et al. 2016; Zilberstein et al. 2020.

44 Cicatiello et al. 2018; Ntasi et al. 2021.

pyrrolidone) (pHEMA/PVP) loaded with trypsin before use. So far, this recently introduced method has not been applied to manuscripts. The analyses carried out concentrated primarily on work of arts and statues, but a transfer to other materials is quite conceivable.[45]

A very simple procedure that can be carried out with standard chemicals, has recently been proposed for the authentication of historical manuscripts. Based on a liquid extraction surface analysis (LESA) approach, 2 µL of a mixture of methanol, water and a small amount of formic acid are added to the manuscript using a pipette. After pulling up and emptying the pipette several times, the solvent is analysed by MS. The focus of this procedure was primarily the analysis of small molecules, but it can be assumed that this approach could also potentially be suitable for the detection of peptides for metaproteomics interpretations.[46]

Compared to the method with the eraser, the other methods require somewhat more time, as the absorbent material must be in contact with the objects for longer. However, a comparison of which of the presented methods is best suited for different materials and issues is still pending.

3.3.2 Sampling in the laboratory: Ambient ionisation and laser ablation techniques

Ambient ionisation techniques, which can be coupled using different mass analysers, are usually carried out in appropriately equipped laboratories, i.e. the artefact must first be transported there. In ambient ionisation sources, the ions are formed outside the MS device. The main advantage is that no sample preparation is required, and sampling is practically non-destructive. Furthermore, these developments usually can also be used to generate spatially resolved chemical images, which is why they are also referred to as mass spectrometry imaging (MSI). For some techniques, however, it should be noted that the object to be analysed must not be too large. In addition, the analysis of larger peptides and proteins with ambient ionisation methods is not that simple, due to their molecular weight, the poor desorption efficiency and the surrounding matrix.[47]

45 Calvano et al. 2020a; Calvano et al. 2020b.
46 Newton et al. 2018.
47 Douglass and Venter 2013.

Using relatively established desorption electrospray ionisation mass spectrometry (DESI-MS), the surface of an artefact is exposed to an electrospray cloud (see Section 3.5) so that the analytes are directly desorbed and ionised. The optimisation of various parameters such as the distance or the angle of the electrospray and the MS inlet is decisive in determining which analyte classes are detected.[48] So far, this method has been used, for example, to detect proteins (app. 20 kDa) and peptides after *in situ* digestion on flint flakes and potsherd samples, but the analysis of historical manuscripts has also been possible.[49] By means of a nanospray DESI (nano-DESI) source, the spatial resolution can also be reduced from approx. 40 µm to 10 µm and splashing effects can be avoided, so that destructive processes are minimised.[50]

Another modification of an electrospray ionisation (ESI) source is the laser ablation electrospray ionisation (LAESI) procedure, in which the sample first has to be slightly moistened. Subsequently, the analytes are ablated from the surface by means of an infrared (IR) laser, in which initially mainly the water molecules evaporate abruptly. During this process, the analytes are carried along into the resulting ablation plume and can then be ionised using an electrospray, consequently, this is a two-stage process.[51] It is also conceivable to replace the usual nanosecond IR laser with a more efficient picosecond IR laser (PIRL) in order to be able to detect larger proteins efficiently while damaging the surrounding analysis material as little as possible, since the lateral resolution (~100µm) is 3-fold smaller than by means of a nanosecond laser.[52] It is also possible to decouple the laser from the ionisation and ablate the samples on site. The collected ablation plume can later be analysed by MS in the laboratory. Such technologies are already used in proteome analysis for medical questions, e.g. for the identification of different tissue types, but have hardly been validated in studies on ancient artefacts. However, these approaches could become increasingly relevant in the near future and as technology develops.

In addition to the techniques described above, there are numerous other research efforts and ambient ionisation techniques e.g. LESA, Matrix-assisted laser desorption electrospray ionisation (MALDESI) or time-of-flight-secondary ion mass spectrometry (SIMS-ToF) that could have the potential to become more

48 Takáts et al. 2005.

49 Heaton et al. 2009; Newton et al. 2018; Schedl et al. 2015; Stephens et al. 2010.

50 Roach et al. 2010.

51 Nemes and Vertes 2007; Stephens et al. 2010.

52 Zou et al. 2015.

relevant for the MS analysis of peptides and proteins of historical documents in the future.[53]

3.4 Mass spectrometric strategies for proteomics analysis

The primary goal of proteome and metaproteome analyses of ancient artefacts is to identify proteins and peptides. This is mainly achieved by clarifying the primary structure, i.e. the amino acid sequence or fragments thereof, and the subsequent comparison and assignment using databases (see Section 3.6). Sequencing of amino acid sequences was traditionally carried out using Edman's method.[54] However, this procedure is both very complex and requires the presence of isolated analytes. Therefore, this procedure is not suitable for the analysis of more complex extracts and, moreover, is not capable of high throughput. Other approaches are based on immunoassays to bind certain sequences, but these are very limiting for qualitative approaches and also not very sensitive.[55] For this reason, MS methods have primarily become established. There are mainly two different approaches:

In the bottom-up approach, proteins and peptides are first digested enzymatically. Usually, the enzyme trypsin is used as the gold standard, which hydrolytically cleaves the amino acids arginine and lysine at the C-terminus. The peptides obtained have a length of 7–20 amino acids and molecular masses of 0.7 kDa–3 kDa. However, the disadvantage of using trypsin is that this enzyme is sometimes too efficient and about 56% of the peptide fragments obtained are too small for further data evaluation. Hence, it may be helpful to try other endopeptidases as well.[56]

The peptide fragments are then separated one-dimensionally or two-dimensionally by liquid chromatography (LC).[57] In contrast to a direct infusion mass spectrometry (DIMS) approach, not all analytes reach the ion source at the same time, but rather with a time delay. In addition to gaining information about the retention time, ion suppression effects are reduced so that a greater number of molecules can be detected overall. This advantage is of particular

