New developments in high-resolution mass spectrometry for proteomic analysis applied to cultural heritage : study of intact proteins, their cross-linking and interaction in artworks and museum objects

Francesca Galluzzi

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

SPÉCIALITÉ CHIMIE ANALYTIQUE ET ENVIRONNEMENTALE

Par Francesca GALLUZZI

New developments in high-resolution mass spectrometry for proteomic analysis applied to cultural heritage: study of intact proteins, their cross-linking and interaction in artworks and museum objects

Sous la direction de: Caroline TOKARSKI

Soutenue le 23 Février 2021

Membres du jury :

M COLLINS Matthew	Professeur, U.Copenhagen/ U. Cambridge	Rapporteur
Mme BIROLO Leila	Professeure, U. Napoli, Federico II	Rapporteur
Mme. ARSLANOGLU Julie	Chargé de recherche, The Metropolitan Museum of Art	Examinateur
M. MENU Michel	Directeur du Département Recherche du C2RMF	Examinateur
Mme. BONADUCE Ilaria	Professeure associée, U. Pisa	Examinateur
M. CHAPOULIE Remy	Professeur, U. Bordeaux Montaigne	Président du jury
M. DUFOURC Erick	Directeur de recherche émérite, U. Bordeaux	Examinateur
Mme TOKARSKI Caroline	Professeure, U. Bordeaux	Directrice de thèse

Your reward will be the widening of the horizon as you climb.
And if you achieve that reward, you will ask no other.
Cecilia Payne-Gaposchk

Acknowledgements

This doctoral research is nearing its end, and the last three years had it all, including a global pandemic. Jokes apart, I wanted to thank all the beautiful people who helped and support me during this journey.

Above all, I would like to express my gratitude to my supervisor Prof. Caroline Tokarski who guided and helped me throughout this thesis.

I sincerely thank Prof. Matthew Collins, University of Copenhagen and University of Cambridge, and Prof. Leila Birolo, University of Federico II, Naples for their willingness to be referees of this work. I would also like to express my gratitude to the other members of my PhD committee, Michel Menu, Julie Arslanoglu, Ilaria Bonaduce, Remy Chapoulie, and Erick Dufourc.

I would also thank the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 722606 that funded my project.

Thanks to my academic family from TEMPERA: prof. Enrico Capellini and all the other professors of the network, and to my favouriteTemperini (Fabiana, Carla, Georgia, Petra, Patrick, Diana, and Miranda) with whom I spent unforgettable and stressful moments during the network training. I learned to not give up on challenges and to expand my comfort zone. I will miss our trips around Europe, but I will not miss the rain in Copenhagen. A huge thanks to Karin Norris and Anne-Sofie Munk Autzen for their incredible work in making all the bureaucratic aspects much more manageable than they were in reality.

I am greatly thankful to Dr Julie Arslanoglu and Dr Federica Pozzi (The Metropolitan Museum of Art, Department of Scientific Research) for the valuable help and contribution and the kind readiness to provide samples from wonderful and precious masterpieces. I also would like to thank Reba F. Snyder (Thaw Conservation Center, The Morgan Library & Museum) for the opportunity to investigate Gainsborough drawings and Maria Fredericks and Frank Trujillo (Thaw Conservation Center, The Morgan Library & Museum) to make possible the analyses on the Coptic manuscript.

Thanks to the MSAP group in Lille, where I spent my first year of PhD: Prof. Rolando, Fabrice, Stephanie, Violaine, Christophe. Thanks particularly to Amandine and Sergui for their important friendship. I hope to see you soon.
A big thanks to all the colleagues at the Proteomic platform in Bordeaux: Nicolas, Corinne, Katell, Butina, Stephane Cl., Jean-William, Anne Marie, Micheal, Landry, and Fatima. Thanks to Catherine Rawlins for her help in the top-down development and Stephan Ch. for his work with the HDX experiments. A special thanks to Catherine for editing any grammatical errors within this thesis. I wish you and Vaclav the best for your PhD.

A huge thanks to Cecile and Shay for being such great colocs in the last two years. After a difficult day at work, it was nice to go home and talk or watch horror movies with you two and Suzie. Thanks for all the soups, raclette, Camembert au four, and other delicious meals.

I am also very grateful to Yoann to have kindly accepted me as a third flatmate for the entire lockdown period and even after, during the thesis preparation.

I would also thank Martina and Anna, who provided the support and friendship I needed even from a distance.

I especially thank my mum and dad, who have always inspired me to do my best and have supported me in every decision I have made. Your unconditional love and care have helped me not only through my PhD but also through life. I am also grateful to my sister, who has always been a great point of reference for me and my endeavours. The last year has been extremely tough without seeing us, but you have always been close to me. Vi voglio bene.

A special thanks to Vincent, the most understanding person I have ever been lucky to meet. He has supported me throughout the last two years, believing in my capacities and giving me the strength not to break down.

New developments in high-resolution mass spectrometry for proteomic analysis applied to cultural heritage: study of intact proteins, their cross-linking and interaction in artworks and museum objects.

Abstract

Proteins in cultural heritage objects constitute a critical source of information. Their detection and characterisation can provide an accurate comprehension of the artist's technique and formulations. A more in-depth insight into the state of conservation and history of the artwork can also be achieved by investigating the molecular level of the degradation mechanisms induced by natural ageing, environmental factors, or either by inappropriate conservation/restoration treatments.

Firstly, the PhD was dedicated to developing proteomic bottom-up and top-down strategies using mass spectrometry to study proteins at trace levels from historical and artistic objects. The potentialities of a combination of these two complementary approaches were highlighted in the study of Gainsborough's drawings, where the presence of milk-based fixatives was revealed with samples collected with minimally invasive techniques. Information achieved with the peptidic analysis, such as the specific breed origin of the detected proteins and their modifications, was enriched through top-down analysis to detect highly modified proteoforms characterised by multiple cleavage patterns.

Secondly, the PhD study aimed to better insight into the proteins' structural and conformational alterations in an artwork. A strategy based on the elaboration of data from the bottom-up analysis was optimised to investigate the protein networks formations through the localisation and characterisation of cross-linked peptidic pairs. The methodology was initially tested on mock-ups of paintings (formulated with lysozyme mixed with lead white pigment) treated with oxidising agents and naturally aged. The detection of different reticulated products led to defining various molecular patterns characteristic of oxidative-based cross-links in the lysozyme protein, which subsequently were detected in more complex samples from historical tempera paintings.

The cross-linking examination, combined with an unbiased modification search, was also effective for a more exhaustive understanding of a Coptic manuscript's conservation history subjected to ancient invasive restoration treatment.

The detection of a great extent of lysine methylations and especially the characteristic fragmentation markers of formaldehyde-based cross-links provided the first analytical evidence of a potential parchment consolidation treatment based on gelatin-formol used by the Vatican library during restoration. The developed approach also offered a more accurate protein identification detecting peptides that passed unnoticed in the classical strategy because they were chemically modified or structurally unreachable.

Finally, the action of some of the most common inorganic pigments on proteins molecular and structural changes was also investigated. Particularly, the role of these inorganic compounds in proteins conformational changes was investigated for the first time through Hydrogen/Deuterium exchange (HDX) studies via mass spectrometry (intact protein approach). The decrease of deuterium exchange observed for certain mixtures suggested the pigment's interposition in the protein solvent accessibility or/and the modification of the molecular conformation.

Key words:

- High resolution mass spectrometry
- Bottom up and top down mass spectrometry
- Artworks, manuscripts and museum objects
- Crosslinking and interactions
- Chemical and posttranslational modificaitons

Nouveaux développements en spectrométrie de masse de haute résolution pour l'analyse protéomique appliquée au patrimoine culturel: étude des protéines intactes, de leurs réticulations et leurs intéractions dans les œuvres d'art et les objets de musées.

Résumé:

Les protéines présentes dans les objets du patrimoine culturel constituent une source d'information essentielle. Leur détection et leur caractérisation peuvent fournir une compréhension précise de la technique et des formulations de l'artiste. Un aperçu plus approfondi de l'état de conservation et de l'histoire de l'œuvre d'art peut également être obtenu grâce à l'étude au niveau moléculaire des mécanismes de dégradation induits par le vieillissement naturel, les facteurs environnementaux ou soit par des traitements de conservation / restauration inappropriés.

En premier lieu, le doctorat était dédié au développement de stratégies protéomiques bottom-up et top-down utilisant la spectrométrie de masse pour l'étude des protéines à l'état de traces à partir d'objets historiques et artistiques. Les potentialités d'une combinaison de ces deux approches complémentaires ont été mises en évidence dans l'étude des œuvres de Thomas Gainsborough, où la présence de fixateurs à base de lait a été révélée avec des échantillons prélevés avec des techniques minimalement invasives. Les informations obtenues par l'analyse peptidique, telles que l'origine biologique des protéines détectées et leurs modifications, ont été enrichies par une analyse top-down avec la détection de protéoformes hautement modifiées caractérisées par des motifs de clivage multiples.

Ensuite, le projet visait à mieux comprendre les modifications structurales et conformationnelles des protéines dans une œuvre d'art. Une stratégie d'analyse bottom-up visant à étudier les réseaux protéiques a été mise au point, intégrant en particulier la localisation et la caractérisation de paires peptidiques réticulées. La méthodologie a été initialement testée sur des peintures modèles (formulées avec du lysozyme mélangé à un pigment blanc de plomb) traitées avec des agents oxydants ou vieillies naturellement.

L'étude de différents produits réticulés du lysozyme a conduit à définir des modèles moléculaires caractéristiques du stress oxydants. Ceux-ci ont ensuite été détectées dans des échantillons plus complexes de peintures tempera historiques.

L'examen des réticulations protéiques a également permis d'obtenir de nouvelles informations sur la composition biologique et la conservation d'un manuscrit Copte ayant été soumis à un ancien traitement de restauration invasif.

La détection importante de méthylation des lysines et des marqueurs de fragmentations caractéristiques des réticulations à base de formaldéhyde, a fourni la première preuve analytique d'un traitement potentiel de consolidation du parchemin à base de gélatine-formol utilisé par la bibliothèque du Vatican pendant la restauration. L'approche développée a également offert une identification plus précise des protéines en détectant les peptides non identifiés par la stratégie classique car modifiés chimiquement ou structurellement inaccessibles. Enfin, l'action de certains pigments inorganiques courants sur les changements moléculaires et structuraux des protéines a également été étudiée. En particulier, le rôle de ces composés inorganiques dans les changements de conformation des protéines a été examiné pour la première fois par des études d'échange hydrogène / deutérium (HDX) par spectrométrie de masse (approche de protéine intacte). La diminution de l'échange de deutérium observée pour certains mélanges suggère l'interposition du pigment dans l'accessibilité aux solvants protéiques, ou / et la modification de la conformation moléculaire.

