Characterisation of degraded proteins in paintings using bottom-up proteomic approaches: new strategies for protein digestion and analysis of data

Sibilla Orsini[1], Avinash Yadav[2,3], Marialaura Dilillo[2], Liam A. McDonnell[2,4], Ilaria Bonaduce[1]

[1]Department of Chemistry and Industrial Chemistry, University of Pisa, Pisa, Italy.

[2] Fondazione Pisana per la Scienza ONLUS, San Giuliano Terme, Pisa, Italy

[3] Scuola Normale Superiore, Pisa, Italy

[4] Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, the Netherlands

* Correspondence to Ilaria.bonaduce@unipi.it

Abstract

Chemical hydrolysis assisted by microwave irradiation has been proposed as an alternative method for the analysis of proteins in highly insoluble matrices. In this work, chemical hydrolysis was applied for the first time to detect degraded proteins from paintings and polychromes. To evaluate the performance of this approach, the number of identified peptides, protein sequence coverage (%) and PSMs were compared with those obtained using two trypsin based proteomics procedures used for the analysis of samples from cultural heritage objects. It was found that chemical hydrolysis allowed the successful identification of all proteinaceous materials in all paint samples analysed except for egg proteins in one extremely degraded sample. Moreover in one sample casein was only identified by chemical digestion. In general, chemical hydrolysis identified more peptides, more PSM's and greater sequence coverage in the samples containing caseins, and often also in animal glue, highlighting the great potential of this approach for the rapid digestion and identification of insoluble and degraded proteins from the field of the cultural heritage.

1. Introduction

Proteinaceous materials, most commonly egg, collagen, and casein from milk, have been extensively used in the past as paint binders and varnishes for polychrome artefacts, panel paintings and murals, or as adhesives for gildings [1,2].

Assessing the presence of proteinaceous materials in artistic and archaeological objects and identifying their nature still represents a challenge for the analytical chemist [3,4]. Modifications of some amino acids, such as oxidation of Met, hydroxylation of Pro, deamidation of Gln and Asn, as well as the partial hydrolysis of the polypeptide chains, are degradation phenomena observed in aged proteins [5-7]. Moreover, it has been shown that ageing of proteins can lead to aggregation, cross-linking by covalent bonds, and the formation of stable complexes with inorganic materials[6,8-11]. All these changes may reduce the solubility of the proteins, thereby adversely affecting the efficiency of protein identification with procedures based on protein solubilisation and proteolytic digestion[12]. Moreover, the database search used for protein identification will underperform if the degradation phenomena are not taken into consideration.

Since its first report in 2006, proteomics techniques have been shown to be *potentially* successful for the identification of proteins in artistic and archaeological objects [3,13-16]; however the proteomics approaches used in biomedical research cannot be simply applied directly to artwork samples because of the extensive protein degradation present in these samples. New field-specific procedures need to be developed to address these field-specific analytical challenges [15,17]. Research so far has focussed on the improvement of the solubilisation step [17-22], and the combined use a protein deglycosidase enzyme to improve the detection of egg [23].

Controlled chemical hydrolysis has been proposed as an alternative approach to identify proteins[24-30]. Protein digestion with trifluoroacetic acid (TFA) 25% (v/v) assisted by microwave irradiation was successfully used to analyse highly insoluble proteins from plant membrane cells [30], casein and BSA [29]. It was demonstrated that chemical hydrolysis increased the number of identified peptides and the protein sequence coverage with respect to enzymatic proteolytic approaches. This approach might be exploited to overcome the lack of access for proteolytic enzymes in insoluble aggregated cross-linked proteins.

Here, we investigated for the first time the potential of chemical hydrolysis by TFA coupled with microwave irradiation to analyse insoluble proteins in samples collected from paintings and polychromies. To evaluate the performances of this approach, each sample was divided into three aliquots, one of which was digested with TFA [30], another pre-treated with PNgaseF and subsequently digested with trypsin [23], and the remaining aliquot was pre-treated with urea and subsequently digested with trypsin [5,31]. Data analysis strategies to improve our capabilities of identifying aged proteins are also discussed.