53 Cleland and Schroeter 2018; Feider et al. 2019; Spraker et al. 2020.

54 Edman 1949.

55 Cartechini et al. 2010; Hendy 2021; Manfredi et al. 2017; Palmieri et al. 2013; Sciutto et al. 2016.

56 Cristobal et al. 2017; Pandeswari and Sabareesh 2019; Swaney et al. 2010; Tsiatsiani and Heck 2015.

57 Delmotte et al. 2007.

relevance for the detection of minor compounds and not only abundant analytes. In most studies, LC separation is performed by reverse phase (RP) chromatography utilising hydrophobic interactions. In this context, water and acetonitrile are the most suitable eluents. A small addition of formic acid to the eluent can be helpful in order to achieve a positive net charge of the peptides, which in turn has a positive effect on the ionisation. In addition, hydrophilic interactions with potential silanol residues of the stationary phase are suppressed, which in turn leads to greater retention and better peak shapes. However, the use of formic acid carries the risk of causing undesired formylation reactions on the peptides. Therefore, the amount of formic acid should not be too high.[58] In addition to the use of stationary RP, other LC separation principles are also suitable for the analysis of peptides. These include the use of hydrophilic interaction liquid chromatography (HILIC), ion exchange chromatography, or size exclusion chromatography, although with the latter two it must be taken into account that high salt concentrations must be used which are not compatible with the MS instruments, which is why they are of subordinate relevance.[59]

High performance liquid chromatography (HPLC), ultra-high performance liquid chromatography (UHPLC) as well as nano-LC devices are suitable for chromatographic separation. Compared to HPLC systems, significantly higher back pressures can be generated using HPLC systems (up to approx. 1,300 bar). Therefore, by means of UHPLC, stationary phases with smaller particle sizes <3 µm can be used, which results in improved resolution, sensitivity and shorter analysis times. In addition, nano-LC devices are increasingly being used for proteomics studies. The main advantage is improved sensitivity, as more intense, albeit narrower, peaks are produced. This means that low-abundant peptides can still be recorded. Therefore, nano-LC devices are particularly recommended for studies on ancient artefacts, when often only sample material and proteome residues are available. With particularly sensitive methods, detection limits in the low femtomol and sometimes even attomol range can be achieved, whereby the selected mass analyser also has a major influence.[60]

Nano-columns typically have an inner diameter of 50–100 µm and particle size of 1.4–5 µm. They are operated at flow rates of about 100–500 nL/min. The comparatively low flow rates imply lower ion suppression in the subsequent ionisation step, which in turn is also reflected in an increased sensitivity. De-

58 Lenčo et al. 2020.

59 Badgett et al. 2018; Mant et al. 2007.

60 Ivanov et al. 2003; Martin et al. 2000.

pending on the manufacturer, nano-LC systems can withstand back pressures of up to 1,000 bar. Unlike UHPLC or HPLC systems, it is not possible to apply the sample extract directly to the nano-column, since the low flow rates would require a large amount of time for the extract to reach the analytical column. Therefore, the extract is first added to a trap column using a microcharge pump, which allows higher flow rates. This step also allows sample enrichment and purification. Water with little amount of trifluoroacetic acid (TFA) is suitable for loading the trap column in order to bind the peptides to the trap column first, while the salts are rinsed away. TFA is a better ion pair reagent than, for example, formic acid, which is why the analytes bind more strongly to the stationary phase. Subsequently, the trap column is switched into the nano-LC system, using water and acetonitrile together with formic acid as additive as described above, since TFA can lead to strong suppression effects during ionisation. By increasing the acetonitrile content, the peptides then elute gradually from the analytical nano-column. As an alternative to trap columns, injection loops with fixed volumes can also be used, but these are only suitable for small sample volumes and can lead to wide peak shapes. In addition, the loops do not allow an online desalting step.[61] In addition to the 'classic' columns described, chip-based systems are also suitable for achieving purification and separation of the smallest amounts of analytes. Such chips can be replaced very easily using a 'plug and play' procedure and require almost no manual effort. Chip systems can now be purchased commercially from various suppliers. In addition, a number of developments can be expected in this area over the next few years.[62]

The procedure described is also known as shotgun approach, which has become well established in recent years. The shotgun approach differs from procedures in which proteins are first separated using two-dimensional polyacrylamide gel electrophoresis (2D page) before tryptic digestion takes place. The main advantage of the shotgun strategy is that many proteins and peptides can be identified simultaneously with relatively little effort. A disadvantage, however, is that due to the fragmentation PTMs, protein truncations and alternative splicing events of eukaryotic organisms cannot always be taken into account, which on the one hand leads to a limited sequence coverage and on the other hand to a significant loss of information. Despite this drawback, most scientists prefer the bottom-up approach rather than the top-down approach presented below.[63]

61 Noga et al. 2007; Wilson et al. 2015.

62 Vargas Medina et al. 2020.

63 Aebersold and Mann 2003; Dupree et al. 2020; Zhang et al. 2013.

In contrast to the bottom-up approach, the top-down approach aims to analyse intact proteins. Accordingly, there is no enzymatic digestion. After isolating the proteins, which can already be challenging because proteins have lower solubilities than peptides, a chromatographic separation is also carried out either by electrophoresis and/or by LC. However, the separation and purification of proteins is significantly more complex than the separation of peptides and a comparatively large part of the analytes is already lost in this process. In addition, the comparatively low utilisation of top-down strategies is due to the limited size of proteins that can be detected by MS detectors with sufficient sensitivity and resolution.[64] Usually, only proteins <30 kDa are detected.[65] However, a lot of research is currently being carried out on both disadvantages in order to overcome these drawbacks in the future.[66]

In some cases, well-preserved and larger proteins can be detected in ancient fossils, as the surrounding biominerals of bones or teeth have good preservation properties. For example, in 2,000-year-old ancient bone, approx. 30 kDa proteins could be detected and proteins that were still in good condition could also be recorded in the brain tissue of the Tyrolean Iceman 'Ötzi'.[67] For such samples, the use of a top-down approach may be useful. However, the extent to which a top-down approach can provide additional information for proteome and metaproteome analysis of ancient manuscripts has not yet been investigated. Nevertheless, a use of top-down approaches would be conceivable to analyse the chemical composition of comparatively stable proteins such as collagen in leather or parchment in more detail. Further possible applications include studies on protein survival and degradation.[68] Also, whether enzymatic digestion can be dispensed with and whether it is sufficient to analyse only the undigested fragments that arise solely due to the ageing process has not been adequately researched and certainly offers potential for further investigations. However, it must be noted that the last option mentioned is not a top-down approach in the classic sense.

To overcome the disadvantages of bottom-up approaches, which are mainly based on a low sequence coverage, and of top-down approaches, which are more associated with technical limitations, new considerations deal with middle-down approaches as a further strategy. This middle way, although not yet

64 Chen et al., 2018; Padula et al. 2017.
65 Fornelli et al. 2018.
66 Shin et al. 2018; Toby et al. 2016.
67 Bona et al. 2014; Maixner et al. 2013.
68 Hendy et al. 2021.

widely used, involves enzymatic digestion with special proteases that are less efficient than trypsin, yielding longer peptide fragments of about 20-100 amino acids that have a molecular mass of 2.5 kDa-10 kDa. Because larger fragments are obtained overall, both the number of peptides and the complexity of the sample extracts decrease. As a result, the requirements for chromatographic separation are reduced, too. At the same time, more unique fragments are formed so that the sequence coverage increases and PTMs as well as proteoforms can be analysed more easily.[69] So far, the middle-down strategy has been used to a very limited extent. Nevertheless, there could also be great potential here for the proteomics analysis of ancient documents.