Mots clés:

- Spectrométrie de masse de haute résolution
- Analyses protéomiques «bottom up» et «top down»
- Œuvres d'Art, manuscrits et Objets de musée
- Réticulations protéiques et interactions
- Modifications chimiques et posttraductionnelles

UNITÉ DE RECHERCHE

Institute of Chemistry and Biology of Membrane and Nano-Objects (CBMN)
UMR n° 5248 CNRS

Spectrométrie de masse des macromolécules biologiques

University of Bordeaux
146, rue Léo Saignat
Bâtiment Plateforme Génomique Fonctionnelle
33076 Bordeaux Cedex
France

Résumé détaillé

Cette thèse de doctorat visait à approfondir la compréhension des composés protéiniques dans les œuvres d'art à travers la mise en œuvre de diverses stratégies basées sur la protéomique par spectrométrie de masse, dont certaines non encore adoptées dans les études du patrimoine culturel. Les diverses méthodologies proposées visaient à identifier les protéines à partir d'échantillons à l'état de trace ainsi qu'à divulguer de nouvelles informations moléculaires qui n'avaient pas encore été pleinement explorées dans les études sur le patrimoine culturel.

Le **chapitre I** donne un aperçu introductif du contexte de recherche de cette étude doctorale. Les protéines sont une source essentielle d'informations dans les études d'art, d'artefacts archéologiques et paléontologiques. L'étude et la caractérisation de ces molécules permettent une connaissance précise du matériau, des informations sur la technique de fabrication de l'artiste et sur l'état de conservation de l'objet. Les principales techniques d'analyse, à la fois non et micro-invasives, mises en œuvre dans l'étude des protéines du patrimoine culturel ont été brièvement discutées. L'accent est ensuite mis sur les grandes lignes de la stratégie protéomique combinée aux analyses par spectrométrie de masse qui sont réalisées dans cette thèse de doctorat. Cette approche a récemment émergé comme une approche réussie pour une compréhension approfondie des œuvres d'art anciennes, fournissant l'identification des protéines ainsi que leur détermination fiable des espèces et la caractérisation de leurs modifications. Le chapitre se termine par une vue d'ensemble de l'instrumentation de spectrométrie de masse utilisée pour l'analyse.

Le **chapitre II** traite de la protéomique Bottom-up (BUP) qui est la stratégie dominante pour la caractérisation des protéines et l'étude de leurs modifications post-traductionnelles. Depuis les résultats pionniers dans les années 2000, cette stratégie basée sur la digestion des protéines avant l'analyse spectrométrique de masse, s'est grandement améliorée dans l'étude des protéines anciennes menant à une compréhension plus approfondie des échantillons artistiques et historiques. Néanmoins, les recherches sont toujours à la recherche d'améliorations supplémentaires pour réduire l'impact de l'échantillonnage sur l'œuvre d'art, améliorer la sensibilité de l'analyse et limiter les interférences externes.

Dans ce chapitre, une analyse méthodologique ascendante optimisée a été discutée en application de l'étude de deux œuvres d'art protéiniques différentes (i) des peintures murales de la région de Nubie (vallée du Nil moyen) datées entre le 6ème et le 14ème siècle après JC, et (ii) des dessins de paysages réalisés par l'artiste anglais Thomas Gainsborough au XVIIIe siècle. L'objectif commun était la réalisation d'une identification simple du contenu protéique malgré l'hétérogénéité significative et les conditions de dégradation des œuvres d'art et la faible quantité d'analyte collecté. Dans l'étude des dessins de Gainsborough, la concentration de l'échantillon a été particulièrement réduite grâce à la mise en œuvre de deux techniques mini-invasives consistant en un frottement doux de la surface à travers des gommes en PVC et des films de polissage fins. De plus, un examen attentif a été mené sur les protéines PTM pour comprendre l'état de conservation des objets.

Le **chapitre III** décrit une nouvelle méthodologie basée sur la protéomique top-down pour l'identification des protéoformes à partir de la collection d'analytes traces dans les œuvres artistiques et historiques. Le protocole présenté, établi avec des caséines standard disponibles dans le commerce (caséines alpha et bêta), consiste en une préparation d'échantillon basée sur un filtre optimisée pour éliminer les substances interférentes sans diminution de la concentration d'analyte, suivie d'une analyse MS / MS en mode intact avec fragmentation EThcD. La stratégie a également été évaluée sur des échantillons de colle à base de caséine plus anciens (datés de 1900 à 1950) conduisant à la détection de diverses protéoformes de caséine tronquées caractérisées par des modifications caractéristiques des processus de vieillissement et de dégradation, telles que l'oxydation et la désamidation. Enfin, le protocole a été mis en œuvre dans l'étude du dessin de Gainsborough «Paysage avec cheval et charrette descendant une colline» dont l'échantillon a été collecté avec une technique mini-invasive. Malgré la quantité incroyablement infime de l'échantillon, plusieurs protéoformes de caséines, à la fois tronquées et hautement modifiées, ont été identifiées. Les résultats suggèrent l'application du lait comme fixateur de pigment, soutenant les écrits de l'artiste et les données ascendantes sur le même dessin. En outre, de nouvelles connaissances ont été obtenues sur l'espèce biologique du lait et sur la dégradation des échantillons, prouvant la complémentarité des approches ascendantes et descendantes et montrant que le TDP pourrait fournir une enquête plus approfondie sur les protéines et les PTM dans les artefacts artistiques.

Dans le **chapitre IV** sont présentées les études réalisées pour améliorer la compréhension des altérations structurelles et conformationnelles des protéines dans le liant protéinique appliqué dans les peintures à la détrempe.

Deux techniques innovantes, non encore réalisées dans les études du patrimoine culturel, ont été menées en combinant l'analyse spectrométrique de masse avec l'échange hydrogène / deutérium (analyse des protéines intactes) et avec des études de réticulation (approche ascendante). Les études ont été réalisées sur un système modèle de lysozyme pour simplifier la complexité de la composition en protéines de l'œuf. Les études HDX-MS visaient à enquêter sur le rôle de certains des pigments inorganiques les plus courants (blanc de plomb / blanc de zinc / cinabre / plomb rouge) sur les changements moléculaires et structurels des protéines, avant le vieillissement des protéines. La dynamique de l'échange isotopique dans le lysozyme non pigmenté a été comparée aux résultats de modèles de peinture à base de lysozyme pour avoir un aperçu des différences structurelles des protéines. Des informations complémentaires sur le vieillissement des protéines et les dégradations par oxydation ont été obtenues en effectuant des études de réticulation avec analyse MS. Une nouvelle stratégie basée sur l'élaboration de données à partir des analyses ascendantes a été optimisée pour étudier les formations des réseaux protéiques à travers la localisation et la caractérisation de paires peptidiques réticulées. La méthodologie a été initialement testée sur des maquettes de peintures (formulées avec du lysozyme mélangé à un pigment blanc de plomb) qui ont été traitées avec des agents oxydants et vieillies naturellement. La détection de différentes formations réticulées a conduit à définir divers modèles moléculaires caractéristiques des réticulations à base d'oxydation dans le lysozyme, qui ont ensuite été détectés dans des échantillons plus complexes de peintures à la tempera historiques.

Le **chapitre V** traite des modifications chimiques et structurelles qui peuvent se produire dans les protéines d'œuvres d'art historiques et artistiques suite à des causes anthropiques, telles que par exemple l'introduction de produits réactifs lors de traitements de conservation. L'étude de ces produits de dégradation pourrait fournir une vision plus complète de l'œuvre d'art et de son histoire de conservation. Dans ce chapitre, les analyses protéomiques se sont étendues au-delà de l'identification classique des protéines ascendantes, en examinant les modifications chimiques et structurelles des composés protéiniques induites dans un manuscrit préalablement soumis à des traitements de restauration invasifs. L'application d'une recherche de modification non biaisée suivie d'une étude structurale des paires de peptides réticulés a fourni les premières preuves analytiques de l'application de formol lors des traitements de conservation des parchemins dans la bibliothèque du Vatican. La recherche a également démontré que l'approche plus globale peut affiner l'interprétation des résultats en atteignant des données qui ne sont pas détectées dans l'approche standard parce qu'elles sont chimiquement modifiées ou structurellement inaccessibles.

Abbreviations

ABC: Ammonium bicarbonate
ACN: Acetonitrile
aDNA: Ancient deoxyribonucleic acid
AF: Acid formic
AGC: Automatic gain control
ALC: Average local confidence
BUP: Bottom-up proteomics
CH: Cultural heritage
ChaFRADIC: Charged-base fractional diagonal chromatography
CHAPS: 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate
CID: Collision induced dissociation
D_2O: Deuterated solvent
Da: Dalton
DDA: Data-dependent acquisition
DDT: 1,4-Dithiothreitol
DIA: Data-independent acquisition
DCA: Deoxycholic acid
DHA: Dehydroalanine
DTT: Dithiothreitol
ECD: Electron-capture dissociation
eFasp: Enhanced filter aided sample preparation
ER-FTIR: External reflection–Fourier transform infrared spectroscopy
ESI: Electrospray ionization
ETD: Electron-transfer dissociation
EThcD: Electron-Transfer/Higher-Energy Collision Dissociation
ev: Electron volts
eZooMS: Electrostatic Zooarchaeology by Mass Spectrometry
FDR: False discovery rate
FFPE: Formalin-fixed paraffin-embedded tissues
FT-IR: Fourier-transform infrared spectroscopy
FT-ICR: Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry
FWMH: Full width at half maximum
GC: Gas chromatography
GELFrEE: gel-eluted liquid fraction entrapment electrophoresis
g/L: Gramme for litre
H_2O_2: Hydrogen peroxide
H_2O: Water
HCCA: α-Cyano-4-hydroxycinnamic acid
HCD: Higher-energy C-trap dissociation
HCl: Hydrochloric acid
HDX-MS: Hydrogen deuterium exchange mass spectrometry
HgS: Cinnabar pigment
HMW: High molecular weight
HILIC: Hydrophilic interaction liquid chromatography
HTS: high throughput DNA sequencing techniques
IAA: Iodoacetamide
IRMPD: Infrared multiple photon dissociation
kDa: Kilo Dalton
LC: Liquid Chromatography