2.Experimental section

2.1 Samples

Six samples from polychromes, canvas and mural paintings with different conservation conditions and geographical origins were divided into three aliquots and analysed by three different proteomics approaches. The sample code, the typology of the artwork, its age and geographical origin, as well as the nature of the proteinaceous materials present (previously identified by GC-MS analysis[32]) are reported in Table S-1 of Supporting Information S1.

2.2 Analytical procedures

Each sample was homogenized with a pestle in a mortar, divided into three aliquots of identical mass (2 mg), and analysed according to the scheme reported in Figure 1, with three different protocols: UREA, PNGaseF, and TFA, based on the procedures reported in previous publications[5,17,23,30,31]. 6M Urea was used as denaturing agent to increase the solubilisation of the proteins in UREA protocol before tryptic digestion. A deglycosylation step with an enzyme, PNGaseF, was introduced in the pre-treatment of the sample before tryptic digestion in order to cleavage the bonds between protein and carbohydrates. While in the TFA protocol the proteins are solubilised and chemical digested with acid agent (TFA) at high concentration. The details are reported in the Supporting Information S2.
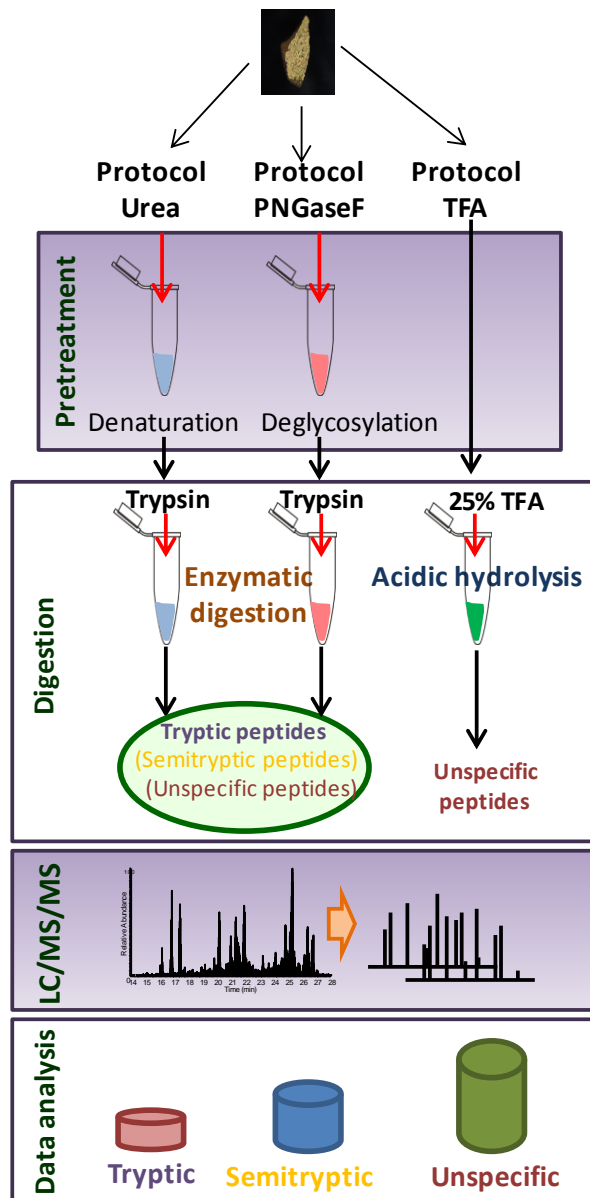
Figure 1: Scheme of the three analytical approaches: protocol with Urea, protocol with PNGaseF and protocol with TFA.

2.3 Data analysis

Raw MS/MS data were converted to peak lists in Mascot Generic Format (MGF) using Proteome Discoverer version 1.4 (Thermo Scientific). MGF files were searched on Mascot[33] server version 2.5 (Matrix Science) through Mascot Daemon version 2.5.1 (Matrix Science). A customised protein sequence database was created with protein sequences available from Uniprot [34]. A total of 239 protein sequences of the most abundant proteins contained in proteinaceous materials used as paint binders (egg, milk and collagen-based) was included in the target database. The protein sequences included were:

- Collagen Type I: alpha- 1, alpha- 2, alpha- 3, and alpha- 4 from rabbit, cow, sheep, goat and fish. Collagen Type II: alpha-1 from rabbit, cow, sheep, goat and fish [17,35,36].

