

Supporting information.

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

Sibilla Orsini¹, Avinash Yadav^{2,3}, Marialaura Dilillo², Liam A. McDonnell^{2,4}, Ilaria Bonaduce¹

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

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

³Scuola Normale Superiore, Pisa, Italy

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

* Correspondence to ilaria.bonaduce@unipi.it

| | |
|---|-----------|
| S.1 Samples | 3 |
| Table S-1. Sample code, geographical origin and dating of artistic samples. | 3 |
| S.2 Analytical procedures..... | 4 |
| S.2.1 Chemicals and Reagents | 4 |
| S.2.2 Analytical protocol based on using urea followed by trypsin digestion (UREA) | 4 |
| S.2.3 Analytical protocol based on using PNGaseF followed by trypsin digestion (PNGase)..... | 4 |
| S.2.4 Analytical protocol based on the use of TFA (TFA) | 4 |
| S.2.5 Sample desalting | 4 |
| S.2.6 LC-MS/MS | 5 |
| S.3 Results | 6 |
| Table S-2: Comparison of protocols and database search methods for analysis of CP16 th AD. | 6 |
| Table S-3: Comparison of protocols and database search methods for analysis of EP15 th AD. | 7 |
| Table S-4: Comparison of protocols and database search methods for analysis of MP14 th AD. | 8 |
| Table S-5: Comparison of protocols and database search methods for analysis of MP13 th AD. | 9 |
| Table S-6: Comparison of protocols and database search methods for analysis of MP2 nd AD..... | 10 |
| Table S-7: Comparison of protocols and database search methods for analysis of MP4 th BC..... | 11 |
| S.3.1 Variable modifications | 12 |
| Table S-8: Numbers of modified peptides found in CP16 th AD and MP2 nd AD. | 12 |
| S.3.2 Examples of protein sequence coverage obtained using the different digestion protocols and different database search modes | 13 |
| S.3.3 Effect of identification threshold in tryptic search..... | 15 |
| Table S-9: Effect of database search stringency on number of identified peptides..... | 15 |
| S.3.4 Analysis of a standard sample of casein from bovine milk..... | 16 |
| Table S-10: Comparison of protocols and database search methods for the analysis of a standard sample of casein from bovine milk. | 16 |

S.1 Samples

Samples from polychromies, canvas and mural paintings with different conservation conditions and geographical origins are described in Table S-1, indicating 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 analysis¹) are reported in Supporting Information Table S-1.

Table S-1. Sample code, geographical origin and dating of artistic samples.

| Code | Type | Origin | Age | Description | GC/MS ID ¹ |
|----------------------------|-----------------|----------|-----------------------------|--|-----------------------|
| CP16thAD | Canvas Painting | Italy | 16 th century AD | easel painting on canvas | collagen |
| EP15thAD | Easel Painting | Greece | 15 th century AD | Greek icon | casein, collagen, egg |
| MP13thAD | Mural Painting | Italy | 13 th century AD | restoration glue on a mural painting | casein, collagen |
| MP14thAD | Mural Painting | Italy | 14 th century AD | gilding on a mural painting of a church | egg |
| MP4thBC | Mural Painting | Turkey | 4 th century BC | painted decoration of an internal wall of a house | unknown |
| MP2ndAD | Mural Painting | Pakistan | 2 nd century AD | painted decoration of an external wall of a civic building | unknown |

S.2 Analytical procedures

S.2.1 Chemicals and Reagents

Trypsin from Bovine Pancreas, Dithiothreitol (DTT), trifluoroacetic acid (TFA), formic acid (FoAc), urea, ammonium bicarbonate (AmBic), water and ACN were purchased from Sigma Aldrich. Peptide-N-glycosidase F (PNGaseF) was obtained from Roche Custom Biotech. All samples were treated in Eppendorf Protein LoBind tubes (1.5 mL) purchased from Sigma-Aldrich.

S.2.2 Analytical protocol based on using urea followed by trypsin digestion (UREA)

The sample was treated according to the protocol reported previously^{2,3}. In brief, the sample (50-200 µg) was treated with 10 µL of 6M Urea for 10 min at room temperature and then sonicated for 20 min at room temperature. The solution was then diluted 1:6 with water, followed by the addition of 0.5 µL of trypsin (1 µg/ µL), and left to incubate for 15-18 hours at 37°C. To stop the proteolytic digestion the solution was acidified by the addition of 20 µL of 10% formic acid.

S.2.3 Analytical protocol based on using PNGaseF followed by trypsin digestion (PNGase)

The sample was treated according to the protocol reported previously⁴. In brief, the sample (50-200 µg) was partially dissolved in 50 µL of 50 mM ammonium bicarbonate, 3 µL of PNGaseF solution added, and the solution left to incubate at room temperature for 2 hours then stopped by boiling for 2 minutes. For proteolytic digestion 0.5 µL of trypsin (1 µg/ µL) was added, and the solution left to incubate for 15-18 hours at 37°C. The solution was acidified with 10% formic acid to stop the trypsin digestion.