LC-MS/MS: Liquid chromatography combined with mass spectrometry analysis
m: Meter
M: Molaire
m/z: Mass/charge ratio
MALDI: Matrix-assisted laser desorption ionisation
MFGM: Milk fat globule membrane proteins
mg: Milligram
mg/mL: Milligram for millilitre
MHz: Megahertz
min: Minute
mL: Milliliter
MS: Mass spectrometry
MS/MS or MS2: Mass spectrometry tandem
nanoESI: Nano-Electrospray Ionization
NCBI: National Center for Biotechnology Information
NGS: Next Generation Sequencing
$2PbCO_3*Pb(OH)_2$: Lead white pigment
Pb_3O_4: Red lead/ minium pigment
PBS: Phosphate-buffered saline
PCA: Principal component analysis
pH: Potential of hydrogen
pmol/µL: Picomole for microliter
PMF: Peptide mass fingerprinting
PNGase F: Peptide:N-glycosidase F
ppm: Parts per million
PQI: Parchment quality index
PSM: Number of peptide spectrum matches
PTMs: Post-translational modifications
Q: Quadrupole
Q-TOF: Quadrupole Time-of-Flight
NMR: Nuclear magnetic resonance
NMR-MOUSE: nuclear magnetic resonance–mobile universal surface explorer
ROS: Reactive oxygen species
RP-LC: Reverse phase liquid chromatography
RT: Retention time
s: Seconds
SCX: Strong cation exchange
SDS: Sodium Dodecyl Sulphate
SDS- PAGE: Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis
SEC: Size-exclusion chromatography
SH: Sulfhydryls
TCEP: tris(2-carboxyethyl)phosphine
TDP: Top-down proteomics
TFA: Trifluoroacetic acid
TLC: Thin-layer chromatography
ToF: Time of flight
ToF-SIMS: Time-of-Flight Secondary Ion Mass Spectrometry
Tris:tris(hydroxymethyl)aminomethane
µg: Micrograms
µL: Microlitre
UV-VIS: Ultraviolet-Visible Spectroscopy
UVPD: Ultraviolet Photodissociation
V: Volt
v/v: Volume/Volume
w/v: Weight/Volume
XL-MS: Cross-Linking Mass Spectrometry
XRF: X-ray fluorescence
ZooMS: Zooarchaeology by Mass Spectrometry

Table of contents

ABSTRACT I
RESUME: III
RESUME DETAILLE VI
ABBREVIATIONS IX
TABLE OF CONTENTS XI
LIST OF FIGURES XVII
LIST OF TABLES XXII
LIST OF ANNEXES XXIV

CHAPTER I PROTEIN INVESTIGATION IN THE CULTURAL HERITAGE FIELD: FOCUS ON MASS SPECTROMETRY-BASED PROTEOMICS 1

ABSTRACT 2
1. Proteins in cultural heritage artefacts 3
2. Protein investigation in cultural heritage 4
3. Mass spectrometry (MS)-based proteomics 6
3.1 Introduction 6
3.2 Mass spectrometry instrumentation 7
3.2.1 Ion source: Electrospray ionisation ESI and Nanoelectrospray (NanoESI) 7
3.2.2 Orbitrap mass spectrometer 9
Principles of Orbitrap mass analyser 9
Ion capture 11
Ion detection 11
4. Fragmentation in tandem mass spectrometry (MS/MS) 12
4.1 HCD: Higher-energy Collisional Dissociation 14
4.2 Electron-Transfer/Higher-Energy Collision Dissociation (EthcD) 15

CHAPTER II BOTTOM-UP PROTEOMIC ANALYSIS OF HISTORIC AND ARTISTIC SAMPLES 26

ABSTRACT 27
1. Bottom-up approach 28
2. Bottom-up workflow 30
2.1 Sampling 30
2.2 Sample preparation 31

2.3 Protein digestion 31
2.4 LC-MS/MS analysis 32
2.5 Data acquisition 33
2.6 Bioinformatics analysis 33
3. Aim of the research: implementation of bottom-up proteomics in the protein investigation from wall painting and drawing samples 34
4. Methodology 35
4.1 Sample preparation: eFasp procedure 35
4.2 Nano-LC- MS/MS 38
4.3 MS- data processing 38
5. Analysis of wall painting samples from the Nubian region 39
5.1 Historic samples 40
5.2 Analysis and discussion of the results 40
5.2.1 Egg proteins 40
5.2.2 Collagen proteins 44
6. Analysis of Thomas Gainsborough drawings collected with the micro-invasive sampling methods 48
6.1 Introduction 48
6.2 Milk proteins 49
6.3 Samples 51
6.3.1 Sampling method 51
6.3.2 .Standard samples 51
6.3.3 Gainsborough's drawings 52
6.4 Analysis and discussion of the results 52
6.4.1 Standard samples 52
6.4.2 Gainsborough's drawing samples 55
7. Conclusions 61

CHAPTER III TOP-DOWN PROTEOMIC ANALYSIS OF HISTORICAL AND ARTISTIC SAMPLES 70

ABSTRACT 71
1. Top-down approach 72
2. Top-down workflow 73
2.1 Extraction and fractionation 73

2.2 Mass spectrometry analysis 76
2.2.1 Ionisation methods 76
2.2.2 Mass analysers 76
2.2.3 Fragmentations 77
2.3 Data elaboration 77
3. Top-down proteomics applied to cultural heritage studies 80
4. Aim of the research 81
5. Materials 81
5.1 Standard samples implemented for the protocol development 81
5.2 Casco industrial glue 81
5.3 Artistic and historical sample: Thomas Gainsborough drawing 82
6. Methodology 84
6.1 Development of sample preparation based on eFASP protocol 84
6.2 Fractionation (Liquid Chromatography) 86
6.3 MS and MS/MS Methods 86
6.4 Data processing and analysis 87
7. Analysis and discussion of the results 88
7.1 Standard samples implemented for the protocol development 88
7.2 Casco industrial glue 91
7.3 Artistic and historical sample: Thomas Gainsborough drawing 95
8. Conclusion 100

CHAPTER IV PROTEIN STRUCTURAL ANALYSIS IN ARTWORKS BY MASS SPECTROMETRY: HYDROGEN/DEUTERIUM EXCHANGE AND CROSS-LINKING INVESTIGATION 108

ABSTRACT 109
1. Introduction 110
2. Aim of the research 111
3. Hydrogen deuterium exchange combined with mass spectrometry (HDX-MS) 112
4. Hydrogen/Deuterium exchange (HDX) studies via mass spectrometry of painting models formulated with lysozyme and different pigments 115
4.1 Methodology 115
4.1.1 Samples 115
4.1.2 Sample preparation 115

4.1.3 HDX-MS analysis 116
4.2 Analysis and discussion of results 117
4.2.1 Mass spectrometry results at t0 (no incubation in deuterated water) 118
4.2.2 Mass spectrometry results following deuterium incubation at an increasing rate .. 122
5. Investigation of protein cross-linking 128
5.1 Protein cross-linking 128
5.2 Principal causes of cross-linking formations 128
5.2.1 Protein cross-links induced by oxidative reactions 130
Tyrosine-Tyrosine cross-links 130
Tryptophan-trypthophan cross-link 131
Histidine-lysine cross-links 132
5.3 Mass spectrometry-based proteomics for cross-linking investigation 133
5.3.1 Challenges in the cross-link analysis to overcome 134
5.3.2 Fragmentation of cross-linked peptides 136
5.3.3 Data analysis of cross-linked peptides 139
6. Study of protein cross-linking in art paintings using mass spectrometry-based bottom-up proteomics 140
6.1 Materials 140
6.1.1 Formulation of mock-up paints 140
6.1.2 Artistic and historical tempera paintings 140
6.2 Methodology 141
6.2.1 Extraction of proteins from standard painting models 141
6.2.2 Separation with SDS-PAGE (standard painting models) 141
6.2.3 Digestion in gel (standard painting models) 141
6.2.4 LC-MS/MS analysis 142
6.2.5 Data analysis 142
6.3 Analysis and discussion of the results 143
6.3.1 Model paintings 143
Di-tyrosine cross-linking 145
Di-tryptophan cross-linking 147
Tyrosine-tryptophan cross-linking 149
Histidine- lysine cross-linking 150
6.3.2 Artistic and historic tempera paintings 152
7. Conclusions 155

CHAPTER V STUDY OF PROTEIN CROSS-LINKING USING MS: MOLECULAR EVIDENCE OF RESTORATION TREATMENTS APPLIED TO HISTORICAL MANUSCRIPTS 164

ABSTRACT 165
1. Introduction 166
2. Parchment 166
2.1 General aspects 166
2.2 Analytical techniques implemented in parchment analysis 169
3. Impact of ancient restoration treatments on the analysis results 171
3.1 Example of an invasive restoration treatment: gelatin-formol 174
4. Formaldehyde reaction with proteins 175
4.1 Formaldehyde action on collagen proteins 175
4.2 Formaldehyde as a cross-linker: theory and mechanism 176
4.3 Structural study of formaldehyde cross-linking in a model protein 179
4.3.1 Sample preparation and analysis 180
4.3.2 Analysis and discussion of the results 180
5. Proteomic analysis of a formaldehyde glue treatment of a Coptic manuscript 183
5.1 Introduction of Coptic literature 183
5.2 Hamuli collection of Coptic codes 184
5.3 Methodology 185
5.3.1 Samples 185
5.3.2 Sample preparation 185
5.3.3 nLC-MS/MS analysis 185
5.3.4 Data Analysis 186
5.4 Protein/Species Identification 186
5.4.1 Data analysis 186
5.4.2 Results 187
Bovinae subfamily-Bos Taurus *187*
Caprinae subfamily - Capra hircus *188*
Oryctolagus cuniculus *191*
5.5 Investigation of proteins modifications 195
5.5.1 Data analysis 195
5.5.2 Results from the analysis 195
5.6 Cross-links investigation and localisation 198

5.6.1 Data analysis 198
5.6.2 Result analysis 199
Identification of reticulated peptides characteristic for a particular biological species 200
Interactions among collagens from different biological breeds 205
Interactions among different collagens from the same biological breed 207
5.7 Discussion of the results achieved from the three investigations 208
6. Conclusions 209