- Egg yolk: apovitellenin-I, apolipoproteins B and A1, vitellongenins 1, 2 and 3, vitelline membrane and Ig lambda from chicken and duck [37]. Egg white: ovalbumin (54%), ovotransferrin (12%), ovomucoid (11%), ovomucin (3.5%), and lysozyme (3.4%) [37].
- Milk: caseins (casein-αS1 (33.6%) (w/w), casein-β (27.2%), casein-αS2 (10.1%), casein-κ (8.4%)) and β-lactoglobulin (8.3%) and α-lactoglobulin (2.4%) [38] from cow, sheep and goat.

A database of 116 common contaminant proteins (cRAP) was obtained from http://www.thegpm.org/crap/ and combined with the target database for all searches. For each Mascot search the precursor mass tolerance was set to 10 ppm and fragment mass tolerance set to 0.6 Da. The database search was conducted in tryptic and semi-tryptic modes with a maximum of 2 missed cleavages, and unspecific mode with 0 missed cleavages. Samples treated with TFA chemical hydrolysis were searched only in unspecific mode whereas samples treated with protocols PNGaseF and UREA were searched in tryptic, semi-tryptic and unspecific modes. Oxidation (M), Hydroxylation (P) and Deamidated (NQ) were selected as variable modifications for all searches. Oxidation of Cys, Hist, and Trp, and hydroxylation of Lys were not included as variable modifications as preliminary investigations demonstrated they only occurred to a minor extent and did not aid the identification of proteins, because the increased number of variable modifications decreases the statistical power with which proteins may be identified. Supporting Information S.3.1 shows a comparison of protein ID's using the four selected variable modifications with those obtained using all eight variable modifications. After database search, Mascot search results were filtered with an expectation value less than 0.05. Peptide hits from contaminant proteins were filtered out. False discovery rate (FDR) for all searches were controlled at 5 %.

3. Results and Discussion

A selection of samples collected from paintings and polychromies with different ages and conservation status[39] were analysed according to the analytical approach described in Figure 1. The identification of proteins from aged paintings and polychromes is not straightforward because aging can introduce a variety of non-standard modifications. For example it is known that aging may cause partial hydrolysis[6,17,40], which randomly cleaves the proteins at any amino acid position[17]. Tryptic digestion of aged proteins may produce tryptic peptides from non-hydrolysed proteins (i.e. c-terminal Arginine or Lysine[41]), semi-tryptic peptides from partially hydrolysed proteins, and unspecific peptides if the hydrolysed sequence did not contain any tryptic site. Thus, tryptic digest of samples which have aged significantly may contain tryptic, semi-tryptic and unspecific peptides. These three classes of peptides are best identified with a database search performed in tryptic, semi-tryptic or unspecific modes. Database search of trypsin treated samples which have undergone aging has been shown to perform better in semi-tryptic mode than in tryptic mode[17], because most proteins are partially hydrolysed is samples from aged works of art.

Peptides obtained with enzymatic and chemical digestion differ in the type of backbone cleavage. Tryptic peptides have a basic N-terminus and a basic residue at the C-terminus, leading to the generation by electrospray ionization of multiply charged peptides and the detection in tandem mass spectrometry experiments of N-terminal and C-terminal fragments; in contrast the non-specific peptides generated by chemical hydrolysis need not contain a basic amino acid, and so may contain only the basic N-terminus, leading to preference for N-terminal fragments in tandem mass spectrometry experiments.