S.2.4 Analytical protocol based on the use of TFA (TFA)

The sample was treated according to the protocol reported previously⁵. In brief, 20 µL of 25% (v/v) trifluoroacetic acid was added to 50-200 µg of sample followed by 0.1 µL of 30mg/ml of dithiothreitol. Then the sample vials were placed in a plastic beaker, which contained water in sufficient volume to immerse the sample solution (in the Eppendorf tube), and the beaker placed in a domestic microwave oven (Smeg, model number SA37X). After ten minutes of microwave irradiation at 450 W power, 80 µL of water was added to the sample. The non-specific protein digest was centrifuged at 14000 rpm for 10 min and the supernatant collected.

For samples containing a high amount of calcium carbonate, especially those from mural paintings, we added 25% (v/v) trifluoroacetic acid until CO₂ gas production ceased. Then the solution was dried under N₂ flow, and 20 µL of 25% (v/v) trifluoroacetic acid was added to the sample and treated as described above.

S.2.5 Sample desalting

Peptide desalting was performed using an AssayMAP Bravo platform (Agilent technologies) with C18 cartridges (Agilent, 5 µL bed volume) using the peptide cleanup V2 protocol. Briefly, C18 cartridges were primed with ACN, equilibrated with 50 µL of 0.1% formic acid, 100 µL of diluted samples were loaded at 5

μL/min; two cup wash and a cartridge wash were performed with 50 μL at 10 μL/min, followed by a stringent syringe wash with ACN and peptide elution with 30 μL of 80% ACN and 0.1% FoAC at 5 μL/min. The solutions were dried in a SpeedVac and re-suspended in 10 μL of 10% formic acid. An aliquot of the solution (1-3 μg) was then injected in an LC-MS/MS instrument for protein identification.

S.2.6 LC-MS/MS

The LC-MS/MS analysis was performed using an EASY-nLC 1000 coupled to an Orbitrap Fusion mass spectrometer (both Thermo Scientific GmbH, Bremen, Germany). The peptides were first desalted on a C18 trap column (Acclaim® PepMap100, 75 μm x 2 cm, 3μm particle size, 100Å pore size) and then separated on a 75 μm x 50 cm C18 analytical column (Acclaim® Rapid Separation Liquid Chromatography (RSLC) column, 2μm particle size, 100 Å pore size) using a flow rate of 300 nL/min and a 35min run. Briefly, peptides were loaded at 800 bar followed by a non-linear gradient: 0min, 2%B; t=15min, 35%B; t=23min, 95%B; t=35min, 95%B. Buffer A consisted of 0.1% formic acid in water and Buffer B of 0.1% formic acid in ACN. Data-dependent analysis was performed in a top-speed mode with a 2 second cycle time. MS scans were acquired in the Orbitrap at 120 K resolution, m/z range 375-1500, AGC target of 5×10^5 and 100 ms maximum injection time. Monoisotopic precursors with a charge state from +2 to +8 and intensity greater than 5×10^3 were subject to MS/MS in the ion trap using collision-induced dissociation (CID): 1.6 m/z quadrupole isolation window, 35% NCE, AGC target of 5×10^3 , 300 ms maximum injection time and dynamic exclusion of 20s.

S.3 Results

The selection of paintings and polychromies described in Table S-1 were analyzed with the protocols and database search methods reported in Figure 1. The number of identified peptides, protein sequence coverage (%) and PSMs were chosen as output parameters to compare the data. The results are reported in Tables S-2 to S-7.

Table S-2: Comparison of protocols and database search methods for analysis of CP16thAD.

Identified peptides, protein sequence coverage (%) and PSMs of proteins from animal glue of Bos taurus in CP16thAD. The largest number of identified peptides, protein sequence coverage (%) and number of PSMs for each protein are indicated in red. Results are reported for different database search modes: T for tryptic cleavage, S for semitryptic and U for unspecific.

| Prot. | # identified peptides | | | | | | sequence coverage (%) | | | | | | PSMs | | | | | |
|--------|-----------------------|-----|-----|--------|-----|-----|-----------------------|------|----|----|--------|----|------|-----|------|-----|-----|-----|
| | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA | Urea | | | TFA |
| | T | S | U | T | S | U | U | T | S | U | T | S | U | U | T | S | U | U |
| COL1A1 | 28 | 188 | 162 | 48 | 245 | 195 | 588 | 32 | 48 | 43 | 52 | 61 | 55 | 62 | 92 | 298 | 229 | 894 |
| COL1A2 | 19 | 125 | 104 | 26 | 166 | 141 | 329 | 23 | 41 | 37 | 29 | 50 | 48 | 53 | 39 | 153 | 123 | 470 |
| COL3A1 | 16 | 44 | 30 | 3 | 73 | 51 | 119 | 23 | 30 | 24 | 4 | 39 | 34 | 44 | 34 | 65 | 45 | 180 |
| COL2A1 | 1 | 15 | 11 | - | 17 | 11 | 58 | 9 | 8 | 3 | - | 10 | 6 | 18 | 5 | 17 | 13 | 81 |
| COL4A1 | - | - | - | - | 2 | 4 | 29 | - | - | - | - | 1 | 2 | 16 | - | - | - | 31 |