CONCLUSION AND PERSPECTIVE 225
ANNEXE: SAMPLES INVESTIGATED 228

List of Figures

Figure I-1 Schematic representation of a mass spectrometer. 7
Figure I-2 Principles of electrospray ionisation (positive detection mode). 8
Figure I-3 A cut-away model of the Orbitrap: a) central electrode, b) outer electrode, c) simulation of the ions in red. Figure extracted from Scigelova and Makarov, 2006. 10
Figure I-4 Scheme of Q Exactive plus and of Orbitrap Fusion Lumos. 12
Figure I-5 Potential fragmentations in the peptide backbone with the corresponding nomenclature of the forming fragment ions and the MS/MS fragmentation. Figure elaborated from Zhurov et al., 2013. 14
Figure II-1 Scheme of a bottom-up proteomic strategy. 28
Figure II-2 Scheme of bottom-up eFasp methodology. 37
Figure II-3 MS/MS spectrum of (A) the triply charged ion at m/z 761.0650 (Δm= -0.7 ppm) belonging to the ovalbumin in the sample Gh_01 (Ghazali Monastery, North Church dated second half of the 7th century). (B) the doubly charged ion at m/z 878.4036 (Δm= 1.2 ppm) related to lysozyme peptide 46-61 in sample D_BV_6. 41
Figure II-4 Lysozyme sequence coverage from the sample Gh_01 (red pigment area). The sample is from the Ghazali Monastery, North Church, in the second half of the 7th-century. 42
Figure II-5 MS/MS of doubly charged precursor ion at m/z 663.3191 (Δm= 0.4 ppm) related to the peptide 40-51 of the lysozyme from sample D_H_A_18 (Old Dongola, House A 8th century). 43
Figure II-6 MS/MS of (A) doubly charged ion at m/z 654.8245 (Δm= -1.0 ppm) attributed to ovotransferrin in the sample US. 22_8c (church on Us Island, 10th/11th century). (B) doubly charged ion m/z 723.8826 (Δm= 0.1 ppm) attributed to the peptide ^{642}VGATGEIFVVnSPR^{655} of vitellogenin 1 from the sample Se_5_2 (Selib, end of 6th/7th century). 44
Figure II-7 MS/MS spectrum of the doubly charged ion at m/z 443.7227 (Δm= 0.6 ppm) from the sample US. 22_8c white (church on Us Island, 10th/11th century). The ion was assigned to the peptide ^{361}GSEGPQGVR^{369} of the α1 type I collagen. 45
Figure II-8 Distribution of samples under investigations and the detected proteinaceous compounds 46
Figure II-9 Milk fat globule membrane. Figure extracted from Lordan et al., 2017. 50
Figure II-10 Sampling techniques using PVC –free erasers and fine polishing films. 51
Figure II-11 MS/MS spectra from the sample A (mock-up standard, PVC eraser of whole milk) (A) of the doubly charged ion at m/z 976.4818 (Δm=2.0 ppm) attributed to $_{\alpha S1}$-CN (119 – 134) peptide. (B) of the doubly charged ion at m/z 818.3922 (Δm=1.4 ppm) attributed to beta- lactoglobulin (141 -154) peptide. 53

Figure II-12 MS/MS spectra from the sample S1C2 (2014.32): (A) of the di-charged charged ion at m/z 999.0391 (Δm=1.4 ppm) presenting y and b fragments of kappa casein 90-107 peptide. The peptide sequence presents oxidation on tryptophan W98 deamidation on asparagine N103; (B) of the di-charged charged ion at m/z 854.9502 (Δm=1.5 ppm) presenting y and b fragments of the lactadherin peptide ^{233}INLFDTPLETQYVR^{246}..... 56
Figure II-13 MS/MS spectra of peptides identified in sample S2C2 from Coastal Scene with Shipping, Figures, and Cows (ca. 1780; 2017.89). (A) The doubly charged ion at m/z 831.3934 (Δm=0.8 ppm) enabled the identification of the 121-134 peptide of alpha S1 casein; (B) The triply charged ion at m/z 694.0209 (Δm=0.5 ppm) enabled the identification of the 118-134 peptide of alpha S1 casein; (C) MS/MS spectrum from the reference sample of modern milk subject to heating..... 58
Figure III-1 Schematic representation of top-down vs bottom-up approaches. Figure elaborated from Catherman et al., 2014. 72
Figure III-2 Casco Powdered Casein Glue. 82
Figure III-3 Landscape with Horse and Cart Descending a Hill. The sample collected for the top-down analysis is the III, 63_S1B2 (6 μm grit).. 83
Figure III-4 Scheme of top-down eFasp methodology. 85
Figure III-5 70-minute gradient curve and flow rate. 86
Figure III-6 NanoLC nano ESI-Orbitrap MS spectrum of commercially available casein standards investigation.. 89
Figure III-7 NanoLC nanoESI-Orbitrap spectra related to Casco glue sample. (A) TIC chromatogram with a focus on casein S1. (B) the charge state distribution selected for the MS2 (5^+ m/z 637.519, 5^+ m/z 815.643 and 7^+ m/z 1401.244, respectively). At the bottom, the fragment maps are reported with the matching fragment ions and the modifications' localisation.. 92
Figure III-8 EThcD spectra of the 5^+ charged ion m/z 815.643 from the Casco glue, presenting dual ion series (z/c and y/b) of the proteoform α_{S1}CN (25-60) with a deamidation localised on the asparagine residue (N32). 93
Figure III-9 MS spectrum of hydrolysed κ-CN forms of κ-CN variant A (86-190) and its oxidised forms identified using a top-down approach applied to a sample taken from a historic Gainsborough drawing.. 96
Figure III-10 The isotopic pattern of the 8^+ charged ion of the truncated proteoform κ-CN variant A (86-190) modified by 1 oxidation, 2 phosphorylations and an increasing number of deaminations, eight, nine and ten respectively (A) theoretical patterns and (B) experimental patterns observed in MS spectra obtained from the analysis of the Gainsborough artwork sample. 97
Figure III-11 The truncated proteoform 86-190 of the κ-CN variant A provided the discrimination between *B.taurus* and *B.bison* species, which was not achieved with the BUP studies.. 98
Figure IV-1 Amino acid sequence constituted by H-L-S-C-K-D.. Figure elaborated from Oganesyan et al., 2018 using ChemDraw Prime. 112

Figure IV-2 Painting models formulated by mixing lysozyme with four different pigments often implemented in artistic painting formulations. 115
Figure IV-3 Total ion current (TIC) chromatograms of lysozyme unpigmented and painting models formulated with four pigments widely implemented in artistic productions: lead white, zinc white, cinnabar and red lead. All samples showed two main peaks of elutions. 118
Figure IV-4 MS spectra from the first main peak of elution in all samples under investigation at time zero (no deuterium exchange).. 119
Figure IV-5 MS spectra second main peak of elution in all samples under investigation at time zero (no deuterium exchange).. 121
Figure IV-6 MS spectra of the predominants proteoforms of the first and second main elution peaks in lysozyme with and without white lead pigment after 3600 seconds of incubation with the deuterated solution..... 123
Figure IV-7 Magnification of the peak of 100% detected in lysozyme with and without lead white pigment.. 124
Figure IV-8 Graphs of comparison between the deuterium uptake of unpigmented lysozyme and lead white painting model..... 124
Figure IV-9 Graphs of comparison between the deuterium uptake of unpigmented lysozyme and (A) zinc white model painting, (B) cinnabar model painting.. 126
Figure IV-10 Representation of the potential peptidic constituents obtained from the digestion of a cross-linked protein mixture. The illustration depicts the chains A-B and C of hen egg-white lysozyme and has been realised using UCSF Chimera software. 128
Figure IV-11 Proposed mechanism of Tyr-Tyr cross-links formation (di-tyrosine compound)..... 131
Figure IV-12 Proposed mechanism of Trp-Trp formation (di-tryptophan compound). Figure elaborated from Hägglund et al., 2018 with ChemPrime. 132
Figure IV-13 Proposed mechanisms of formation of His-His, His-Lys and His-Cys cross-links. 133
Figure IV-14 Potential fragmentation patterns in the presence of an intrapeptidic cross-link **(A)** and a pair of reticulated peptides **(B-C)**. The fragmentation might occur at the covalent bond **(B)** or within the peptide backbone **(C)**. Figure elaborated from Maes et al., 2017. 138
Figure IV-15 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of lysozyme protein standard and two naturally aged painting models (lysozyme/ lead white pigment). Two multiforms were observed for both aged samples (dimers and trimers). Furthermore, a tetramer form was highlighted in the mock-up, which was treated with pure H2O2. 144
Figure IV-16 (A) MS2 of the triply charged ion m/z 1031.4832 (Δm=2.19 ppm) attributed to the Tyr-Tyr (-2H) reticulated peptide pair ^{46}nTDGSTDY53GILQInSR^{61} - ^{22}GY23SLGNWVcAAK^{33}. (B) MS spectrum representing the oxidation forms of the triply charged ion m/z 1031.4832 (potential di-tyrosine).. 146

Figure IV-17 (A) MS/MS spectrum of the triply charged ion m/z 888.0911 (MW 2661.2501 Da Δm=1.20 ppm) attributed to the homodimeric form (^{22}GYSLGNW28VcAAK^{33})$_2$. (B) Table of the two series of y-ions fragments (detected with and without loss of -2.0156 Da). 148
Figure IV-18 MS/MS spectrum of the triply charged ion m/z 1191.5618 (MW 3571.6619 Da) from the untreated painting model. The value was attributed to the inter-peptide cross-link (Tyr-Trp) between the peptides ^{46}NTDGSTDY53GILQInSR^{61} and ^{97}KIVSDGNGMnAW108VAwR^{112}.. 149
Figure IV-19 MS/MS spectrum of the 4$^+$charged ion at m/z 374.1953 from the treated model painting. The values were attributed to the reticulated form of (^{15}H15GLDNYR^{21}) – (^{1}K1VFGR^{5}), having a covalent bond between His15 and Lys1 residues.. 150
Figure IV-20 The MS/MS spectrum of the triply charged ion at m/z 1169.5380 (MW 3505.5908 Da), detected in the tempera on wood (Dicredi). The charged ion was attributed to the homodimer (^{46}NTDGSTDY53GILQInSR^{61}) involving a covalent bond between two Tyr53 of different lysozyme proteins..... 153
Figure V-1 The hierarchical structure of collagen: the amino acid sequence is composed of a repetition of three amino acid residues (glycine-X-X2, which often are proline and hydroxyproline). Figure extracted from Nijhuis et all., 2019..... 168
Figure V-2 Proposed chemistry of the formaldehyde cross-linking reaction: Figures elaborated from Sutherland et al., 2008. 178
Figure V-3 Alternative proposed chemistry of the formaldehyde cross-linking reaction mechanism. Figure extracted from Tayri-Wilk et al., 2020. 179
Figure V-4 MS/MS spectrum of the 5+ charged ion m/z 66.7227. The fragmentation pattern strongly suggests the presence of an inter-peptidic cross-link between R112 of the residue α ^{98}IVSDGNGmnAWVAWRNR^{114} with the K116 of the peptide β ^{115}cKGTDVQAWIR^{125}.. 181
Figure V-5 MS/MS spectrum of the charged ion 4$^+$ m/z 858.1727 corresponding to the miss-cleaved peptide ^{97}KIVSDGNGmnAWVAWRNRcKGTDVQAWIR^{125}. 182
Figure V-6 MS/MS of the doubly charged precursor ion at m/z 1399.1604 attributed to the peptide (674 – 706) of collagen 2(I) from sample S1. attributed uniquely to *Bos taurus* species. 188
Figure V-7 MS/MS spectrum of the triply charged ion at m/z 1031.8372 (Δm 1.9 ppm) detected in the sample S1 attributed to the peptide 845-877 of collagen 2(I) specific uniquely for the *Capra hircus* species.. 190
Figure V-8 MS/MS spectrum of the doubly charged ion at m/z 649.8341 (Δm 1.8 ppm) detected in the sample A_S1 attributed to the peptide 341-354 of collagen 2(I) to either *Oryctolagus cuniculus* (European rabbit), *Suricata suricatta* (meerkat), or *Carlito syrichta* (Philippine tarsier). 191
Figure V-9 The calculated rate of non deamidated asparagine and glutamine per sample in collagen type 1 (COL1A1 and COL1A2)..... 196
Figure V-10 Pie charts representing the extent of different types of formaldehyde-based cross-links in the three collagen proteins under investigation..... 200