The peptide search space for the database search is dependent on the number of proteins in the database as well as the enzyme cleavage rule, number of missed cleavages and the number of dynamic modifications. The search space for tryptic peptides is a subset of the semi-tryptic search space, which itself is a subset of unspecific search space. The unspecific search space is larger than the semi-tryptic search space, which itself is larger than a tryptic search space (unspecific>semi-tryptic>>tryptic). The increase in search space from tryptic to semi-tryptic or from semi-tryptic to unspecific significantly increases the required computational time and adversely affects statistical confidence. For example the number of candidate peptides of the human proteome using a tryptic search are 2.9E6 (enzyme: trypsin; minimum peptide length: 7; maximum peptide length: 40; # of permitted missed cleavages: 2), which increases to 4.6E7 for a semitryptic search (enzyme: trypsin; minimum peptide length: 7; maximum peptide length: 40; # of permitted missed cleavages: 2) and 4.0E8 for an unspecific search (enzyme: unspecific; minimum peptide length: 7; maximum peptide length: 40). We expected our samples to have undergone partial hydrolysis due to aging, thus we decided to conduct the database search in tryptic, semi-tryptic and unspecific modes. If we utilised the full proteome from all organisms commonly used to produce the proteinaceous materials historically used for paintings, the resulting semi-tryptic search space would be very large, leading to long analysis times, low statistical probabilities of peptide spectral matches and low statistical confidence; similarly, the unspecific search would become practically and statistically intractable. To overcome this, we restricted the proteins in our target database to the most abundant proteins expected to be present in the proteinaceous materials used for paintings and polychromies. The restricted number of proteins in the target database allowed us to analyse the data obtained with the three analytical protocols (protocols UREA, PNGaseF and TFA) in unspecific, semi-tryptic and tryptic modes. The numbers of identified peptides, protein sequence coverage (%) and peptide spectrum matches (PSMs) were chosen as output parameters to compare the data obtained from the samples treated with the three different protocols, and results are summarised in Tables S-2 – S-7.

The results demonstrated that the UREA, PNGaseF and TFA based methods enabled the identification of the proteinaceous materials present in each sample. The only exceptions were egg proteins from the 2nd century mural painting (MP2[nd]AD), which were only identified using the PNGaseF protocol, and milk proteins from the 15th century easel painting (EP15[th]AD), which were identified only with the TFA-based chemical hydrolysis protocol.

The denaturation provided by the chaotrope urea led to increased protein sequence coverage, numbers of peptides and PSMs in samples containing collagen, such as those collected from the 4th century and 13th century mural paintings (MP4thBC, Table S-7; MP13thAD, Figure 2, Table S-5), as well as the 15th century easel painting (EP15thAD, Table S-3). Chemical hydrolysis using TFA enabled the identification of caseins in samples containing milk, for example from the 15th century easel painting (EP15thAD, Table S-3) and 13th century mural painting (MP13thAD, Figure 2, Table S-5), giving the greatest protein sequence coverage and number of PSMs.
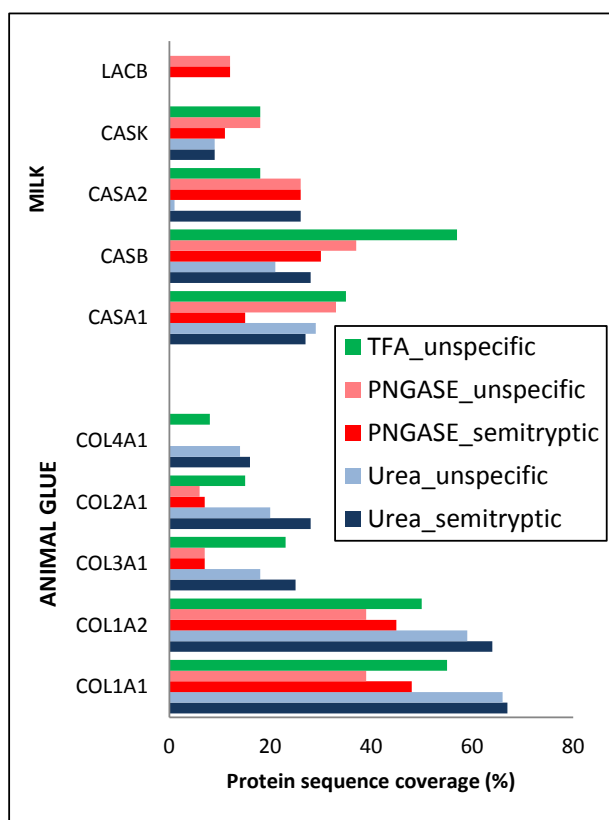

Figure 2: Bar chart of protein sequence coverage (%) of collagen and casein from *Bos taurus* from a 13th century mural painting (details provided as MP13thAD, Table S-5).