Table S-3: Comparison of protocols and database search methods for analysis of EP15thAD.

Identified peptides, protein sequence coverage (%) and PSMs of animal glue proteins from *Bos taurus*, egg from *Gallus gallus* and milk from *Ovis aries* in EP15thAD. The largest number of identified peptides, protein sequence coverage (%) and number of PSMs for each protein are indicated in red. Results are reported for different database search modes: T for tryptic cleavage, S for semitryptic and U for unspecific.

| Prot. | # identified peptides | | | | | | | sequence coverage (%) | | | | | | | PSMs | | | | | | |
|---------------|-----------------------|-----|-----|--------|-----|-----|-----|-----------------------|----|----|--------|----|----|-----|------|-----|-----|--------|-----|-----|-----|
| | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA |
| | T | S | U | T | S | U | U | T | S | U | T | S | U | U | T | S | U | T | S | U | U |
| COL1A1 | 52 | 227 | 177 | 52 | 150 | 114 | 215 | 56 | 59 | 55 | 56 | 58 | 53 | 42 | 265 | 478 | 339 | 253 | 323 | 239 | 227 |
| COL1A2 | 32 | 158 | 132 | 38 | 97 | 77 | 93 | 40 | 48 | 46 | 50 | 51 | 47 | 34 | 107 | 243 | 190 | 97 | 148 | 120 | 96 |
| COL3A1 | 23 | 68 | 52 | 27 | 44 | 34 | 25 | 32 | 44 | 39 | 41 | 43 | 36 | 23 | 65 | 121 | 87 | 62 | 74 | 54 | 25 |
| COL2A1 | 5 | 17 | 16 | 2 | 9 | 9 | 30 | 6 | 9 | 9 | 2 | 5 | 5 | 11 | 13 | 28 | 20 | 8 | 14 | 11 | 31 |
| COL4A1 | - | 5 | 5 | - | 3 | - | 7 | - | 4 | 4 | - | 2 | - | 5 | - | 6 | 5 | - | 3 | - | 7 |
| VIT2 | 43 | 101 | 95 | 57 | 164 | 160 | 266 | 26 | 35 | 34 | 33 | 44 | 40 | 39 | 104 | 181 | 164 | 207 | 347 | 321 | 285 |
| VIT1 | 22 | 31 | 28 | 43 | 91 | 79 | 102 | 12 | 15 | 14 | 26 | 30 | 27 | 21 | 33 | 44 | 41 | 84 | 138 | 125 | 105 |
| APOV1 | 3 | 15 | 15 | 3 | 19 | 18 | 44 | 33 | 52 | 52 | 33 | 55 | 54 | 55 | 20 | 34 | 32 | 30 | 54 | 48 | 51 |
| APOA1 | 6 | 5 | 4 | 7 | 7 | 7 | - | 16 | 16 | 16 | 21 | 21 | 21 | - | 6 | 5 | 4 | 7 | 7 | 7 | - |
| APOB | - | 8 | 9 | - | 16 | 15 | 1 | - | 2 | 2 | - | 4 | 4 | 1 | - | 8 | 9 | - | 21 | 19 | 1 |
| VIT3 | 3 | 3 | 3 | 3 | 5 | 5 | 1 | 7 | 7 | 7 | 11 | 12 | 12 | 1 | 3 | 3 | 3 | 4 | 6 | 6 | 1 |
| MUC5B | - | - | - | - | 2 | 1 | - | - | - | - | - | 1 | 1 | - | - | - | - | - | 2 | 1 | - |
| OVAL | - | - | - | - | 2 | 2 | - | - | - | - | - | 1 | 1 | - | - | - | - | - | 2 | 2 | - |
| TRFE | 6 | 8 | 7 | 6 | 6 | 5 | 2 | 8 | 10 | 9 | 9 | 9 | 9 | 1 | 7 | 8 | 7 | 6 | 6 | 5 | 2 |
| LYSC | 3 | 4 | 4 | 3 | 3 | 3 | - | 13 | 18 | 18 | 10 | 10 | 10 | - | 3 | 4 | 4 | 5 | 5 | 5 | - |
| CASB | - | - | - | - | - | - | 18 | - | - | - | - | - | - | 27 | - | - | - | - | - | - | 26 |
| CSN2 | - | - | - | - | - | - | 2 | - | - | - | - | - | - | 13 | - | - | - | - | - | - | 3 |
| LACB | - | - | - | - | - | - | 2 | - | - | - | - | - | - | 3 | - | - | - | - | - | - | 2 |

Table S-4: Comparison of protocols and database search methods for analysis of MP14thAD.

