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.

© 2015 Elsevier B.V. All rights reserved.

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/© 2015 Elsevier B.V. All rights reserved.

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.

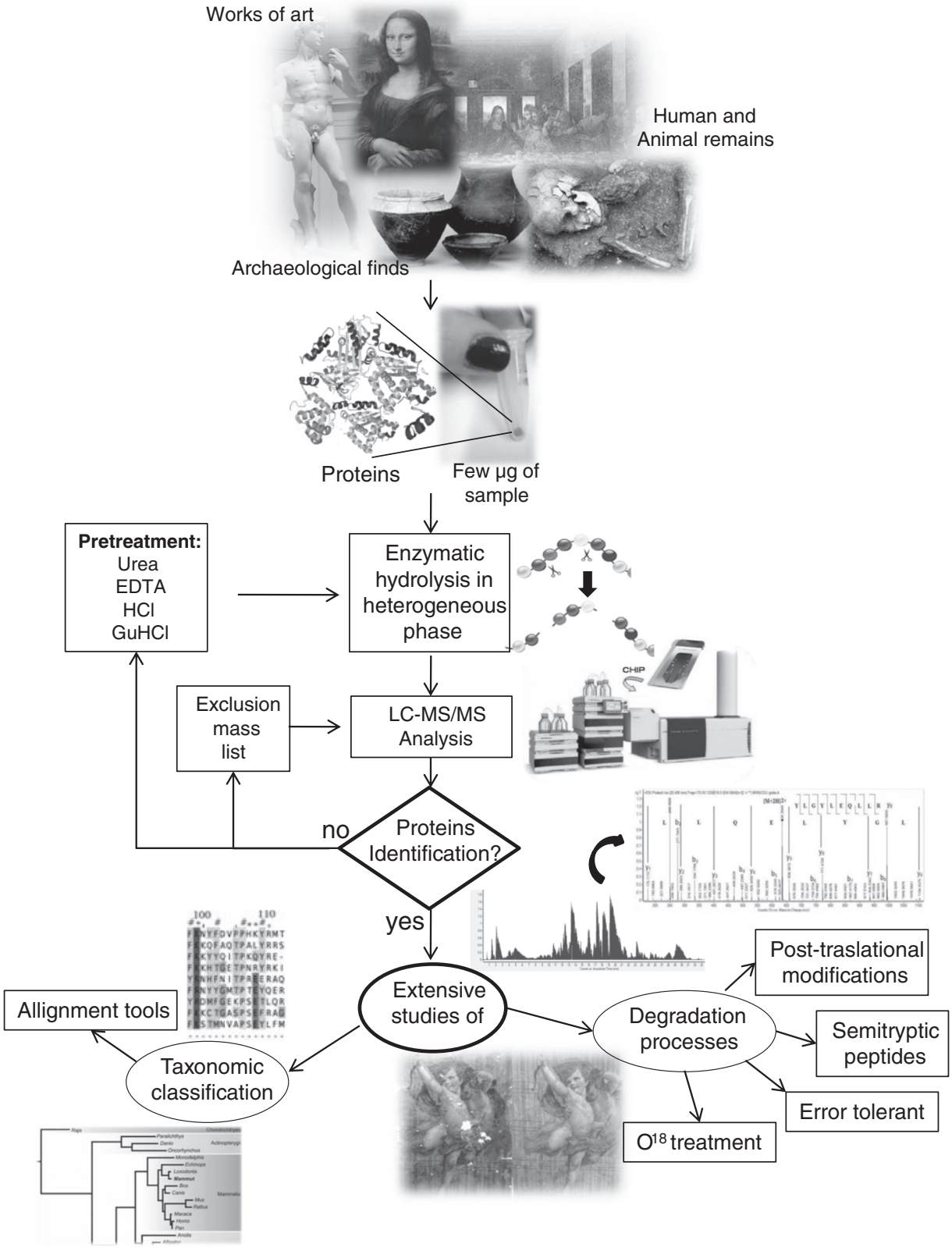

Fig. 1. General flow chart of proteomic investigations.