In those samples that contained egg the use of the deglycosidase PNGaseF led to the largest number of protein identifications, e.g. mucin-5B (MUC5B) and ovalbumin (OVAL) from the 15th century easel painting (EP15thAD, Table S-3) and vitellogenin-2 (VIT2) and lysozyme C (LYSC) from the 2nd century mural painting (MP2ndAD, Figure 3). As expected, PNGaseF protocol gave the greatest sequence coverage and PSMs for glycosylated proteins (Table 4), such as vitellogenin-1 (VIT1) (8 glycosylation sites), vitellogenin-2 (VIT2) (7 glycosylation sites), vitellogenin-3 (VIT3) (2 glycosylation sites), apolipoprotein B (APOB) (1 glycosylation site), ovalbumin (OVAL)(1 glycosylation site), ovalbumin-related protein Y (OVALY) (4 glycosylation sites) and mucin-5B (MUC5B) (16 glycosylation sites), which are all proteins found in egg.
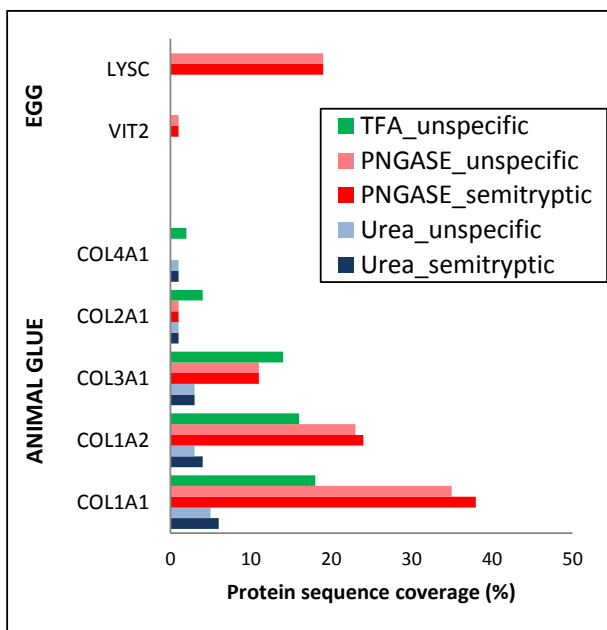
Figure 3: Bar chart of protein sequence coverage (%) of collagen from *Bos taurus* and egg from *Gallus gallus* in MP2ndAD.

The results show that the TFA based chemical hydrolysis allowed the identification of all the proteinaceous materials found in every sample, with the only exception being the egg proteins in the 2nd century mural painting (MP2ndAD, (Table S-6). In samples containing animal glue as proteinaceous material, such as the 15th century easel painting (EP15thAD, Table S-3), the 2nd century mural painting (MP2ndAD, Table S-6), and the 16th century easel painting (CP16thAD, Table S-2) the TFA-based chemical hydrolysis led to the identification of a much greater number of peptides, as well as comparable or greater sequence coverage and number of PSMs with respect to the results using the trypsin based protocols with UREA and PNGaseF.
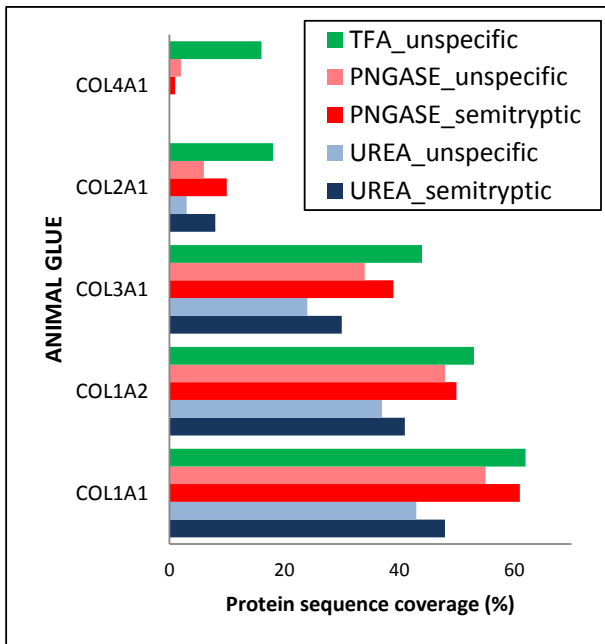
Figure 4: Bar chart of protein sequence coverage (%) of collagen from *Bos taurus* of CP16[th]AD.