Identified peptides, protein sequence coverage (%) and PSMs of egg from *Gallus gallus* in MP14thAD. The largest number of identified peptides, protein sequence coverage (%) and number of PSMs for each protein are indicated in red. Results are reported for different database search modes: T for tryptic cleavage, S for semitryptic and U for unspecific.

| Protein | # identified peptides | | | | | | sequence coverage (%) | | | | | | PSMs | | | | | | | | |
|---------|-----------------------|----|----|--------|----|----|-----------------------|------|----|----|--------|----|------|-----|------|-----|----|--------|-----|-----|-----|
| | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA |
| | T | S | U | T | S | U | U | T | S | U | T | S | U | U | T | S | U | T | S | U | U |
| VIT2 | 22 | 80 | 73 | 77 | 72 | 62 | 33 | 14 | 23 | 23 | 24 | 26 | 23 | 8 | 43 | 109 | 98 | 125 | 121 | 105 | 36 |
| VIT1 | 9 | 25 | 22 | 25 | 22 | 19 | 8 | 6 | 9 | 9 | 12 | 12 | 11 | 3 | 10 | 28 | 24 | 33 | 29 | 27 | 8 |
| APOV1 | 3 | 15 | 16 | 15 | 15 | 14 | 8 | 33 | 33 | 33 | 51 | 43 | 43 | 30 | 7 | 21 | 21 | 26 | 26 | 24 | 9 |
| APOB | 2 | 3 | 2 | 3 | 4 | 1 | - | 1 | 30 | 27 | 2 | 35 | 10 | - | 2 | 3 | 2 | 3 | 4 | 1 | - |
| VIT3 | - | 1 | - | 4 | 2 | 1 | - | - | 1 | - | 8 | 5 | 1 | - | - | 1 | - | 4 | 2 | 1 | - |
| MUC5B | 2 | 2 | 2 | 2 | 2 | 3 | - | 1 | 1 | 1 | 2 | 2 | 2 | - | 2 | 2 | 2 | 2 | 2 | 3 | - |
| VMO1 | 1 | 1 | 1 | 2 | 2 | 3 | - | 1 | 1 | 1 | 10 | 10 | 8 | - | 2 | 2 | 2 | 4 | 4 | 5 | - |
| OVAL | 1 | 6 | 14 | 6 | 8 | 10 | 11 | 2 | 8 | 17 | 8 | 11 | 17 | 10 | 1 | 6 | 15 | 6 | 8 | 10 | 13 |
| OVALY | 1 | 5 | 4 | 3 | 5 | 4 | - | 1 | 7 | 5 | 5 | 7 | 5 | - | 1 | 5 | 4 | 3 | 5 | 4 | - |
| OVALX | 1 | 2 | 1 | 1 | 1 | 1 | - | 1 | 2 | 1 | 1 | 2 | 1 | - | 1 | 2 | 1 | 1 | 1 | 1 | - |
| TRFE | 5 | 10 | 9 | 7 | 17 | 12 | 1 | 8 | 12 | 10 | 10 | 17 | 12 | 1 | 6 | 11 | 10 | 11 | 18 | 13 | 1 |
| LYSC | 2 | 6 | 6 | 4 | 9 | 8 | 4 | 6 | 14 | 14 | 5 | 24 | 23 | 7 | 5 | 11 | 10 | 17 | 24 | 22 | 4 |

Table S-5: Comparison of protocols and database search methods for analysis of MP13thAD.

Identified peptides, protein sequence coverage (%) and PSMs of collagen and milk from *Bos taurus* in MP13thAD. The largest number of identified peptides, protein sequence coverage (%) and number of PSMs for each protein are indicated in red. Results are reported for different database search modes: T for tryptic cleavage, S for semitryptic and U for unspecific.