3.5 Technology requirements for proteomics analyses using mass spectrometry

Since the development of the first MS instruments by Aston and Thomson at the beginning of the twentieth century, a wide variety of techniques and designs have been developed for various scientific issues and applications, and yet they are all based on a similar principle: (i) The desorption of the analytes into the gas phase including their ionisation, (ii) the separation of the ions according to their mass-to-charge ratio (*m/z*) and (iii) the detection.[70] Nevertheless, the different types are more or less suitable for the analysis of proteins and peptides; in the following sections we will briefly discuss the most important ones.

Some potential ion sources for the desorption and ionisation of the analytes have already been explained in Section 3.3.2. However, the most commonly used ion sources for the detection of proteins and peptides are ESI and matrix-assisted laser desorption/ionisation (MALDI). Both techniques are soft ionisation sources in which the analytes fragment only weakly, if at all. MALDI is based on the use of UV-absorbing substances, the so-called matrix, which are mixed with the analytes. The matrix must be present in a large excess so that a solid matrix crystal is formed by co-crystallisation. Subsequently, the matrix is then ablated using a pulsed UV laser, and the analytes are entrained and ionised, although the exact ionisation process is not yet fully understood. However, different theories exist. MALDI predominantly leads to the formation of singly charged adducts such as, for example, $[M+H]^+$, $[M+Na]^+$ or $[M+K]^+$.[71]

69 Cristobal et al. 2017; Pandeswari and Sabareesh 2019.

70 Smoluch and Silberring 2019.

71 Dreisewerd 2003; Karas and Hillenkamp 1988.

Compared to MALDI, online couplings with LC principles are possible using ESI. The eluate of the LC is sprayed under atmospheric pressure by means of nitrogen within an electric field. As the solvent evaporates, the charged analytes accumulate on the surface of the individual droplets. Due to the high charge density, the individual droplets disintegrate like an explosion (Coulomb explosion) until only the charged analytes are present and passed into the MS.[72] Characteristic of the ESI process is the formation of pseudo ($[M+H]^+$, $[M+H]^-$) and adduct molecules ($[M+Na]^+$, $[M+NH_4]^+$ etc.) as well as the induction of multiple charges ($[M+H]^{2+}$, $[M+3H]^{3+}$etc.), especially with higher mass molecules like proteins. With certain additives such as weak acids or salts to the eluents, the ionisation processes can be influenced and thus possibly higher signal intensities can be achieved.[73] Nanospray-ESI sources enable the formation of a stable spray and maximum sensitivity even at low flow rates, which is particularly relevant when only small amounts of the analytes are available.[74] Since MALDI and ESI are based on different ionisation processes and are complementary to each other, it is recommended to use both techniques in order to record the maximum information content from the proteome of a sample.[75]

To determine the *m/z* ratios of proteins and peptides, high-resolution analysers such as time-of-flight (ToF), orbitrap or fourier-transform ion-cyclotron-resonance (FT-ICR) analysers are primarily used. The various mass analysers differ mainly in resolution, accuracy, measuring range, scan rate and price.[76] ToF analysers are combined with both ESI and MALDI ion sources. They are characterised by fast scan rates and are therefore suitable for fast LC separations. In addition, very large *m/z*-ratios can still be analysed with them. Compared to ToF analysers, orbitraps are comparatively compact, but do not allow as fast scan rates. In addition, the mass range is limited to about *m/z* 6,000. However, they have a significantly better resolution and MS^n spectra can be recorded. The highest mass resolution is achieved with FT-ICR devices. They are also the most expensive both to purchase and to maintain, as they are operated with superconducting magnets that are cooled with liquid nitrogen and helium. However, scan rates are the lowest compared to the other two analysers. Usually, different analysers are coupled with each other so that MS/MS or MS^n experiments are possible.

72 Yamashita and Fenn 1984.

73 García 2005; Leitner et al. 2007; Nshanian et al. 2018.

74 Karas et al. 2000; Wilm and Mann 1994.

75 Nadler et al. 2017.

76 Creydt and Fischer 2020.

In recent years, the market leaders have also equipped some of their LC-ESI-QToF devices with ion mobility spectroscopy (IMS) cells. Depending on the device manufacturer (Agilent, Bruker or Waters), different designs are available, all with the aim of introducing an additional orthogonal separation in order to be able to distinguish compounds that have the same (isobars) or a very similar *m/z* ratios. The ion mobility of a molecule is influenced by the mass and the charge but also by its size and shape (collision cross section, CCS). Especially in proteomics studies, this technology offers a high added value, both at the protein level, to distinguish structural conformers and especially at the peptide level in shotgun approaches to increase the sequence coverage, as isobar fragments can be separated better and background noise is reduced, which leads to an improved signal-to-noise ratio. Furthermore, this technique can also be very helpful in identifying PTMs.[77] Even if, to our knowledge, no comprehensive studies using IMS devices have yet been carried out on ancient artefacts, it can be assumed that this technology will have a high added value in future studies.

3.6 Data evaluation

The identification of peptides and their associated proteins in proteomics and metaproteomics experiments is often challenging and can be very time-consuming. In the simplest case, with PMF approaches, which are usually carried out with MALDI-ToF devices, a bottom-up experiment is carried out and, after enzymatic digestion, the peptides obtained are measured by MS in full-scan mode (Fig. 3A). In this way, no sequence data are obtained, but a fingerprint that is dependent on both the peptide or protein and the enzyme. The protein can then be identified on the basis of a comparison of the peak list obtained with the corresponding sequence databases. The ZooMS method already presented is based on this procedure (see Section 2). The prerequisite for this simple and quick procedure is, on the one hand, that there are as few impurities or PTMs as possible and, on the other hand, that the corresponding sequences are available in the protein or genome databases used, such as Swiss-Prot or NCBI. Suitable search programs are, for example Mascot,[78] MS-FiT,[79] ProFound[80]

77 Dodds and Baker 2019; Winter et al. 2019.

78 Perkins et al. 1999.

79 https://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard (accessed on 24 Jan. 2021).