It is interesting to note that in general the number of peptides identified using an unspecific database search was less than that obtained using a semi-tryptic search, whereas the number of peptides identified with a semi-tryptic search was greater than using a tryptic search. As example, the peptides identified with the UREA protocol of COL1A1 of EP15[th]AD were 177 in unspecific search mode, 227 in semi-tryptic and 52 in tryptic mode and sequence coverage is reported in Supporting Information S.3.2. The results clearly confirm that partial hydrolysis of proteins occurs during aging of the paintings[17]. For this reason the largest number of identified peptides in samples digested with trypsin were obtained using a semi-tryptic database search. The unspecific database search would also have included these peptides within the search space, but was compromised by the higher threshold score for positive identification, due to the larger search space in unspecific mode. Peptides that were identified in tryptic searches may not pass the higher threshold for the larger unspecific search space, and similarly when comparing the semi-tryptic search to unspecific search. This can be seen in Tables S-2 to S-7, which demonstrate that the number of peptides, protein sequence coverage and number of PSMs obtained using an unspecific search mode were generally less than those obtained using semi-tryptic mode and tryptic mode (an example is shown in Supporting Information S-4). The samples MP2ndAD and MP14thAD showed marginally better results using a tryptic search than semitryptic and unspecific searches but if the peptide identification was made stricter, to reflect the larger databases of the semitryptic and unspecific searches, the results became comparable to those obtained with semitryptic and unspecific searches. More details are reported in Supporting Information S3.3.

Differently from collagen and egg proteins, milk proteins showed a higher number of identified peptides and protein sequence coverage in unspecific mode than in semi-tryptic mode (Supporting Information S-5). In order to confirm that this behaviour is typical of milk proteins, a standard sample of casein from bovine milk

was analysed using the same protocols and same database search modes. The results demonstrated that the number of peptides, protein sequence coverage and PSMs are the same or increase when using an unspecific database search (Supplementary Information S.3.4). An explanation can be found in the fact that solubilisation of casein requires the use of alkaline media, which may cause hydrolysis of the polypeptide chain. The TFA-based chemical hydrolysis was very efficient at identifying casein proteins, also for the standard sample of casein from bovine milk. It was noted however that the detection of LACB was poor. This protein is the least abundant in the mixture (casein-αS1 (33.6%) (w/w) and casein-β (27.2%), followed by casein-αS2 (10.1%), casein-κ (8.4%), β-lactoglobulin(LACB) (8.3%) and α-lactoglobulin (2.4%)[42]). The comparison of the TFA-based chemical hydrolysis with the trypsin based methods indicates that the unspecific nature of the chemical digestion may render it more biased toward abundant proteins. Data-dependent tandem mass spectrometry selects the more abundant peptides for MS/MS, the larger number of peptides generated by unspecific chemical digestion means the mass spectrometer will select more peptides from the more abundant proteins, and thus fewer from the lower abundant proteins (exacerbated here by the relatively short LC-gradients).

Conclusion

In this study a new protocol for the proteomic analysis of insoluble and degraded proteins from samples collected from paintings and polychromes is reported. The protocol is based on the chemical digestion of proteins with 25% TFA assisted by microwave irradiation, and is proposed as an alternative to the established trypsin based approaches. The principal advantages are circumvention of the protein solubilisation, denaturation and enzymatic digestion steps used in established, enzyme-based approaches. The protocol successfully identified all proteinaceous materials in all samples with the exception of egg proteins in one extremely degraded sample. Furthermore chemical digestion identified more peptides, more PSM's and greater sequence coverage in the samples containing caseins. In particular in one sample casein was only identified by chemical digestion. The major advantage over enzymatic digestion relies on the fact that TFA can access more of the polypeptide chain in insoluble proteins characterised by inter- and intra-molecular aggregation, cross-linking and stable complexes with inorganic species (as found in paintings). It should be noted that while the presence of unspecific cleavages in samples digested with trypsin can be used to indicate protein aging, the unspecific cleavages resulting from TFA based chemical digestion means it cannot be used in this manner.