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

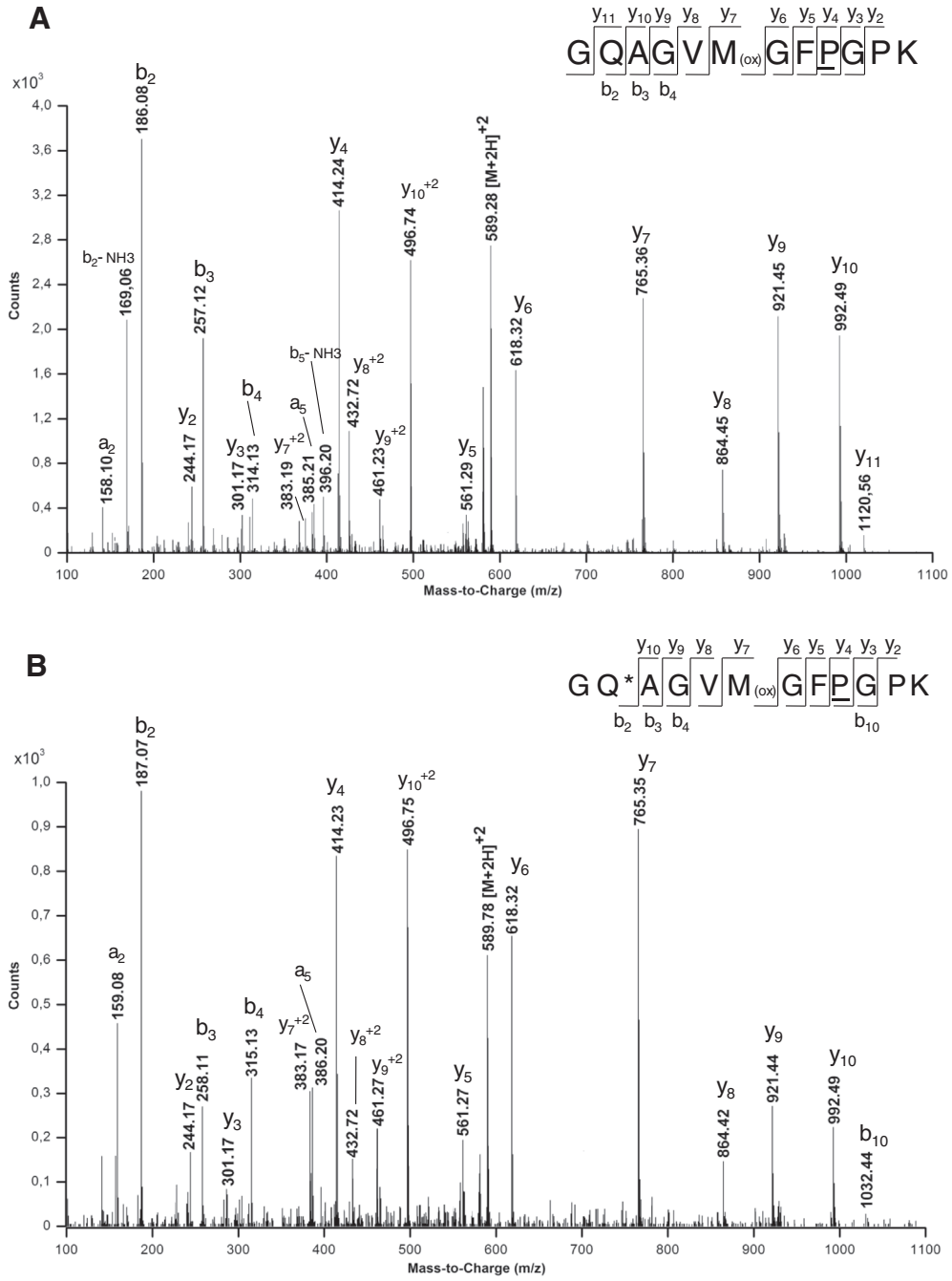

Fig. 2. MS/MS spectra of the doubly charged ions at *m*/*z* 589.28 (A) of the peptide GQAGVMGFPGPK of human collagen alpha-1(I) (P02452) and its deamidated form at *m*/*z* 589.78 (B) identified in the analysis of the sample of human bone dated 1st A.D. The product ions are indicated with the observed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