80 Zhang and Chait 2000.

or PeptIdent.[81] Sometimes it can be helpful to use different programs and databases, as the tools rely on different algorithms and quality parameters for the calculations.[82]

If PMF is not sufficient for reliable identification, e. g. because the extract is not pure enough, or if the peptide sequence is to be determined more precisely, peptide fragment fingerprinting (PFF) approaches are particularly suitable (Fig. 3B).

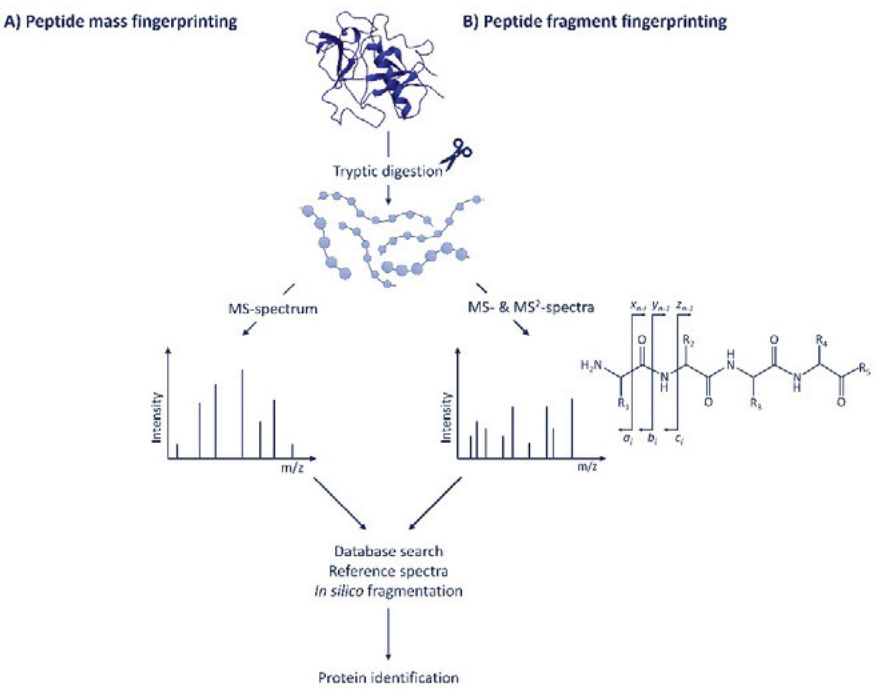

Fig. 3: Workflow of A) PMF and B) PFF. While only MS spectra are recorded for PMF experiments, PFF studies enable additional structural information by generating MS^2 or MS^n spectra.

For this purpose, the peptides are subjected to mass spectrometric fragmentation after enzymatic digestion, so that even smaller fragment ions are

81 NCSC US, PeptIdent <http://www.pdg.cnb.uam.es/cursos/BioInfo2004/pages/visualizacion/programas_manuales/spdbv_userguide/us.expasy.org/tools/peptident.html> (accessed on 24 Jan. 2021).

82 Damodaran et al. 2007; Dupree et al. 2020; Henzel et al. 2003; Zengin et al. 2017.

obtained. In this way, structural information for a database comparison is obtained. The Mascot software can also be used for this step. In addition to many other programs, SEQUEST[83] is a helpful tool in this context, especially when spectra with low signal-to-noise ratios are available. PFF experiments are also suitable for *de novo* sequencing, which is relevant if peptides are not registered in databases, and PTMs can be more easily traced.[84]

In addition to the procedures explained, as in many other research areas, bioinformatic methods for pattern recognition are becoming increasingly relevant.[85] These have the advantage, on the one hand, that the data can be evaluated automatically and, on the other hand, that more signals and multivariate relationships can be taken into account so that sub-differentiations can be made between the various sample groups. Overall, more signals can be used in this way than with a pure database comparison, which assumes that the signals are already known. At the same time, relevant marker signals can be extracted, the structure of which can be particularly relevant for further interpretations. While such approaches have so far hardly been pursued in manuscript research using proteomics, the work on other ancient artefacts or with other technological platforms is already more advanced.[86] This procedure assumes that reference samples are measured using the same method, but the larger such a database becomes, the more information can be made available. It can be assumed that some developments in this area can be expected in the next few years.

4 Conclusions

Although paleoprotomics is a comparatively young discipline, numerous breakthroughs have already been made in manuscript research. Above all, the progress from destructive sampling to non-invasive procedures is of great value and enables research in this area to be progressed increasingly, as the acquisition of samples as a whole is made significantly easier. Further future developments on the part of MS and also chromatographic methods will make it possible to measure with ever greater sensitivity, so that details of the proteome can be recorded better and better. In this regard, we see great potential in the establishment of IMS technologies. Some successes could already be recorded

83 Eng et al. 1994.

84 Dupree et al. 2020; Na and Paek 2020; Zengin et al. 2017.

85 Creydt and Fischer 2020.

86 Alvarez et al. 2019; Bacci et al. 2001; Gu and Buckley 2018; Navas et al. 2008.

on the data evaluation side. Nevertheless, there are still some challenges to be mastered in this area in order to achieve a greater coverage of the sequences and improve the database research. Furthermore, with the implementation of bioinformatic methods it should be possible to obtain a large amount of additional information via a manuscript, since signals that have not yet been interpreted can be explained better. There is further potential in expanding the currently predominantly applied bottom-up approaches to middle-down and top-down strategies, since in this way less information could be lost.

Overall, the proteomics and metaproteomics analysis of ancient manuscripts offers a great deal of added information, also in connection with other, comparatively more established methods. Based on the successes already achieved, further progress can be expected in the next few years, even if there is still much to be done.

Acknowledgements

This publication was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2176 'Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures', project no. 390893796 and was conducted within the scope of the Centre for the Study of Manuscript Cultures (CSMC) at Universität Hamburg.

References

Abelson, Philip (1954a), 'Paleobiochemistry: organic constituents of fossils', Carnegie Institution of Washington, *Yearbook*, 53: 97–101.

Abelson, Philip (1954b), 'Amino acids in fossils', *Science*, 119/3096: 576.

Aebersold, Ruedi and Matthias Mann (2003), 'Mass spectrometry-based proteomics', *Nature*, 422/6928: 198–207.

Alvarez, Angel Martin Fernandez, Julie Bouhy, Marc Dieu, Catherine Charles and Olivier Deparis (2019), 'Animal species identification in parchments by light', *Scientific Reports*, 9/1: 1825.

Bacci, Mauro, Martha Fabbri, Marcello Picollo and Simone Porcinai (2001), 'Non-invasive fibre optic fourier transform-infrared reflectance spectroscopy on painted layers: Identification of materials by means of principal component analysis and mahalanobis distance', *Analytica Chimica Acta*, 446/1: 15–21.