The differences in sequence coverage obtained from the same sample using different protocols are most likely due to the changing nature of the proteins in the samples, which will depend on the sample composition, the material preparation and artistic or manufacturing technique, as well as the conservation history of the object. For example the ability of the chaotrope urea to aid protein solubilization is likely to be

acutely dependent on the degree of the protein aggregation/cross-linking. Similarly the success of PNGase F pretreatment will be dependent on the degree of N-linked glycosylation. The small size of TFA is considered to make it less prone to being affected by cross-linking and thus to perform better when heavily cross-linked proteins are present. Moreover, the TFA based analytical protocol is simple, rapid and economical, and has the advantage of being tuneable, as reaction time and temperature may be adjusted to maximise chances of success in identifying proteins in extremely degraded samples. For these reasons this new approach based on the use of TFA will be extremely helpful as an alternative or as part of an integrative approach (TFA, Trypsin, PNGase F) for the study of proteins from samples from paintings and other cultural heritage objects.

The study also highlights the critical nature of data analysis for the identification of aged proteins in samples from cultural heritage, and thus also highlights the need for new tailored approaches that take into consideration protein changes resulting from anthropogenic and natural degradation.

References

(1) Colombini, M. P.; Modugno, F. *Organic Mass Spectrometry in Art and Archaeology*; Wiley, 2009.
(2) Mills, J.; Raymond, W. *Organic Chemistry of Museum Objects*, 2nd ed.: Oxford, 1994.
(3) Dallongeville, S.; Garnier, N.; Rolando, C.; Tokarski, C. *Chem. Rev.* **2016**, *116*, 2-79.
(4) Bonaduce, I.; Ribechini, E.; Modugno, F.; Colombini, M. P. *Top. Curr. Chem.* **2016**, *374*, 6.
(5) Gambino, M.; Cappitelli, F.; Cattò, C.; Carpen, A.; Principi, P.; Ghezzi, L.; Bonaduce, I.; Galano, E.; Pucci, P.; Birolo, L.; Villa, F.; Forlani, F. *J. Biosci.* **2013**, *38*, 397-408.
(6) Duce, C.; Ghezzi, L.; Onor, M.; Bonaduce, I.; Colombini, M. P.; Tine, M. R.; Bramanti, E. *Anal. Bioanal. Chem.* **2012**, *402*, 2183-2193.
(7) Leo, G.; Bonaduce, I.; Andreotti, A.; Marino, G.; Pucci, P.; Colombini, M. P.; Birolo, L. *Anal. Chem.* **2011**, *83*, 2056-2064.
(8) Orsini, S.; Parlanti, F.; Bonaduce, I. *J. Anal. Appl. Pyrolysis* **2017**, *124*, 643-657.
(9) Duce, C.; Bramanti, E.; Ghezzi, L.; Bernazzani, L.; Bonaduce, I.; Colombini, M. P.; Spepi, A.; Biagi, S.; Tine, M. R. *Dalton Trans.* **2013**, *42*, 5975-5984.
(10) Ghezzi, L.; Duce, C.; Bernazzani, L.; Bramanti, E.; Colombini, M. P.; Tiné, M. R.; Bonaduce, I. *J. Therm. Anal. Calorim.* **2015**, *122*, 315-322.
(11) Orsini, S.; Bramanti, E.; Bonaduce, I. *J. Anal. Appl. Pyrolysis* **2018**, *in press*.
(12) Colombini, M. P.; Andreotti, A.; Bonaduce, I.; Modugno, F.; Ribechini, E. *Acc. Chem. Res.* **2010**, *43*, 715-727.
(13) Tokarski, C.; Martin, E.; Rolando, C.; Cren-Olivé, C. *Anal. Chem.* **2006**, *78*, 1494-1502.
(14) Fremont, W.; Sanyova, J.; Saverwyns, S.; Vandenabeele, P.; Moens, L. *Anal. Bioanal. Chem.* **2009**, *395*, 1991-1999.
(15) Leo, G.; Cartechini, L.; Pucci, P.; Sgamellotti, A.; Marino, G.; Birolo, L. *Anal. Bioanal. Chem.* **2009**, *395*, 2269-2280.
(16) Kuckova, S.; Hynek, R.; Kodicek, M. *Anal. Bioanal. Chem.* **2007**, *338*, 201-206.
(17) Vinciguerra, R.; De Chiaro, A.; Pucci, P.; Marino, G.; Birolo, L. *Microchem. J.* **2016**, *126*, 341-348.
(18) He, J.; Yan, H.; Fan, C. *RSC Adv* **2014**, *4*, 42608.
(19) Rao, H.; Li, B.; Yang, Y.; Ma, Q.; Wang, C. *Anal . Methods* **2015**, *7*, 143.
(20) Chambery, A.; Di Maro, A.; Sanges, C.; Severino, V.; Tarantino, M.; Lamberti, A.; Parente, A.; Arcari, P. *Anal. Bioanal. Chem.* **2009**, *395*, 2281-2291.
(21) Barker, A.; Venables, B.; Stevens, S. M.; Seeley, K. W.; wang, P.; Wolverton, S. *J. Archaeol. Method Th.* **2012**, *19*, 407-439.