| Prot. | # identified peptides | | | | | | | sequence coverage (%) | | | | | | | PSMs | | | | | | |
|---------------|-----------------------|-----|-----|--------|-----|----|-----|-----------------------|----|----|--------|----|----|-----|------|-----|-----|--------|-----|-----|-----|
| | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA |
| | T | S | U | T | S | U | U | T | S | U | T | S | U | U | T | S | U | T | S | U | U |
| COL1A1 | 93 | 232 | 189 | 41 | 126 | 87 | 449 | 58 | 67 | 66 | 45 | 48 | 39 | 55 | 534 | 621 | 497 | 123 | 195 | 133 | 475 |
| COL1A2 | 68 | 182 | 151 | 41 | 89 | 67 | 232 | 50 | 64 | 59 | 49 | 45 | 39 | 50 | 256 | 383 | 298 | 88 | 141 | 101 | 239 |
| COL3A1 | 14 | 20 | 15 | 7 | 5 | 4 | 23 | 23 | 25 | 18 | 8 | 7 | 7 | 23 | 19 | 27 | 19 | 7 | 5 | 4 | 24 |
| COL2A1 | 25 | 38 | 26 | 6 | 12 | 10 | 48 | 27 | 28 | 20 | 6 | 7 | 6 | 15 | 49 | 55 | 36 | 9 | 14 | 11 | 51 |
| COL4A1 | - | 6 | 4 | - | - | - | 14 | - | 16 | 14 | - | - | - | 8 | - | 6 | 4 | - | - | - | 15 |
| CASA1 | 1 | 22 | 30 | 1 | 15 | 24 | 17 | 5 | 27 | 29 | 5 | 15 | 33 | 35 | 10 | 83 | 85 | 4 | 61 | 62 | 24 |
| CASB | 5 | 39 | 18 | 5 | 35 | 65 | 142 | 15 | 28 | 21 | 15 | 30 | 37 | 57 | 50 | 217 | 31 | 43 | 144 | 169 | 241 |
| CASA2 | 4 | 11 | 2 | 4 | 11 | 12 | 8 | 12 | 26 | 1 | 12 | 26 | 26 | 18 | 15 | 31 | 6 | 14 | 31 | 34 | 8 |
| CASK | 1 | 5 | 5 | 1 | 6 | 10 | 11 | 3 | 9 | 9 | 3 | 11 | 18 | 18 | 8 | 20 | 11 | 4 | 16 | 23 | 13 |
| LACB | 1 | - | - | 5 | 7 | 7 | - | 3 | - | - | 10 | 12 | 12 | - | 1 | - | - | 7 | 12 | 9 | - |

Table S-6: Comparison of protocols and database search methods for analysis of MP2ndAD.

Identified peptides, protein sequence coverage (%) and PSMs of collagen from *Bos taurus* and egg white from *Gallus gallus* in MP2ndAD. The largest number of identified peptides, protein sequence coverage (%) and number of PSMs for each protein are indicated in red. Results are reported for different database search modes: T for tryptic cleavage, S for semitryptic and U for unspecific.

| Protein | # identified peptides | | | | | | sequence coverage (%) | | | | | | PSMs | | | | | |
|---------|-----------------------|---|---|--------|----|----|-----------------------|------|---|---|--------|----|------|-----|------|----|---|-----|
| | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA | Urea | | | TFA |
| | T | S | U | T | S | U | U | T | S | U | T | S | U | U | T | S | U | U |
| COL1A1 | 8 | 7 | 6 | 33 | 38 | 35 | 45 | 7 | 6 | 5 | 40 | 38 | 35 | 18 | 16 | 11 | 9 | 45 |
| COL1A2 | 5 | 5 | 4 | 23 | 25 | 23 | 23 | 4 | 4 | 3 | 27 | 24 | 23 | 16 | 7 | 7 | 5 | 23 |
| COL3A1 | 1 | 2 | 2 | 4 | 6 | 6 | 16 | 1 | 3 | 3 | 6 | 11 | 11 | 14 | 1 | 2 | 2 | 16 |
| COL2A1 | - | 1 | 1 | 1 | 1 | 1 | 5 | - | 1 | 1 | 1 | 1 | 1 | 4 | - | 1 | 1 | 5 |
| COL4A1 | - | 1 | 1 | - | - | - | 1 | - | 1 | 1 | - | - | - | 2 | - | 1 | 1 | 1 |
| VIT2 | - | - | - | - | 4 | 4 | - | - | - | - | - | 1 | 1 | - | - | - | - | - |
| LYSC | - | - | - | 2 | 2 | 2 | - | - | - | - | 19 | 19 | 19 | - | - | - | 2 | - |

Table S-7: Comparison of protocols and database search methods for analysis of MP4thBC.