Figure V-11 Histograms representing the source of the cross-linked peptides identified in the three collagens under investigation. 201
Figure V-12 The MS fragmentation spectrum of the 4+ charged ion at m/z 717.6157 (Δm: -1.39 ppm) has been attributed to a formaldehyde-based cross-link between two peptides: ^{662}GAAGIpGGKGEKGETGLR^{673} and ^{644}GDIGSpGRDGAR^{661}. The peptide α (662-673) refers to the *Bovinae* subfamily (*Bos taurus, Bos indicus, Bos mutus and Bubalus bubalis*). The peptide β instead is not specific. 202
Figure V-13 Fragmentation of the triply charged ion m/z 820.4151 (Δm= -1.22 ppm) attributed to two peptides ^{490}GPAGPSGTPGEKGPAGER^{507} and ^{967}GSPGPQGVK^{975} bonded by a methylene bridge (+12 Da). The two peptides refer to collagen 1(III). The peptide α (490-507) is characteristic of *Oryctolagus cuniculus* (European rabbit) specie; the peptide β (967-975) is not unique but it does not refer to proteins from *Bovinae* (V 976-> I 976) or *Caprinae* subfamily (V 976-> I 977). 203
Figure V-14 Fragmentation of the ion m/z 915.4427 with a charge 4^+ (Δm 0.11ppm) has been attributed to ^{241}GLpGppGIKGpAGMpGFpGMKGHR^{264} and ^{1083}GppGpQGpRGDK^{1094}, presenting all proline oxidized and covalently bonded by a formaldehyde-based bridge. The two peptides are both from collagen 1(III) but not from the same species; even though none of the peptidic residues is unique for a particular specie, the peptide α (241-264) is attributable to *Oryctolagus cuniculus* but not to *Bovinae* (L242->F244 and I248->M250)or *Caprinae* (I248-> M251). The peptide β (1083-1094) instead might be attributed to *Bovinae* subfamily collagen but not to *Caprinae* (^{1084}GAPGPQGPRGDK^{1095}) or *O. Cuniculus* (^{1081}GAPGPQGPRGDK^{1092}). 206

List of Tables

Table II-1 Selection of proteins identified in the samples from the Nubian wall paintings analysed. 47
Table II-2 List of the milk proteins identified in the standard samples prepared with skim and whole milk on glass and paper surface and then sampled with two micro-invasive methods (PVC-free eraser and μm grit papers). 54
Table II-3 List of the main milk proteins identified in the Gainsborough drawings for selected locations. Values presented in the table are protein sequence coverages expressed as %, and the obtained -10logP protein score is above 40 for all the proteins listed. 57
Table II-4 List of the peptides indicative of the Bovinae subfamily. 60
Table III-1 Protein mixture fractionation in the top-down workflow. 75
Table III-2 Principal software implemented in top-down proteomics. 78
Table III-3 List of the casein proteoform identified in standard casein. 91
Table III-4 List of the principal casein proteoforms identified in Casco powdered industrial casein-glue dated 1900-1950 94
Table III-5 List of the casein proteoforms identified in Thomas Gainsborough drawing *"Landscape with Horse and Cart Descending a Hill"*. For the proteoforms detected manually (without EthcD fragmentation) is reported the peak m/z and the corresponding observed MW (Da) with a 100% abundance. 99
Table IV-1 Challenges encountered in a cross-linking investigation with mass spectrometry proteomics analysis and possible solutions to overcome them. 135
Table IV-2 Cross-links identified in natural aged painting mock-ups (lysozyme/ lead white) 151
Table IV-3 Cross-links identified in four historical tempera based artworks 154
Table V-1 Principal collagen-based glue implemented in the restoration treatments 173
Table V-2 Alignment between the peptide sequences of three Collagen α-1(III) from C. hircus, O. aries and B. taurus. 189
Table V-3 List of the identified proteins in the two samples S1 and A_S1 from the Coptic manuscript "Martyrdom of St. Pteleme" of the Hamuli collection 193
Table V-4 Table of the specie characteristic peptides identified in the two samples S1 and A_S1 from the Coptic manuscript "Martyrdom of St. Pteleme" of the Hamuli collection. 194
Table V-5 Most abundant modifications detected in the two samples. The incidence of a specific modification was calculated, dividing the peptide-to spectrum match (PSM) by the total frequency of the corresponding amino acid residue (%). 195
Table V-6 Unique peptides identified in the samples only as reticulated pair. 204
Table V-7 Interactions among different collagens from the same biological breed 207
Table V-8 List of the cross-links induced by formaldehyde detected in the proteins of collagen 2(I) of different biological breeds (*B. taurus*, *C. subfamily*, *O. cuniculus*, not specific and interactions). 210

Table V-9 List of the cross-links induced by formaldehyde detected in the proteins of collagen 1(III) of different biological breeds (*B. taurus*, *C. subfamily*, *O. cuniculus*, not specific and interactions). 212
Table V-10 List of the cross-links induced by formaldehyde detected in the proteins of collagen 1(I) of different biological breeds (*B. taurus*, *O. cuniculus*, not specific and interactions) 214

List of Annexes

Annexe i Investigated fragments of Nubian wall painting (Middle Nile Valley) dated between the 6th and 14th century AD from the kingdom of Makuria (Northern Sudan and Southern Egypt). Only one sample is from the Kingdom of Alwa (actual central and southern Sudan). The area of the sampling for the MS analyses and the colours are indicated on each fragment. The fragments were collected from different museum collections (including the Sudan National Museum in Khartoum, National Museum in Warsaw, Humboldt University in Berlin, Polish Centre of Mediterranean Archaeology, University of Warsaw) and supplied by Dr Dobrochna Zielinska (*Institute of Archaeology of the University of Warsaw*). 229

Annexe ii Investigated Gainsborough drawings (1750-1780) with the sampling localisation for milk identification via nano-LC-MS/MS. The minimally invasive technique implemented is defined as well. The samples were provided by Dr Julie Arslanoglu and Dr Federica Pozzi (*Department of Scientific Research, The Metropolitan Museum of Art, New York*) and conservator Reba Snyder (*Thaw Conservation Center, The Morgan Library&Museum*, *New York)*. 231

Annexe iii Egg-based historical artworks investigated to detect and localise oxidising-induced cross-links in proteins. The samples were provided by Dr Julie Arslanoglu, *The Metropolitan Museum of Art* (Department of Scientific Research) in New York. The sample D_BV_6 was supplied by Dr Dobrochna Zielinska (Institute of Archaeology of the University of Warsaw). 234

Annexe iv Investigated samples collected from the third page of the Coptic manuscript *"Martyrdom of St. Pteleme"*, which belongs to the Hamuli Collection. The samples were provided by Dr Julie Arslanoglu and Dr Federica Pozzi (*Department of Scientific Research, The Metropolitan Museum of Art, New York*) and conservators Maria Fredericks and Frank Trujillo (*Thaw Conservation Center, The Morgan Library&Museum, New York*). 235

Chapter I Protein investigation in the Cultural heritage field: focus on mass spectrometry-based proteomics

Abstract

The following chapter provides an introductory overview of the research context of this doctoral study. Proteins are a key source of information in the studies of art, archaeological and palaeontological artefacts. The study and characterisation of these molecules provide accurate knowledge of the material, informing the artist's manufacturing technique and state of conservation of the object. The principal analytical techniques, both non- and micro-invasive, implemented in the investigation of proteins in Cultural Heritage, have briefly been discussed. The focus is then given to the outlines of proteomic strategy combined with mass spectrometry analyses performed in this doctoral thesis. The approach has recently emerged as a successful approach for an in-depth understanding of ancient artworks, providing proteins' identification and reliable species determination and characterisation of their modifications. The chapter ends with an overall picture of the mass spectrometry instrumentation employed for the analysis.

1. Proteins in cultural heritage artefacts

Proteins are key molecules in the studies of art, archaeological and palaeontological artefacts [1, 2]. In fossils, ancient bones, and other ancient material such as dental calculus, proteins offer crucial information on ancient and extinct species, human evolution, past migrations and phylogenetic trees [3-7]. Proteins constitute or are trapped in ancient objects of daily-life activities, including clothing, tools, cosmetics or medicines, ceramic vessels or other containers. Their analyses can reveal valuable information on socio-cultural aspects, manufactory, diet, and ancient trade in past populations [8-15]. Protein-based materials were also extensively used in artistic production; e.g. egg, casein, animal glue binders used in wall and easel paintings or applied to stuccoes and sculptures, adhesive for gildings, writing supports (like parchment-based manuscripts), drawing sizing, but also adhesives, consolidates or waterproof coatings used for preservation [16-31].

The study and characterisation of these molecules provide accurate knowledge of the material, inform the artist's manufacturing technique, and state the object's conservation [21, 23, 27, 32-38]. The characterisation of the deterioration products may help in studying the deterioration causes and may assist the definition of adequate preservation strategies (i.e., restoration treatments, control of exposure and environmental parameters) [35, 39-41].

The characterisation of proteinaceous material is an analytical challenge for various reasons: the analyte is often embedded in a complex and heterogeneous matrix consisting of organic and inorganic substances (pigments, binders and other substances contaminants) that interact with each other. Furthermore, proteins are subject to ageing and to degradation processes related to their storage conditions, environmental factors (e.g. temperature, relative humidity, light exposure), interactions with other compounds (external or internal to the object), and invasive or unsuitable conservation conditions and restoration treatments [42-44]. Another critical issue in the study is the low amount of material available for the scientific analysis to guarantee the artefact's minimal invasiveness [1].