Badgett, Majors, Barry Boyes and Ron Orlando (2018), 'Peptide retention prediction using hydrophilic interaction liquid chromatography coupled to mass spectrometry', *Journal of Chromatography A*, 1537: 58–65.

Bona, Agnes, Zoltan Papai, Gabor Maasz, et al. (2014), 'Mass spectrometric identification of ancient proteins as potential molecular biomarkers for a 2000-year-old osteogenic sarcoma', *Plos One*, 9/1: e87215.

Brock, Fiona (2013), 'Radiocarbon dating of historical parchments', *Radiocarbon*, 55/2: 353–363.

Calvano, Cosima Damiana, Elena Rigante, Rosaria Picca, Tommaso Cataldi and Luigia Sabbatini (2020a), 'An easily transferable protocol for in-situ quasi-non-invasive analysis of protein binders in works of art', *Talanta*, 215: 120882.

Calvano, Cosima Damiana, Elena Rigante, Tommaso Cataldi and Luigia Sabbatini (2020b), 'In situ hydrogel extraction with dual-enzyme digestion of proteinaceous binders: the key for reliable mass spectrometry investigations of artworks', *Analytical Chemistry*, 92/15: 10257–10261.

Cappellini, Enrico, Frido Welker, Luca Pandolfi et al. (2019), 'Early pleistocene enamel proteome from dmanisi resolves stephanorhinus phylogeny', *Nature*, 574/7776: 103–107.

Carenzi, Giacomo, Silvia Sacchi, Monica Abbondi and Loredano Pollegioni (2020), 'Direct chromatographic methods for enantioresolution of amino acids: recent developments', *Amino Acids*, 52/6: 849–862.

Cartechini, Laura, Manuela Vagnini, Melissa Palmieri et al. (2010), 'Immunodetection of proteins in ancient paint media', *Accounts of Chemical Research* (American Chemical Society), 43/6: 867–876.

Chen, Bifan, Kyle Brown, Ziqing Lin and Ying Ge (2018), 'Top-down proteomics: Ready for prime time?', *Analytical Chemistry*, 90/1: 110–127.

Chen, Ruru, Mingzhou, Hu, Hailing Zheng et al. (2020), 'Proteomics and immunology provide insight into the degradation mechanism of historic and artificially aged silk', Analytical Chemistry, 92/3: 2435–2442.

Cicatiello, Paola, Georgia Ntasi, Manuela Rossi et al. (2018), 'Minimally invasive and portable method for the identification of proteins in ancient paintings', *Analytical Chemistry*, 90/17: 10128–10133.

Cleland, Timothy and Elena Schroeter (2018), 'A comparison of common mass spectrometry approaches for paleoproteomics', *Journal of Proteome Research*, 17/3: 936–945.

Cleland, Timothy, Elena Schroeter and Mary Schweitzer (2015), 'Biologically and diagenetically derived peptide modifications in moa collagens', *Proceedings of the Royal Society B: Biological Sciences*, 282/1808: 20150015.

Collins, Matthew, Mike Buckley, Helen Grundy et al. (2010), 'Zoo MS: The collagen barcode and fingerprints', *Spectroscopy Europe*, 22/2: 6–10.

Collins, Matthew, Enrico Cappellini, Michael Buckley et al. (2006), 'Analytical methods to detect ancient proteins', in Jan Gunneweg, Charles Greenblatt and Anemie Adriaens (eds), *Bio-and Material Cultures at Qumran*, Stuttgart: Fraunhofer IRB, 33–40.

Colomo, Daniela (2012), 'Writing materials', in Roger Bagnall, Kai Brodersen, Andrew Erskine and Sabine Huebner (eds), *The Encyclopedia of Ancient History*, Wiley Online Library: <https://doi.org/10.1002/9781444338386.wbeah07109>.

Cox, Jürgen and Matthias Mann (2008), 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*, 26/12: 1367–1372.

Creydt, Marina and Markus Fischer (2018), 'Omics approaches for food authentication', *Electrophoresis*, 39/13: 1569–1581.

Creydt, Marina and Markus Fischer (2020), 'Food phenotyping: Recording and processing of non-targeted liquid chromatography mass spectrometry data for verifying food authenticity', *Molecules*, 25/17: 3972.

Cristobal, Alba, Fabio Marino, Harm Post, et al. (2017), 'Toward an optimized workflow for middle-down proteomics', *Analytical Chemistry*, 89/6: 3318–3325.

D'Amato, Alfonsina, Gleb Zilberstein, Svetlana Zilberstein, Benedetto Compagnoni and Pier Giorgio Righetti (2018a), 'Of mice and men: Traces of life in the death registries of the 1630 plague in Milano', *Journal of Proteomics*, 180: 128–137.

D'Amato, Alfonsina, Gleb Zilberstein, Svetlana Zilberstein et al. (2018b), 'Anton Chekhov and Robert Koch cheek to cheek: A proteomic study', *Proteomics*, 18/9: 1700447.

Dallongeville, Sophie, Nicolas Garnier, Christian Rolando and Caroline Tokarski (2016), 'Proteins in art, archaeology, and paleontology: From detection to identification', *Chemical Reviews*, 116/1: 2–79.

Dallongeville, Sophie, Mark Richter, Stephan Schäfer et al. (2013), 'Proteomics applied to the authentication of fish glue: Application to a 17th century artwork sample', *The Analyst*, 138/18: 5357–5364.

Damodaran, Senthilkumar, Troy Wood, Priyadharsini Nagarajan and Richard Rabin (2007), 'Evaluating peptide mass fingerprinting-based protein identification', *Genomics, proteomics & bioinformatics*, 5/3–4: 152–157.

Delmotte, Nathanaël, Maria Lasaosa, Andreas Tholey, Elmar Heinzle and Christian Huber (2007), 'Two-dimensional reversed-phase × ion-pair reversed-phase HPLC: An alternative approach to high-resolution peptide separation for shotgun proteome analysis', *Journal of Proteome Research*, 6/11: 4363–4373.

Dettmer, Katja, Pavel Aronov and Bruce Hammock (2007), 'Mass spectrometry-based metabolomics', *Mass Spectrometry Reviews*, 26/1: 51–78.

Dodds, James and Erin Baker (2019), 'Ion mobility spectrometry: Fundamental concepts, instrumentation, applications, and the road ahead', *Journal of The American Society for Mass Spectrometry*, 30/11: 2185–2195.

Douglass, Kevin and Andre Venter (2013), 'Protein analysis by desorption electrospray ionization mass spectrometry and related methods', *Journal of Mass Spectrometry*, 48/5: 553–560.