Identified peptides, protein sequence coverage (%) and PSMs of collagen from *Bos taurus* in MP4thBC. The largest number of identified peptides, protein sequence coverage (%) and number of PSMs for each protein are indicated in red. Results are reported for different database search modes: T for tryptic cleavage, S for semitryptic and U for unspecific.

| Prot. | # identified peptides | | | | | | | Protein sequence coverage (%) | | | | | | | PSMs | | | | | | |
|---------------|-----------------------|-----|-----|--------|----|----|-----|-------------------------------|----|----|--------|----|----|-----|------|-----|-----|--------|----|----|-----|
| | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA | Urea | | | PNGase | | | TFA |
| | T | S | U | T | S | U | U | T | S | U | T | S | U | U | T | S | U | T | S | U | U |
| COL1A1 | 52 | 211 | 208 | 6 | 34 | 32 | 15 | 50 | 61 | 60 | 5 | 18 | 18 | 10 | 349 | 531 | 525 | 10 | 44 | 40 | 17 |
| COL1A2 | 46 | 161 | 156 | - | 19 | 18 | 31 | 50 | 59 | 58 | - | 12 | 11 | 24 | 202 | 333 | 315 | - | 23 | 20 | 39 |
| COL3A1 | 40 | 86 | 78 | - | 10 | 7 | 14 | 52 | 61 | 56 | - | 12 | 10 | 11 | 166 | 203 | 186 | - | 10 | 8 | 14 |
| COL2A1 | 5 | 15 | 16 | - | 8 | 10 | 1 | 5 | 9 | 9 | - | 6 | 7 | 2 | 23 | 31 | 30 | - | 10 | 12 | 1 |
| COL4A1 | - | 7 | 6 | - | 7 | 6 | 1 | - | 5 | 5 | - | 4 | 3 | 1 | - | 8 | 7 | - | 9 | 8 | 1 |

S.3.1 Variable modifications

Table S.8 shows examples of the incidence of different protein modifications in paintings and polychromies, and the effect of the number of variable modifications on the ability to identify proteins. Two representative samples were selected, CP16thAD (one of the samples with the largest number of identified peptides) and MP2ndAD (a highly degraded sample), and the Mascot search was performed using two sets of variable modifications.

Table S-8: Numbers of modified peptides found in CP16thAD and MP2ndAD.

Number of modified peptides identified for samples CP16thAD and MP2ndAD using the PNGaseF protocol and semitryptic search with **A)** four variable modifications, and **B)** eight variable modifications.

| Variable modification | CP16 th AD | | MP2 nd AD | |
|-----------------------|-----------------------|------|----------------------|-----|
| | A | B | A | B |
| Hydroxylation P | 1178 | 1081 | 211 | 167 |
| Oxidation M | 124 | 120 | 93 | 90 |
| Deamidation N | 137 | 134 | 116 | 110 |
| Deamidation Q | 157 | 146 | 59 | 46 |
| | | | | |
| Hydroxylation K | | 25 | | 15 |
| Oxidation C | | 10 | | 4 |
| Oxidation H | | 4 | | 3 |
| Oxidation W | | 4 | | 4 |

S.3.2 Examples of protein sequence coverage obtained using the different digestion protocols and different database search modes

Sample MP13th AD; COL1A2 sequence coverage with UREA protocol.

MLSFVDTRTLLLLAVTSCSLATCQSLQEATARKGPSGDRGPRGERGPPGPPGRDGDGIPGP
PGPPGPPGPPGLGNNFAAQFDAKGGGPGPMGLMGPRGPPGASGAPGPQGFQGGPPGEPG
EPGQTGPAGARGPPGPPGKAGEDGHPGKPRPGERGVVGPQGARGFPGTPGLPGFKGIR
GHNGLDGLKGQPGAPGVKGEPGAPGENGTPGQTGARGLPGERGRVGAPGPAGARSDG
SVGPVGPAGPIGSAGPPGFPGAPGPKGELGPVGNPGPAGPAGPRGEVGLPGLSGPVGPPG
NPGANGLPGAAGLPGVAGAPGLPGPRGIPGPVGAAGATGARGLVGEPGPAGSKGESG
NKGEPGAVGQPGPPGPSGEEGKRGSTGEIGPAGPPGPPGLRGNPGSRGLPGADGRAGVM
GPAGSRGATGPAGVRGPNGDSGRPGEPGLMGPRGFPGSPGNIPAGKEGPVGLPGIDGRP
GPIGPAGARGEPPNIGFPGPKGPSGDPGKAGEKGHAGLAGARGAPGPDGNNGAQQGPPGLQ
GVQGGKGEQGPAGPPGFQQLPGPAGTAGEAGKPGERGIPGEFGLPGPAGARGERGPPGES
GAAGPTGPIGSRGPSGPPGPDGNKGEPGVVGAPGTAGPSGPSGLPGERGAAGIPGGKGEK
GETGLRGDIGSPGRDGARGAPGAIGAPGPAGANGDRGEAGPAGPAGPRGSPGERGEV
GPAGPNGFAGPAGAAGQPGAKGERGTKGPKGENGPVGPTGPVGAAGPSGPNGPAGS
RGDGGPPGATGFPGAAGRTGPPGPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRSGETG
ASGPPGFVGEKGPSGEPGTAGPPGTPGPQGLGAPGFLGLPGSRGERGLPGVAGSVGEPPG
PLGIAGPPGARGPPGNVGNPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKGERGYPGNA
GPVGAAGAPGPQGPVGPVGKHGNGRGEPPAGAVGPAGAVGPRGPSGPQGIRGDKGEPGD
KGPRGLPGLKGHNGLQGLPGLAGHHGDQGAPGAVGPAGPRGPAGPSGPAGKDGRIGQPGA
VGPAGIRGSQGSQGPAGPPGPPGPPGPPGPSGGGYEFGFDGDFYRADQPRSPTSLRPKDY
EVDATLKSNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIVYCD
FSTGETCIRAQPEDIPVKNWYRNSKAKKHVVWGETINGGTQFEYNVEGVTTKEMATQLAFMR
LLANHASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTN
EWQKTIIEYKTNKPSRLPILDIAPLDIGGADQEIRLNIGPVCFK