2. Protein investigation in cultural heritage

Various methods, either non- or micro-invasive, have been implemented to study proteins of Cultural Heritage samples. For instance, non-invasive spectroscopic techniques have been widely implemented to identify both inorganic and organic materials in artworks. Vibrational spectroscopy techniques such as Fourier-transform infrared (FT-IR) and Raman spectroscopies have been extensively employed, providing straightforward protein recognition by detecting characteristic protein functional groups such as amides [45-53]. Over recent years, significant advancements have been achieved in the instrumentation (e.g. portable instruments [54] or imaging and mapping microscopes [55, 56]) along with the development of new methods; for examples, Surface-enhanced Raman scattering (SERS) [57-59]. Recently, external reflection-Fourier transform infrared spectroscopy (ER-FTIR) has also been implemented to discriminate different paint binders, such as polysaccharides compounds from proteinaceous-based ones [60].

Another technique is the laser-induced fluorescence (LIF) that exploits aromatic amino acids' intrinsic fluorescence (Phe, Tyr, and Trp) to detect proteins. The consequent autofluorescence can also enable identifying degradation products and amino acids cross-linkage reactions [61, 62]. The technique provides a first insight into the material composition (several binders can be discriminated) without the need for sampling or object transportation [63, 64]. The development of multi-spectroscopic mobile devices constituted by LIF and laser-induced breakdown spectroscopy (LIBS) and portable Raman has been successfully applied to analyse binding media [65].

Gas chromatography, coupled to mass spectrometry (GC–MS) [66, 67] and pyrolysis-GC/MS [68-71], are the most applied in the investigation of organic compounds in artworks [16, 18, 25, 32, 72-75]. The hydrolysis of proteins and derivatisation of the produced amino acids followed by their analysis and calculation of their distribution/ratios provide the identification of proteins and their degradation by-products [76]. Developments in sample pre-treatments have also enabled the simultaneous or sequential determination of different analytes classes (e.g. proteinaceous binders, glycerolipids, natural waxes, terpenoid resins, polysaccharide media) investigating a single sample [66, 77]. With the pursuit of improving the study of proteinaceous materials characterised by high thermal stability and low solubility due to ageing and degradation [28, 69], pyrolysis Evolved Gas Analysis Mass Spectrometry (EGA/MS) has been recently optimised.

These techniques lead to a rapid inspection and evaluation of ancient proteins in artistic material without the need for analyte extraction (enabling the detection of less soluble fractions of analytes, such as aggregated proteins) [19, 78].

Similarly to GC-MS, Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS) analysis allows identifying proteins through the detection of amino acids [79]. The technique enables the localisation of proteinaceous compounds in a paint cross-section, which can help in the elucidation of the painter's technique [80] and the discrimination of the original paint area from the restored one [81].

Immunological techniques such as enzyme-linked immunosorbent assay (ELISA) [31, 82, 83] or immunofluorescence-Surface-Enhanced Raman Scattering (immuno-SERS) [58, 84] have acquired increased attention in cultural heritage studies. These methods are characterised by high sensitivity and reliability, allowing the discrimination and localisation (mapping) of target proteins in a heterogeneous matrix [31, 85-87]. The antibody exploits the antigen-antibody binding reaction to discriminate a specific protein target, also contributing to the assessment of the degradation. [88]. The development of portable instruments, such as chemiluminescent (CL) combined with the lateral flow immunoassay (LFIA) technology, provide further progress by enabling the simultaneous detection of more than one protein (for example, ovalbumin and collagen) directly in situ [89].

In the two last decades, proteomics combined with mass spectrometry (MS)-based analysis emerged as a successful tool for the reliable species identification and the characterisation of molecular alteration patterns in ancient and artistic objects [23, 90-93]. MS-based proteomics allows the identification of proteins and especially the characterization of the biological species of origin, which cannot be obtained with the other techniques mentioned. Beyond that, proteomics leads to a detailed study of protein modifications (PTMs) and interactions for a more-in depth knowledge of degradation processes related to ageing and specific conditions (storage, environmental factors, pollution, restoration treatments). Following the pioneering studies performed on a few micrograms of the samples [23, 90-92], MS-based proteomic strategies continue to increasingly develop to provide the protein characterisation from a minimal amount of samples (achievement of non (or minimally)-invasive sampling [24, 94]) as well as to transcend the protein identification, pursuing a higher understanding of the protein degradation [35, 36, 39, 95].

3. Mass spectrometry (MS)-based proteomics

3.1 Introduction

The term proteome coined in 1994 [96] refers to *"*the complete set of proteins expressed by the genome of a cell, a tissue, and an organism at a precise moment of its development and in a precise environment*"*. Contrarily to the genome, the proteome is not unique for every organism and proteins present in a cell, as well as their function, location, and structure are not stable, and they can change. Proteomics is the study of the proteome that aims to identify, characterise and quantify the proteins of a defined cell or system [97-99]. Additionally, the discipline pursues the investigation of protein isoforms, modifications (such as post-translational modifications, PTMs), their interactions, their structural organisations and the complexes in which they are involved [100].

In proteomics, mass spectrometry allows the determination of the exact molecular weight of proteins or their proteolytic fragmentation products (peptides) by measuring the mass-to-charge ratio (m/z) of produced gas-phase ions [101].

In a classic proteomic approach, proteins are extracted using soft procedures to avoid their hydrolysis into amino acids. If required, gel electrophoresis such as Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) can be conducted to separate complex proteins samples [102-104]. Before the mass spectrometry analysis, the macromolecules can be enzymatically or chemically cleaved into peptides or investigated as intact proteins. Furthermore, depending on the methodology implemented, the analyte solution can be analysed directly with the mass spectrometer (for example, with MALDI instruments) or formerly separated through liquid chromatography (LC-MS) [105, 106]. Once the analysis is completed, bioinformatics tools integrating protein and genomic databases assist the protein identification.

The mass spectrometry analysis can be performed by measuring the ion mass (MS analysis) [100, 107, 108], or it can also involve the fragmentation of the ions (MS/MS or tandem analysis). In this strategy, the mass determination of fragmented ions originating from peptide or proteins is provided. The MS/MS method is commonly referred to as a bottom-up approach [109, 110] to investigate peptides or a top-down approach if intact proteins are examined [111, 112].

3.2 Mass spectrometry instrumentation

A mass spectrometer is conventionally composed of 3 elements: an ionisation source, a mass analyser, and an ion current detector (Figure I-1). The molecules, in the liquid or solid phase, are firstly transferred to the gas phase and ionised. Then the ions are separated in an analyser according to their mass-to-charge ratio (m / z). The ions' movement in the analyser is ensured under electric and/or magnetic fields. The speed and trajectory of ions towards the detector hence depend on their characteristic m / z. Finally, the ions hit the detector that converts these events into an electrical signal.

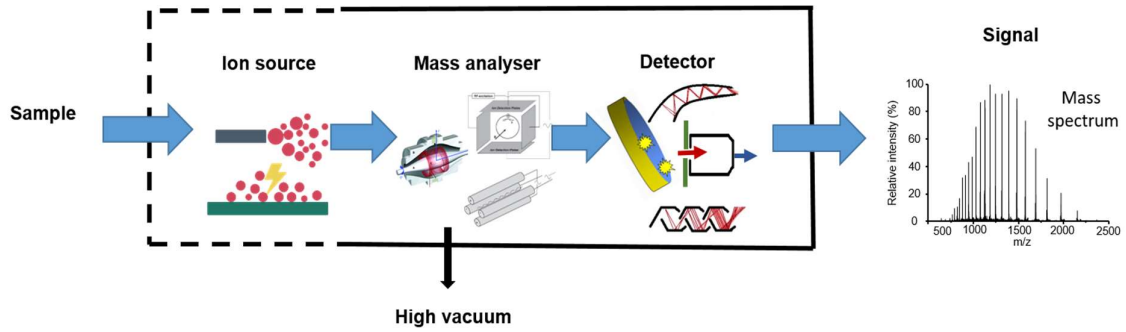


Figure I-1 Schematic representation of a mass spectrometer.

In the present doctoral thesis, the investigations were conducted using an ESI MS/MS high-resolution Orbitrap technology coupled with nano-flow HPLC (reversed-phase). The principles of ESI ionisation and Orbitrap technology will be described in the following sections.

3.2.1 Ion source: Electrospray ionisation ESI and Nanoelectrospray (NanoESI)

Developed by Fenn and Yamashita in the early 1980s (Nobel prize in 2002), Electrospray ionisation (ESI) is a soft ionisation method that ionises the molecules without fragmenting or degrading them [113, 114]. Similarly to Matrix-Assisted Laser Desorption Ionization (MALDI), this characteristic is principally employed in proteomics analyses, providing multi-charged ions.

The solution of LC eluant containing the analyte in ESI is transferred to the gas phase based on a desolvation/ionisation phenomenon (Figure I-2). The analytes are injected through a narrow metal capillary tube under atmospheric pressure.

The application of a high electric field (106 V.m-1) between the capillary and the counter-electrode induce the separation of opposite charges and the production of multi charged droplets. The generation of cations is generated in the presence of a negative electric field and vice-versa. The ion source's heat and the use of desolvation gas (often N2) promote evaporation of the solvent. The charge droplets' volume is reduced until the Rayleigh limit is reached: the Coulomb repulsion forces exceed the surface tension forces, and the droplets exploded into smaller droplets. This phenomenon called the Coulomb explosion recurs until the solvent's evaporation and the formation of ions in the gas phase.

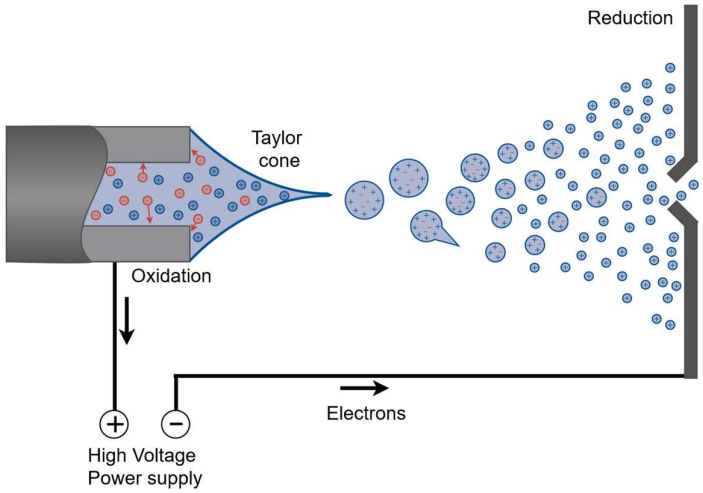


Figure I-2 Principles of electrospray ionisation (positive detection mode).

Two models are accepted to explain the production of desolvated ions in the gas phase: the residual charge model (CRM) describes the ions as progressively desolvated [115]. The alternative model proposes an “ion evaporation model” where ions can be directly desorbed from the droplets [116].

In ESI, the molecules with high molecular weight have multiple ionisable sites to form multiply charged ions. The smaller values of the measured m/z increase the instrument's sensibility and extend the mass range measured [114, 117, 118].