Dreisewerd, Klaus (2003), 'The desorption process in MALDI', *Chemical Reviews*, 103/2: 395–426.

Dupree, Emmalyn, Madhuri Jayathirtha, Hannah Yorkey et al. (2020), 'A critical review of bottom-up proteomics: The Good, the bad, and the future of this field', *Proteomes*, 8/3: 14.

Edman, Pehr (1949), 'A method for the determination of amino acid sequence in peptides', *Archives of biochemistry*, 22/3: 475.

Eng, Jimmy, Ashley McCormack and John Yates (1994), 'An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database', *Journal of the American Society for Mass Spectrometry*, 5/11: 976–989.

Feider, Clara, Anna Krieger, Rachel DeHoog and Livia Eberlin (2019), 'Ambient ionization mass spectrometry: Recent developments and applications', *Analytical Chemistry*, 91/7: 4266–4290.

Fiddyment, Sarah, Natalie Goodison, Elma Brenner et al. (2021), 'Girding the loins? Direct evidence of the use of a medieval parchment birthing girdle from biomolecular analysis', *Royal Society Open Science*, 8/3: 202055.

Fiddyment, Sarah, Bruce Holsinger, Chiara Ruzzier et al. (2015), 'Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting', *Proceedings of the National Academy of Sciences*, 112/49: 15066–15071.

Fiddyment, Sarah, Matthew Teasdale, Jiří Vnouček et al. (2019), 'So you want to do biocodicology? A field guide to the biological analysis of parchment', *Heritage Science*, 7/1: 35.

Fornelli, Luca, Timothy Toby, Luis Schachner et al. (2018), 'Top-down proteomics: Where we are, where we are going?', *Journal of proteomics*, 20/175: 3–4.

Fulton, Tara and Beth Shapiro (2019), 'Setting Up an Ancient DNA Laboratory', in Beth Shapiro, Axel Barlow, Peter Heintzman, Michael Hofreiter, Johanna Paijmans and André Soares (eds), *Ancient DNA. Methods in Molecular Biology*, New York: Humana Press, 1–13.

García, María Concepción (2005), 'The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography–electrospray mass spectrometry', *Journal of Chromatography B*, 825/2: 111–123.

Gilbert, Thomas, Hans-Jürgen Bandelt, Michael Hofreiter and Ian Barnes (2005), 'Assessing ancient DNA studies', *Trends in Ecology & Evolution*, 20/10: 541–544.

Gu, Jincui, Qingqing Li, Boyi Chen et al. (2019), 'Species identification of Bombyx mori and Antheraea pernyi silk via immunology and proteomics', *Scientific Reports*, 9/1: 9381.

Gu, Muxin and Michael Buckley (2018), 'Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting', *BMC Bioinformatics*, 19/1: 241.

Heaton, Karl, Caroline Solazzo, Matthew Collins, Jane Thomas-Oates and Edmund Bergström (2009), 'Towards the application of desorption electrospray ionisation mass spectrometry (DESI-MS) to the analysis of ancient proteins from artefacts', *Journal of Archaeological Science*, 36/10: 2145–2154.

Hendy, Jessica (2021), 'Ancient protein analysis in archaeology', *Science Advances*, 7/3: eabb9314.

Henzel, William, Colin Watanabe and John Stults (2003), 'Protein identification: the origins of peptide mass fingerprinting', *Journal of the American Society for Mass Spectrometry*, 14/9: 931–942.

Hill, Ryan, Matthew Wither, Travis Nemkov et al. (2015), 'Preserved proteins from extinct bison latifrons identified by tandem mass spectrometry; Hydroxylysine glycosides are a common feature of ancient collagen', *Molecular & Cellular Proteomics*, 14/7: 1946–1958.

Ivanov, Alexander, Li Zang and Barry Karger (2003), 'Low-attomole electrospray ionization MS and MS/MS analysis of protein tryptic digests using 20-µm-i.d. polystyrene–divinylbenzene monolithic capillary columns', *Analytical Chemistry*, 75/20: 5306–5316.

Karas, Michael, Ute Bahr and Thomas Dülcks (2000), 'Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine', *Fresenius' Journal of Analytical Chemistry*, 366/6: 669–676.

Karas, Michael and Franz Hillenkamp (1988), 'Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons', *Analytical Chemistry*, 60/20: 2299–2301.

Kasso, Tuuli, Markku Oinonen, Kenichiro Mizohata, Jaakko Tahkokallio and Tuomas Heikkilä (2020), 'Volumes of worth—Delimiting the sample size for radiocarbon dating of parchment', *Radiocarbon*, 63/1: 105–120.

Kostyukevich, Yury, Anna Bugrova, Vitaly Chagovets et al. (2018), 'Proteomic and lipidomic analysis of mammoth bone by high-resolution tandem mass spectrometry coupled with liquid chromatography', *European Journal of Mass Spectrometry*, 24/6: 411–419.

Kuckova, Stepanka, Hynek, Radovan and Milan Kodicek (2007), 'Identification of proteinaceous binders used in artworks by MALDI-TOF mass spectrometry', *Analytical and Bioanalytical Chemistry*, 388/1: 201–206.

Leitner, Alexander, Joachim Emmert, Katja Boerner and Wolfgang Lindner (2007), 'Influence of solvent additive composition on chromatographic separation and sodium adduct formation of peptides in HPLC–ESI MS', *Chromatographia*, 65/11: 649–653.

Lenčo, Juraj, Maria Khalikova and František Švec (2020), 'Dissolving peptides in 0.1% formic acid brings risk of artificial formylation', *Journal of Proteome Research*, 19/3: 993–999.

Leo, Gabriella, Ilaria Bonaduce, Alessia Andreotti et al. (2011), 'Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings', *Analytical Chemistry*, 83/6: 2056–2064.

Li, Li, Gong Decai, Yao Zhengyuan and Wang Jie (2019), 'A preliminary study of the decline in solubility of ancient silk protein', *Polymer Degradation and Stability*, 169: 108988.

Li, Li, Yuxuan Gong, Hao Yin and Decai Gong (2015), 'Different types of peptide detected by mass spectrometry among fresh silk and archaeological silk remains for distinguishing modern contamination', *Plos One*, 10/7: e0132827.

Lim, Megan and Kojo Elenitoba-Johnson (2004), 'Proteomics in pathology research', *Laboratory Investigation*, 84/10: 1227–1244.

Llamas, Bastien, Guido Valverde, Lars Fehren-Schmitz et al. (2017), 'From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era', *STAR: Science & Technology of Archaeological Research*, 3/1: 1–14.