| | |
|---|-------------|
| — | tryptic |
| — | semitryptic |
| — | unspecific |

An analysis of the identified peptides of COL1A2 revealed that four peptides identified using an unspecific search were not identified in semitryptic mode. These peptides were due to protein hydrolysis. On the other hand, thirty-one semi-tryptic peptides were not identified using an unspecific search because they did not pass the threshold score (because the much larger search space significantly reduced the statistical significance of a match).

Sample MP13th AD; CASB sequence coverage with UREA protocol.

MKVLILACLVALALAREQEELNVVGETVESLSSSEESITHINKKIEKFQSEEQQQTEDELQDKIH
PFAQAQSLVYPFTGPIPNLPLTQTTPVVPPFLQPEIMGVPKVKETMVPKHKEMPFPKY
PVEPFTESQSLTLTDVEKLHLPLPLVQSWMHQPPQPLPPTVMFPPQSVLSLSQPKVLPVPQKA
VPQRDMPIQAFLLYQEPVLGPVRGPFPILV

| | |
|---|-------------|
| — | tryptic |
| — | semitryptic |
| — | unspecific |

Sample MP13th AD; CASB sequence coverage with PNGaseF protocol.

MKVLILACLVALALAREQEELNVVGETVESLSSSEESITHINKKIEKFQSEEQQQTEDELQDKIH
PFAQAQSLVYPFTGPIPNLPLTQTTPVVPPFLQPEIMGVPKVKETMVPKHKEMPFPKY
PVEPFTESQSLTLTDVEKLHLPLPLVQSWMHQPPQPLPPTVMFPPQSVLSLSQPKVLPVPQKA
VPQRDMPIQAFLLYQEPVLGPVRGPFPILV

| | |
|---|-------------|
| — | tryptic |
| — | semitryptic |
| — | unspecific |

The protein sequence coverage and the number of identified peptides of CASB obtained using an unspecific database search was greater than that obtained using semitryptic and tryptic searches. This is probably due to the slaked lime that was used during a 20th century restoration⁶, which will have led to hydrolysis of the protein.

S.3.3 Effect of identification threshold in tryptic search.

A stricter database search may be performed by increasing the Mascot identification score threshold or decreasing the Mascot expectation value threshold. The expectation value (E) is derived directly from the score and the threshold:

$$E = P_{threshold} * \exp ((S_{threshold} - score) / 10))$$

in which $P_{threshold}$ is a probability threshold, S the peptide score and $S_{threshold}$ the threshold score. The expectation value is the number of times you could expect to get this score or better by chance. To determine the effect of a stricter search on the number of peptides identified in the tryptic search the peptide identifications were filtered by decreasing the peptide expectation value threshold from 0.05, to 0.025 and 0.01. Results are summarized in Table S-9:

Table S-9: Effect of database search stringency on number of identified peptides.

Number of identified peptides, sequence coverage (%) and number of PSMs for sample MP2ndAD analyzed using the PNGaseF protocol and searched in tryptic, semitryptic and unspecific mode using different thresholds of the peptide expectation value threshold. T is for tryptic, S for semitryptic and U for unspecific.

| | Number identified peptides | | | | | Sequence coverage (%) | | | | | Number PSMs | | | | |
|---------------|-------------------------------------|-------|------|------|------|-------------------------------------|-------|------|------|------|-------------------------------------|-------|------|------|------|
| | peptide expectation value threshold | | | | | peptide expectation value threshold | | | | | peptide expectation value threshold | | | | |
| | 0.05 | 0.025 | 0.01 | 0.05 | 0.05 | 0.05 | 0.025 | 0.01 | 0.05 | 0.05 | 0.05 | 0.025 | 0.01 | 0.05 | 0.05 |
| | T | T | T | S | U | T | T | T | S | U | T | T | T | S | A |
| COL1A1 | 33 | 30 | 27 | 38 | 35 | 40 | 38 | 33 | 38 | 35 | 70 | 61 | 55 | 64 | 60 |
| COL1A2 | 23 | 21 | 18 | 25 | 23 | 27 | 25 | 20 | 24 | 23 | 45 | 41 | 32 | 41 | 35 |
| COL3A1 | 4 | 4 | 4 | 6 | 6 | 6 | 4 | 4 | 11 | 11 | 6 | 6 | 6 | 9 | 8 |
| COL2A1 | 1 | - | - | 1 | 1 | 1 | - | - | 1 | 1 | 2 | - | - | 1 | 1 |
| COL4A1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| VIT2 | - | - | - | 4 | 4 | - | - | - | 1 | 1 | - | - | - | 5 | 4 |
| LYSC | 2 | 2 | 2 | 2 | 2 | 19 | 19 | 19 | 19 | 19 | 2 | 2 | 2 | 2 | 2 |