A major improvement in the performance of electrospray ionization arose from the possibility of reducing the flow rate of the solution used to create the spray. The first nanospray sources (20 nL.min-1) appeared in 1996 [119]. The small diameter of the capillary at low flow rates ensures the spray's stability through an easy formation of small droplets [120].

The miniaturization of ESI sources has the significant advantage of increasing the concentration of analytes that elute from a liquid chromatography column to enter the mass spectrometer, enhancing, in turn, the sensitivity of the analysis.

Furthermore, the use of low flow allows an increment of the time analysis and the measurement accuracy without increasing the sample volume injected [121].

3.2.2 Orbitrap mass spectrometer

The mass analyser is literally and figuratively the core of the mass spectrometer. In a proteomic analysis, the primary characteristic of an analyser is the sensitivity, resolution, mass accuracy, mass range and quality of the fragmentation spectra. The analysers most commonly used are the ion traps (IT), the time-of-flight analyser (TOF), the quadrupole analyser (Q), and the Fourier transform analysers (FT-ICR, Orbitrap) [118, 122-130]. Their combination resulted in hybrid instruments that leverage each of the assembled analysers [131].

The roots of Orbitrap rely on the principle of orbital trapping (charged particles can be trapped in electrostatic fields) introduced by Kingdon [132] in 1923. Nonetheless, it is only in 2000, with the work conducted by Makarov [133, 134], that the ion trapping in the quadro-logarithmic field showed its capability of providing mass spectra. These findings resulted in the emergence of the first Orbitrap mass analyser. In 2005 the instrument became commercially available by Thermo Fisher.

Since then, a wide range of Orbitrap-based instruments have been produced with a continuous improvement of instrumental performances (such as resolution or speed of the analysis) and sold worldwide for various applications, including environmental and clinical analyses, metabolomics, lipidomics and proteomics. The instrument is defined as a high-resolution mass spectrometer since it provides a resolution of 15,000 – 500,000 (full width at half maximum FWHM) at m/z 200 and an accuracy lower than one ppm.

Principles of Orbitrap mass analyser

Three electrodes compose the orbitrap: a spindle-like inner electrode that traps the charged ions in an orbital movement around it, and two outer barrel-like electrodes, facing each other inwards and blocked by a central dielectric ring [135] (Figure I-3).

All three electrodes are designed to generate an electrostatic field, e.g., a quadro-logarithmic potential distribution [136] described by the Mathieu equation:

$$U\,(r,z) = \frac{k}{2}\left(z^2 - \frac{r^2}{2}\right) + \frac{k}{2}(R_m)^2 * \ln\left[\frac{r}{R_m}\right] + C$$

where r, z are cylindrical coordinates, R_m is the characteristic radius, C is a constant, k is the field curvature, defined by the shape of the central electrode and the applied voltage [133].

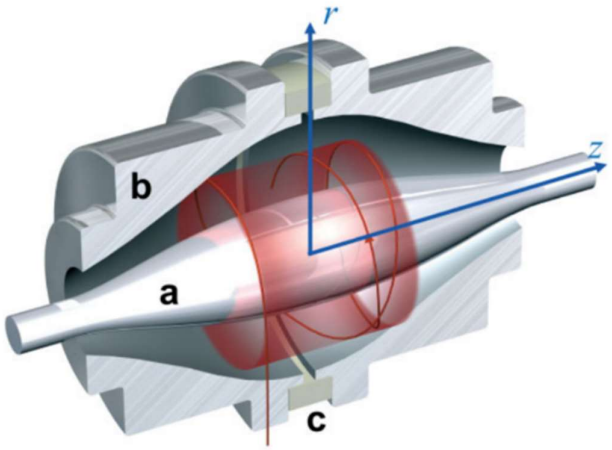


Figure I-3 A cut-away model of the Orbitrap: a) central electrode, b) outer electrode, c) simulation of the ions in red. Figure extracted from Scigelova and Makarov, 2006.

The trapped ions are induced to form stable trajectories turning around the central electrode and oscillating along its length. The motion of the trapped ion presents three frequencies:

- axial oscillations frequency ω, along the z-axis
- radial oscillation frequency ω_r between the maximum and minimum radii
- rotational frequency $\omega_{\varphi,}$ around the core electrode

Contrarily to the rotational and radial frequencies dependent on the radius R, axial frequencies are independent of the ions' initial speeds and coordinates. Hence, these frequencies depend exclusively on the m/z ratio and can be used for its determination [137, 138]:

$$\omega = \sqrt{\frac{q}{m}k}$$

The constant k comes from the Mathieu equation and varies with the voltage between the core and the external electrodes. The axial oscillation frequencies can be detected with the measurement of the image current on the outer orbitrap electrodes.

Ion capture

Before reaching the Orbitrap, the ions produced by the continuous ion source are first stored in a linear trap and then transferred to a C-shaped ion trap called C-trap. This latter allows the accumulation of the ions followed by a radial injection that minimizes the initial kinetic energy in the z-direction. The ions are trapped through the application of electrical potential at the electrostatic gate and a trap electrode located in the other extremity.

For the ions' extraction, the radiofrequency potential applied is turned off, pulsed voltages are then applied to the electrodes, which extract the ions by pushing them orthogonally to the curved axis through a slot placed in the internal hyperbolic electrode. The ions are channelled into the Orbitrap through the curvature of the C-Trap and subsequent lenses, whose purpose is to accelerate and compact them into small "packets" of the specific mass-to-charge ratio. Ion packets of individual m/z are injected into the Orbitrap at a certain distance from the equator (closer to one of the extremities of the trap) and begin a coherent axial oscillation.

Their capture is achieved through the principle of '*electrodynamic-squeezing'* [129], in which they undergo a constant increase in electric field strength. The potential of the outer electrons is left stable while the one on the central electrode is reduced (increasing the electric voltage). These conditions drive the ion packets to move toward the central electrode, allowing the entrance of new ions (with higher m/z). The ions' *squeezing* is stopped when the desired radius is reached (commonly equidistant from central and external electrodes). The voltage is then stabilised to avoid mass deviation in the detection.

Ion detection

Once all ions with different m/z ratios have entered the Orbitrap, the packet of ions with the same m/z will continue to oscillate along the radial direction together, remaining in phase.

In contrast, radial and rotational motion frequencies will start to vary slightly based on small differences in the initial parameters. This de-phasing provokes the redistribution of the ion packets in a ring shape with the ions uniformly distributed along its circumference. This ion ring, moving from one outer electrode to the other, will induce an image current on each of the outer electrodes, which in turn will be amplified differentially, producing a signal. This latter will be digitalised and converted from a time-domain signal to a mass-to-charge ratio spectrum through Fourier transform [139].

In this doctoral thesis, two different versions of Orbitrap instrumentations were used: the QExactivePlus with a resolution of more than 140.000 (m/z 200) and the Orbitrap Fusion Lumos Tribrid architecture (quadrupole mass filter, linear ion trap and Orbitrap mass analysers) with a resolution of 500.000 (m/z 200) (Figure I-4). The mass accuracy for both is estimated < 1 ppm.

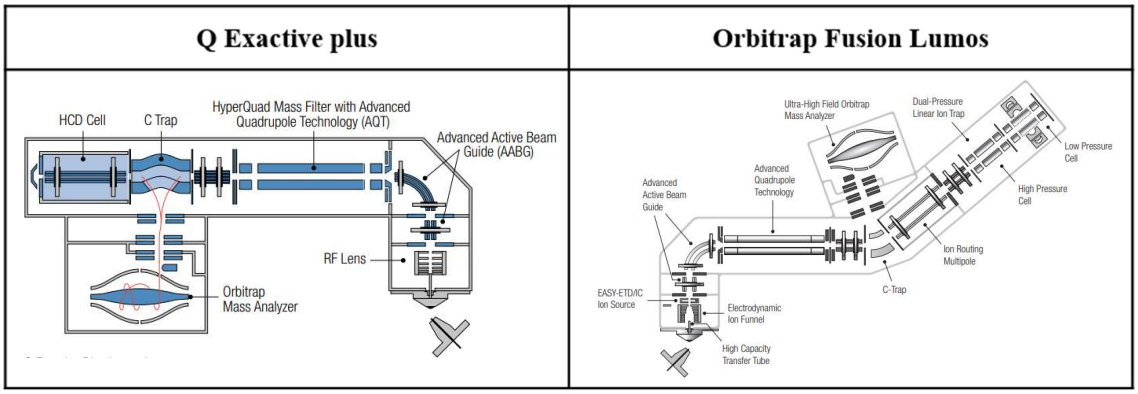


Figure I-4 Scheme of Q Exactive plus and of Orbitrap Fusion Lumos.

4. Fragmentation in tandem mass spectrometry (MS/MS)

Tandem mass spectrometry (MS/MS) is a two-step technique in which specific ions (precursor ions) are fragmented (product ions), providing the determination of the whole protein or peptide sequence and the structure of the corresponding precursor ion [140, 141].

In the MS/MS analysis, a first mass analyser separates the ions according to their m/z ratio; then the selected ions are fragmented in a collision cell (located between the two mass analysers), the fragmented ions are then sorted according to their m/z ratio in a second mass analyser and detected. The resulting MS2 spectrum comprises all the m/z values constituting the primary sequence of the precursor ion differing one from the other by one amino acid [142]. Ion fragmentation can be performed (i) in space when two or more mass analysers are sequentially combined to each other's (Q–q-Q, Q-Tof, QIT), or (ii) in time when one single mass analyser is involved for the selection and the following fragmentation (Ion Trap or FTICR) [143].

Different molecular dissociation techniques have been developed to induce the formation of fragment ions through the specific cleavage of the repeating amide bonds in one or more of the three potential positions: CαCO (a-,x-) and CO-NH (b-,y-) and NH-Cα (c-, z-) (Figure I-5). These fissions can be heterolytic or homolytic, and they can include the relocation of one or more hydrogens, holding the charge at the C- (x,y,z ions) or N-terminus (a,b,c ions).

The main fragmentation methods are:

1. Collisional-based fragmentations. Collision-induced dissociation (CID) [140] and higher-energy dissociation (HCD) [144, 145] are two techniques based on the ion collision with inert gas, such as nitrogen molecules or argon or helium atoms, with the consequent formation of b- and y- ions.

2. Electron fragmentation. These techniques consist in the reaction among electrons (<0.2 eV) and the precursor peptide ion with the consequent formation of a positive precursor ion ([M+nH] (n-1) +) and the following cleavage of N-Cα bonds, producing c and z type ions. In electron transfer dissociation (ETD) [146, 147], an electron is transferred by an anion to a protonated analyte. The electron capture dissociation ECD [148] instead consists of exothermic interactions of thermal electrons with the protonated analyte. Contrary to CID, which can cause the loss of certain labile PTMs (e.g., phosphorylations), these methodologies led to their preservation [146].