Maixner, Frank, Thorsten Overath, Dennis Linke et al. (2013), 'Paleoproteomic study of the Iceman's brain tissue', *Cellular and Molecular Life Sciences*, 70/19: 3709–3722.

Manfredi, Marcello, Elettra Barberis, Fabio Gosetti et al. (2017), 'Method for noninvasive analysis of proteins and small molecules from ancient objects', *Analytical Chemistry*, 89/6: 3310–3317.

Mant, Colin, Yuxin Chen, Zhe Yan et al. (2007), 'HPLC analysis and purification of peptides', in Gregg Fields (ed), *Peptide Characterization and Application Protocols*, Totowa: Humana Press, 3–55.

Martin, Susan, Jeffrey Shabanowitz, Donald Hunt and Jarrod Marto (2000), 'Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry', *Analytical Chemistry*, 72/18: 4266–4274.

Marvasi, Massimiliano, Duccio Cavalieri, Giorgio Mastromei, Antonella Casaccia and Brunella Perito (2019), 'Omics technologies for an in-depth investigation of biodeterioration of cultural heritage', *International Biodeterioration & Biodegradation*, 144: 104736.

Mazza, Roberta (2019), 'Dating early Christian papyri: Old and new methods – Introduction', *Journal for the Study of the New Testament*, 42/1: 46–57.

Mellacheruvu, Dattatreya, Zachary Wright, Amber Couzens et al. (2013), 'The CRAPome: a contaminant repository for affinity purification–mass spectrometry data', *Nature Methods*, 10/8: 730–736.

Na, Seungjin and Eunok Paek (2020), 'Computational methods in mass spectrometry-based structural proteomics for studying protein structure, dynamics, and interactions', *Computational and Structural Biotechnology Journal*, 18: 1391–1402.

Nadler, Wiebke Maria, Dietmar Waidelich, Alexander Kerner et al. (2017), 'MALDI versus ESI: The impact of the ion source on peptide identification', *Journal of Proteome Research*, 16/3: 1207–1215.

Navas, Natalia, Julia Romero-Pastor, Eloisa Manzano and Carolina Cardell (2008), 'Benefits of applying combined diffuse reflectance FTIR spectroscopy and principal component analysis for the study of blue tempera historical painting', *Analytica Chimica Acta*, 630/2: 141–149.

Nemes, Peter and Akos Vertes (2007), 'Laser ablation electrospray ionization for atmospheric pressure, in vivo, and imaging mass spectrometry', *Analytical Chemistry*, 79/21: 8098–8106.

Newton, James, Gordon Ramage, Nikolaj Gadegaard et al. (2018), 'Minimally-destructive atmospheric ionisation mass spectrometry authenticates authorship of historical manuscripts', *Scientific Reports*, 8/1: 10944.

Noga, Marek, Filip Sucharski, Piotr Suder and Jerzy Silberring (2007), 'A practical guide to nano-LC troubleshooting', *Journal of Separation Science*, 30/14: 2179–2189.

Nshanian, Michael, Rajeswari Lakshmanan, Hao Chen, Rachel Ogorzalek Loo and Joseph Loo (2018), 'Enhancing sensitivity of liquid chromatography–mass spectrometry of peptides and proteins using supercharging agents', *International Journal of Mass Spectrometry*, 427: 157–164.

Ntasi, Georgia, Daniel Kirby, Ilaria Stanzione et al. (2021), 'A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects', *Journal of Proteomics*, 231: 104039.

Orlando, Ludovic, Aurélien Ginolhac, Guojie Zhang et al. (2013), 'Recalibrating equus evolution using the genome sequence of an early middle pleistocene horse', *Nature*, 499/7456: 74–78.

Orsini, Sibilla, Federica Parlanti and Ilaria Bonaduce (2017), 'Analytical pyrolysis of proteins in samples from artistic and archaeological objects', *Journal of Analytical and Applied Pyrolysis*, 124: 643–657.

Ostrom, Peggy, Michael Schall, Hasand Gandhi et al. (2000), 'New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry', *Geochimica et Cosmochimica Acta*, 64/6: 1043–1050.

Padula, Matthew, Iain Berry, Matthew Rourke et al. (2017), 'A comprehensive guide for performing sample preparation and top-down protein analysis', *Proteomes*, 5/2: 11.

Palmieri, Margarita, Manuela Vagnini, Lucia Pitzurra, Bruno Brunetti and Laura Cartechini (2013), 'Identification of animal glue and hen-egg yolk in paintings by use of enzyme-linked immunosorbent assay (ELISA)', *Analytical and Bioanalytical Chemistry*, 405/19: 6365–6371.

Pandeswari, Boomathi and Varatharajan Sabareesh (2019), 'Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry', *RSC Advances*, 9/1: 313–344.

Perkins, David, Darryl Pappin, David Creasy and John Cottrell (1999), 'Probability-based protein identification by searching sequence databases using mass spectrometry data', *Electrophoresis*, 20/18: 3551–3567.

Rabin, Ira and Oliver Hahn (2020), 'Detection of Fakes: The Merits and Limits of Non-Invasive Materials Analysis', in Cécile Michel and Michael Friedrich (eds), *Fakes and Forgeries of Written Artefacts from Ancient Mesopotamia to Modern China* (Studies in Manuscript Cultures, 20), Berlin: De Gruyter, 281–290.

Ramsøe, Abigail, Vivian van Heekeren, Paola Ponce et al. (2020), 'DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes', *Journal of Archaeological Science*, 115: 105080.

Righetti, Pier, Gleb Zilberstein and Svetlana Zilberstein (2020), 'EVA Technology and Proteomics: A Two-Pronged Attack on Cultural Heritage', *Journal of Proteome Research*, 19/8: 2914–2925.

Righetti, Pier, Gleb Zilberstein and Alfonsina D'Amato (2019), 'What Sherlock sorely missed: the EVA technology for cultural heritage exploration', *Expert Review of Proteomics*, 16/9: 1–10.

Roach, Patrick, Julia Laskin and Alexander Laskin (2010), 'Nanospray Desorption Electrospray Ionization: an Ambient Method for Liquid-Extraction Surface Sampling in Mass Spectrometry', *The Analyst*, 135/9: 2233–2236.

Rodriguez-Valera, Francisco (2004), 'Environmental genomics, the big picture?', *FEMS microbiology letters*, 231/2: 153–158.

Rosini, Elena, Paola D'Antona and Loredano Pollegioni (2020), 'Biosensors for D-Amino Acids: Detection Methods and Applications', *International Journal of Molecular Sciences*, 21/13: 4574.