S.3.4 Analysis of a standard sample of casein from bovine milk

Casein from bovine milk was dissolved in water by adding a few drops of 2.5M NH₃, until a clear solution was obtained (5 mg/ml). The solution was then applied onto a glass slide and allowed to dry. Samples were then collected with a scalpel and analyzed. The number of identified peptides, protein sequence coverage and PSMs are reported in Table S-10.

Table S-10: Comparison of protocols and database search methods for the analysis of a standard sample of casein from bovine milk.

Identified peptides, protein sequence coverage (%) and PSMs of casein from *Bos taurus* in a reference sample of casein analysed with UREA, PNGaseF and TFA. The highest values of identified peptides, protein sequence coverages (%) and PSMs for each protein are reported in red. Results are reported for different search modes: T for tryptic, S for semitryptic, and U for unspecific. The TFA protocol was run in duplicate (U₁ and U₂).

| | # identified peptides | | | | | | | | | Sequence coverage (%) | | | | | | | | | PSMs | | | | | | | | |
|--------------|-----------------------|----|----|--------|----|----|----------------|----------------|--|-----------------------|----|----|--------|----|----|----------------|----------------|--|------|-----|-----|--------|-----|-----|----------------|----------------|--|
| | Urea | | | PNGase | | | TFA | | | Urea | | | PNGase | | | TFA | | | Urea | | | PNGase | | | TFA | | |
| | T | S | U | T | S | U | U ₁ | U ₂ | | T | S | U | T | S | U | U ₁ | U ₂ | | T | S | U | T | S | U | U ₁ | U ₂ | |
| CASA1 | 8 | 14 | 16 | 8 | 22 | 24 | 78 | 79 | | 13 | 21 | 21 | 10 | 15 | 15 | 41 | 46 | | 21 | 65 | 69 | 24 | 110 | 113 | 164 | 155 | |
| CASB | 21 | 55 | 66 | 17 | 31 | 33 | 217 | 234 | | 20 | 35 | 37 | 21 | 34 | 36 | 62 | 63 | | 87 | 253 | 247 | 41 | 62 | 60 | 433 | 462 | |
| CASA2 | 15 | 30 | 40 | 12 | 18 | 20 | 16 | 22 | | 16 | 26 | 27 | 10 | 18 | 18 | 20 | 24 | | 42 | 75 | 77 | 31 | 40 | 42 | 37 | 37 | |
| CASK | 6 | 18 | 29 | 15 | 37 | 40 | 32 | 28 | | 5 | 21 | 21 | 16 | 28 | 28 | 25 | 25 | | 36 | 97 | 91 | 84 | 153 | 154 | 46 | 44 | |
| LACB | 18 | 23 | 23 | 22 | 48 | 49 | 6 | 1 | | 25 | 41 | 41 | 26 | 33 | 33 | 6 | 1 | | 24 | 39 | 38 | 52 | 160 | 162 | 15 | 1 | |

References

- (1) Bonaduce, I.; Colombini, M. P. *Rapid Communications in Mass Spectrometry* **2003**, *17*, 2523-2527.
- (2) 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.
- (3) Lluveras-Tenorio, A.; Vinciguerra, R.; Galano, E.; Blaensdorf, C.; Emmerling, E.; Perla Colombini, M.; Birolo, L.; Bonaduce, I. *PLOS ONE* **2017**, *12*, e0172990.
- (4) Vinciguerra, R.; Galano, E.; Vallone, F.; Greco, G.; Vergana, A.; Bonaduce, I.; Marino, G.; Pucci, P.; Amoresano, A.; Birolo, L. *Analytical Chemistry* **2015**, *87*, 10178-10182.
- (5) Zhong, H.; Marcus, S. L.; Li, L. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 471-481.
- (6) Andreotti, A.; Baracchini, C.; Bonaduce, I.; Caleca, A.; Colombini, M.; Paolucci, A. In *Proceedings of 2008 ICOM-CC Triennial Conference*, 2014, pp 825-832.