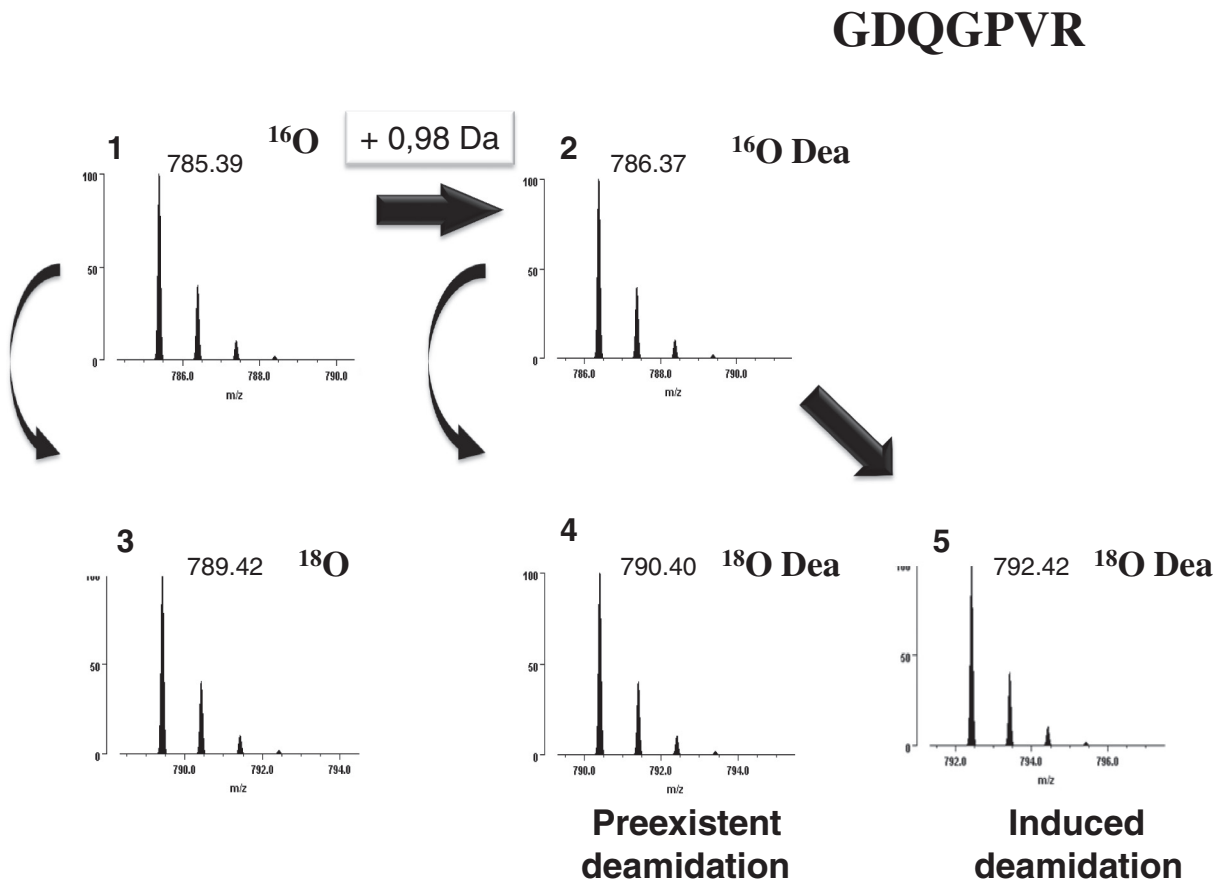

Fig. 3. Scheme of expected mass shifts upon enzymatic digestion in $H_2^{18}O$ of the peptide GDQGPVGR (MH^+ 735.39) of human collagen alpha-1(I) (P02452). Upon digestion with trypsin in $H_2^{18}O$ up to 2 ^{18}O are incorporated at the C-terminus of newly generated peptides with a mass shift of about 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) 1509300652060005 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.