Schedl, Andreas, Philipp Korntner, Thomas Eichinger et al. (2015), 'Detection of Cellulose-Derived Chromophores by Ambient Ionization-MS', *Analytical Chemistry*, 88/2: 1253-1258.

Schroeter, Elena and Timothy Cleland (2015), 'Glutamine deamidation: An indicator of antiquity, or preservational quality?', *Rapid Communications in Mass Spectrometry*, 30/2: 251–255.

Schulz, Anna, Silke Lautner, Jörg Fromm and Markus Fischer (2018), 'Not stealing from the treasure chest (or just a bit): Analyses on plant derived writing supports and non-invasive DNA sampling', *Plos One*, 13/6: e0198513.

Sciutto, Giorgia, Martina Zangheri, Silvia Pratiet al. (2016), 'Immunochemical Micro Imaging Analyses for the Detection of Proteins in Artworks', *Topics in Current Chemistry*, 374/3: 32.

Shin, Hyun-Cheol, Diana Deterra, Jonghoo Park et al. (2018), 'Ultra-high mass multimer analysis of protein-1a capping domains by a silicon nanomembrane detector', *Journal of Proteomics*, 175: 5–11.

Smoluch, Marek and Jerzy Silberring (2019), 'A Brief History of Mass Spectrometry', in Marek Smoluch, Giuseppe Grasso, Piotr Suder and Jerzy Silberring (eds), *Mass Spectrometry: An Applied Approach* (Wiley series on mass spectrometry), Hokboken: John Wiley, 5–8.

Solazzo, Caroline (2019), 'Characterizing historical textiles and clothing with proteomics', *Conservar Património*, 31: 97–114.

Spraker, Joseph, Gordon Luu and Laura Sanchez (2020), 'Imaging mass spectrometry for natural products discovery: a review of ionization methods', *Natural Product Reports*, 37/2: 150–162.

Stephens, Catherine, Bindesh Shrestha, Hannah Morris et al. (2010), 'Minimally Invasive Monitoring of Cellulose Degradation by Desorption Electrospray Ionization and Laser Ablation Electrospray Ionization Mass Spectrometry', *The Analyst*, 135: 2434–2444.

Swaney, Danielle, Craig Wenger and Joshua Coon (2010), 'Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics', *Journal of Proteome Research*, 9/3: 1323–1329.

Takáts, Zoltán, Justin Wiseman and Graham Cooks (2005), 'Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology', *Journal of Mass Spectrometry*, 40/10: 1261–1275.

Teasdale, Matthew, Sarah Fiddyment, Jiří Vnouček et al. (2021), 'The York Gospels: a 1000-year biological palimpsest', *Royal Society Open Science*, 4/10: 170988.

Toby, Timothy, Luca Fornelli and Neil Kelleher (2016), 'Progress in Top-Down Proteomics and the Analysis of Proteoforms', *Annual review of analytical chemistry*, 9/1: 499–519.

Toniolo, Lucia, Alfonsina D'Amato, Riccardo Saccenti, Davide Gulotta and Pier Giorgio Righetti (2012), 'The Silk Road, Marco Polo, a bible and its proteome: A detective story', *Journal of Proteomics*, 75/11: 3365–3373.

Tsiatsiani, Liana and Albert Heck (2015), 'Proteomics beyond trypsin', *The FEBS Journal*, 282/14: 2612–2626.

van der Werf, Inez Dorothé, Cosima Damiana Calvano, Giulia Germinario, Tommaso Cataldi and Luigia Sabbatini (2017), 'Chemical characterization of medieval illuminated parchment scrolls', *Microchemical Journal*, 134: 146–153.

Vargas Medina, Deyber Arley, Edvaldo Vasconcelos Soares Maciel and Fernando Mauro Lanças (2020), 'Miniaturization of liquid chromatography coupled to mass spectrometry. 3. Achievements on chip-based LC–MS devices', *TrAC Trends in Analytical Chemistry*, 131: 116003.

Vilanova, Cristina and Manuel Porcar (2020), 'Art-omics: multi-omics meet archaeology and art conservation', *Microbial Biotechnology*, 13/2: 435–441.

Welker, Frido, Jazmín Ramos-Madrigal, Petra Gutenbrunner, et al. (2020), 'The dental proteome of Homo antecessor', *Nature*, 580/7802: 235–238.

Wilm, Matthias and Matthias Mann (1994), 'Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?', *International Journal of Mass Spectrometry and Ion Processes*, 136/2: 167–180.

Wilson, Steven Ray, Tore Vehus, Henriette Sjaanes Berg and Elsa Lundanes (2015), 'Nano-LC in proteomics: recent advances and approaches', *Bioanalysis*, 7/14: 1799–1815.

Winter, Daniel, Marc Wilkins and William Donald (2019), 'Differential Ion Mobility–Mass Spectrometry for Detailed Analysis of the Proteome', *Trends in Biotechnology*, 37/2: 198–213.

Yamashita, Masamichi and John Fenn (1984), 'Electrospray ion source. Another variation on the free-jet theme', *The Journal of Physical Chemistry*, 88/20: 4451–4459.

Zengin, Talip, Sercan Kılıç, Ufuk Yenigün, Serhat Erdoğmuş and Ömür Baysal (2017), 'A Smart Way for Talking With Proteins; Proteomics', MOJ Proteomics & Bioinformatics, 5/3: 86-92.

Zhang, Wenzhu and Brian Chait (2000), 'ProFound: An Expert System for Protein Identification Using Mass Spectrometric Peptide Mapping Information', *Analytical Chemistry*, 72/11: 2482–2489.

Zhang, Yaoyang, Bryan Fonslow, Bing Shan, Moon-Chang Baek and John Yates (2013), 'Protein analysis by shotgun/bottom-up proteomics', *Chemical reviews*, 113/4: 2343–2394.

Zilberstein, Gleb, Uriel Maor, Emmanuil Baskin, Alfonsina D'Amato and Pier Righetti (2016), 'Unearthing Bulgakov's trace proteome from the Master i Margarita manuscript', *Journal of Proteomics*, 152: 102–108.

Zilberstein, Gleb, Svetlana Zilberstein, Uriel Maor and Pier Giorgio Righetti (2020), 'Surface analysis of ancient parchments via the EVA film: The Aleppo Codex', *Analytical Biochemistry*, 604: 113824.

Zou, Jing, Francis Talbot, Alessandra Tata et al. (2015), 'Ambient Mass Spectrometry Imaging with Picosecond Infrared Laser Ablation Electrospray Ionization (PIR-LAESI)', *Analytical Chemistry*, 87/24: 12071–12079.