(22) Barker, A.; Venables, B.; Stevens, S. M.; Seeley, K. W.; Wang, P.; Wolverton, S. *J. Archaeol. Method Theory* **2012**, *19*, 407-439.

(23) Vinciguerra, R.; Galano, E.; Vallone, F.; Greco, G.; Vergana, A.; Bonaduce, I.; Marino, G.; Pucci, P.; Amoresano, A.; Birolo, L. *Anal. Chem.* **2015**, *87*, 10178-10182.

(24) Zhong, H. Y.; Zhang, Y.; When, Z. H.; Li, L. *Nat. Biotechnol.* **2004**, *22*, 1291-1296.

(25) Yassine, M. M.; Gou, N.; Zhong, H.; Li, L.; Lucy, C. A. *Anal. Chim. Acta* **2007**, *597*, 41-49.

(26) Wang, N.; MacKenzie, L.; De Souza, A. G.; Zhong, H.; Goss, G.; Li, L. *J. Proteome Res.* **2007**, *6*, 263-272.

(27) Swatkoski, S.; Russell, S. C.; Edwards, N.; Fenselau, C. *Anal. Chem.* **2006**, *78*, 181-188.

(28) Gonzalez, M. H.; Souza, G. B.; Oliveira, V.; Forato, L. A.; Nòbrega, J. A.; Nogueira, A. R. A. *Talanta* **2009**, *79*, 396-401.

(29) Wang, N.; Li, L. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1573-1587.

(30) Zhong, H.; Marcus, S. L.; Li, L. *J. Am. Soc.Mass Spectrom.* **2005**, *16*, 471-481.

(31) Lluveras-Tenorio, A.; Vinciguerra, R.; Galano, E.; Blaensdorf, C.; Emmerling, E.; Perla Colombini, M.; Birolo, L.; Bonaduce, I. *PLOS ONE* **2017**, *12*, e0172990.

(32) Bonaduce, I.; Colombini, M. P. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2523-2527.

(33) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cot, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.

(34) UniProt_Consortium. *Nucleic Acids Res.* **2015**, *43*, D204–D212.

(35) Schweitzer, M.; Hill, C. L.; Asara, J. M.; Lane, W. S.; Pincus, S. H. *J. Mol. Evol.* **2002**, *55*, 696-705.

(36) Wadsworth, C.; Buckley, M. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 605-615.

(37) Mine, Y. *Trends Food Sci. Technol.* **1995**, *6*, 225-232.

(38) Heck, J. M. L.; Schennink, A.; van Valenberg, H. J. F.; Bovenhuis, H.; Visker, M. H. P. W.; van Arendonk, J. A. M.; van Hooijdonk, A. C. M. *J. Dairy Sci.* **2009**, *92*, 1192-1202.

(39) Orsini, S.; Parlanti, F.; Bonaduce, I. *J. Anal. Appl. Pyrolysis* **2017**, *124*, 643-657.

(40) Warinner, C.; Hendy, J.; Speller, C.; Cappellini, E.; Fischer, R.; Trachsel, C.; Arneborg, J.; Lynnerup, N.; Craig, O. E.; Swallow, D. M.; Fotakis, A.; Christensen, R. J.; Olsen, J. V.; Liebert, A.; Montalva, N.; Fiddyment, S.; Charlton, S.; Mackie, M.; Canci, A.; Bouwman, A., et al. *Sci Rep.* **2014**, *4*, 7104.

(41) Bensadek, D.; Monigatti, F.; Steen, J. A. J.; Steen, H. *Int. J. Mass Spectrom.* **2007**, *268*, 181-189.

(42) Sigma_Aldrich. In *Product Information*.