3. Photo-dissociation. Ultraviolet photo-dissociation 193 nm (UVPD) consists of the absorption of pulsed UV laser from the amide backbone, which causes a uniform cleavage all over the entire amino acid sequence and the formation of a great variety of ions types (x, y, z, a, b, c). This technique can potentially provide the highest sequence coverage due to a rich array of fragment ions leading to the most confident proteoform characterisation [149].

4. Infrared multiphoton dissociation (IRMPD). The technique is based on releasing an infrared laser to protein parent ions producing b- and y- ions. Additionally, to CID, a better m/z trapping range and identification of PTMs are provided [150, 151].

5. Different fragmentation techniques have been recently coupled into hybrid fragmentation methods to improve their singular features. For example, ETD has been combined with collisional methods CID or HCD (ETciD and EThcD) to achieve a higher sequence coverage over a broad range of charge states and a more detailed PTMs map [152, 153].

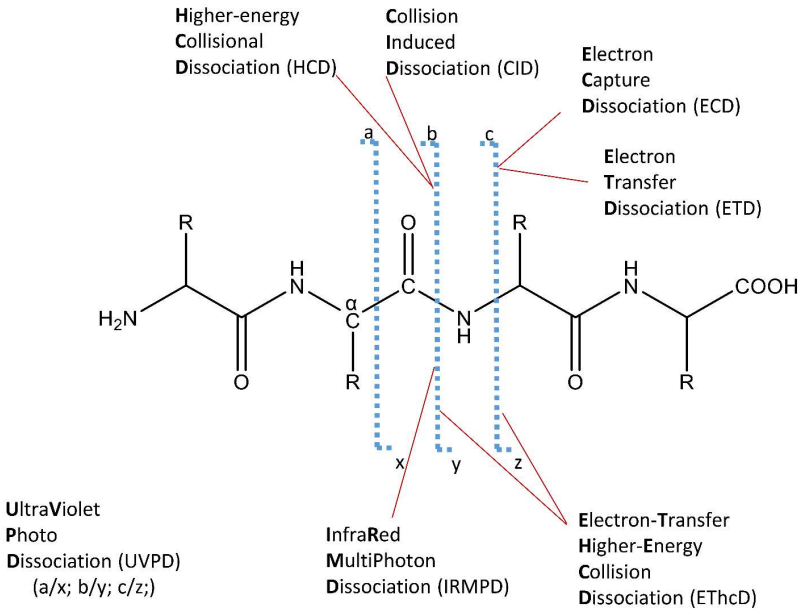


Figure I-5 Potential fragmentations in the peptide backbone with the corresponding nomenclature of the forming fragment ions and the MS/MS fragmentation. Figure elaborated from Zhurov et al., 2013.

In the presented doctoral research Higher-energy Collisional Dissociation (HCD) was applied for the studies via bottom-up approach (Chapter II), and Electron-Transfer/Higher-Energy Collision Dissociation (EthcD) was implemented in top-down investigations (Chapter III).

4.1 HCD: Higher-energy Collisional Dissociation

High-energy collisional dissociation (HCD) developed in 2007 as an optimisation of the CID available on Orbitrap technology [144]. The fragmentation is produced in a collision cell octopolaire designed to reach higher Radio Frequency (RF) voltage and, in turn, higher collisional energies than CID, which is conducted in a linear ion trap [154, 155]. In the Orbitrap analyser, the produced fragment ions are first trapped and cooled inside the multipole and then moved to the C-trap and into the Orbitrap to be detected. The precursor ion can experience multiple fragmentations in these conditions, producing a greater fragment ion signal (whether an over-fragmentation is prevented) [156].

Additionally, with HCD, the loss of low molecular weight fragments (lower than one-third of the precursor ion) is prevented [144]. Therefore, a high resolution and high mass accuracy can be provided with MS/MS spectra covering almost the whole mass range (including the low mass area with a2 and b2 ions, immonium ions) [157].

4.2 Electron-Transfer/Higher-Energy Collision Dissociation (EthcD)

Electron-Transfer/Higher-Energy Collision Dissociation is a hybrid fragmentation technique that combines electron transfer dissociation (ETD) with beam-type collision-induced dissociation (HCD) to take advantages of each of the two methods. The elevated phosphate neutral-losses constrain the collisional fragmentation while the electron fragmentation preserves labile modifications, but it principally detects high charge states, missing several doubly charged ions [158].

The hybrid technique can be performed in the recent orbitrap instruments, and it consists of first ETD fragmentation in the linear ion trap, followed by the transfer of all ions (both precursors and products) to the HCD collision cell, for the second process of fragmentation. The resulting fragmentation spectrum shows the information gathered from HCD and ETD together, encompassing b-, c-, y-, and z-ions [159, 160]. Higher protein coverage and a more secure localisation of post-translational modifications (PTMs) are achieved compared to the individual techniques [161]. These results prompted the significant exploitation of EthcD, mostly for top-down studies [152, 162, 163]. Research also showed that this technique is especially beneficial for *de novo* sequencing [164, 165]. The optimisation of database search engines is a current concern to provide a complete analysis of dual ion series and offer a higher rate of identification success in the study of an extensive data set [166].

Reference

1. Dallongeville, S., N. Garnier, C. Rolando ,C. Tokarski, *Proteins in art, archaeology, and paleontology: from detection to identification.* Chemical reviews, 2016. **116**(1): p. 2-79.
2. Vinciguerra, R., A. De Chiaro, P. Pucci, G. Marino ,L. Birolo, *Proteomic strategies for cultural heritage: From bones to paintings.* Microchemical Journal, 2016. **126**: p. 341-348.
3. Cappellini, E., A. Prohaska, F. Racimo, F. Welker, M.W. Pedersen, M.E. Allentoft, P. de Barros Damgaard, P. Gutenbrunner, J. Dunne ,S. Hammann, *Ancient biomolecules and evolutionary inference.* Annual review of biochemistry, 2018. **87**: p. 1029-1060.
4. Welker, F., *Palaeoproteomics for human evolution studies.* Quaternary Science Reviews, 2018. **190**: p. 137-147.
5. Welker, F., M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo ,A. Kramarz, *Ancient proteins resolve the evolutionary history of Darwin's South American ungulates.* Nature, 2015. **522**(7554): p. 81-84.
6. Welker, F., G.M. Smith, J.M. Hutson, L. Kindler, A. Garcia-Moreno, A. Villaluenga, E. Turner ,S. Gaudzinski-Windheuser, *Middle Pleistocene protein sequences from the rhinoceros genus Stephanorhinus and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae.* PeerJ, 2017. **5**: p. e3033.
7. Horn, I.R., Y. Kenens, N.M. Palmblad, S.J. Van der Plas-Duivesteijn, B.W. Langeveld, H.J. Meijer, H. Dalebout, R.J. Marissen, A. Fischer ,F. Vincent Florens, *Palaeoproteomics of bird bones for taxonomic classification.* Zoological Journal of the Linnean Society, 2019. **186**(3): p. 650-665.
8. Warinner, C., J. Hendy, C. Speller, E. Cappellini, R. Fischer, C. Trachsel, J. Arneborg, N. Lynnerup, O.E. Craig ,D.M. Swallow, *Direct evidence of milk consumption from ancient human dental calculus.* Scientific reports, 2014. **4**: p. 7104.
9. Hendy, J., C. Warinner, A. Bouwman, M.J. Collins, S. Fiddyment, R. Fischer, R. Hagan, C.A. Hofman, M. Holst ,E. Chaves, *Proteomic evidence of dietary sources in ancient dental calculus.* Proceedings of the Royal Society B: Biological Sciences, 2018. **285**(1883): p. 20180977.
10. Solazzo, C., W.W. Fitzhugh, C. Rolando ,C. Tokarski, *Identification of protein remains in archaeological potsherds by proteomics.* Analytical Chemistry, 2008. **80**(12): p. 4590-4597.
11. Gleba, M., *Sheep to textiles: approaches to investigating ancient wool trade*. 2014, Harrassowitz.
12. Sinet-Mathiot, V., G.M. Smith, M. Romandini, A. Wilcke, M. Peresani, J.-J. Hublin ,F. Welker, *Combining ZooMS and zooarchaeology to study Late Pleistocene hominin behaviour at Fumane (Italy).* Scientific reports, 2019. **9**(1): p. 1-13.
13. Cappellini, E., M.T.P. Gilbert, F. Geuna, G. Fiorentino, A. Hall, J. Thomas-Oates, P.D. Ashton, D.A. Ashford, P. Arthur ,P.F. Campos, *A multidisciplinary study of archaeological grape seeds.* Naturwissenschaften, 2010. **97**(2): p. 205-217.
14. Brandt, L.Ø., K. Haase ,M.J. Collins, *Species identification using ZooMS, with reference to the exploitation of animal resources in the medieval town of Odense.* Danish Journal of Archaeology, 2018. **7**(2): p. 139-153.
15. Shevchenko, A., Y. Yang, A. Knaust, H. Thomas, H. Jiang, E. Lu, C. Wang ,A. Shevchenko, *Proteomics identifies the composition and manufacturing recipe of the 2500-year old sourdough bread from Subeixi cemetery in China.* Journal of proteomics, 2014. **105**: p. 363-371.
16. Colombini, M.P. ,F. Modugno, *Organic mass spectrometry in art and archaeology*. 2009: John Wiley & Sons.

17. Cennini, C., C.D.A. Cennini ,G. De Beer, *The craftsman's handbook*. Vol. 2. 1954: Courier Corporation.
18. Colombini, M.P., R. Fuoco, A. Giacomelli ,B. Muscatello, *Characterization of proteinaceous binders in wall painting samples by microwave-assisted acid hydrolysis and GC-MS determination of amino acids.* Studies in conservation, 1998. **43**(1): p. 33-41.
19. Linn, R., I. Bonaduce, G. Ntasi, L. Birolo, A. Yasur-Landau, E.H. Cline, A. Nevin ,A. Lluveras-Tenorio, *Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings.* Angewandte Chemie, 2018. **130**(40): p. 13441-13444.
20. Fiddyment, S., M.D. Teasdale, J. Vnouček, É. Lévêque, A. Binois ,M.J. Collins, *So you want to do biocodicology? A field guide to the biological analysis of parchment.* Heritage Science, 2019. **7**(1): p. 35.
21. Dallongeville, S., M. Koperska, N. Garnier, G. Reille-Taillefert, C. Rolando ,C. Tokarski, *Identification of animal glue species in artworks using proteomics: application to a 18th century gilt sample.* Analytical chemistry, 2011. **83**(24): p. 9431-9437.
22. Pozzi, F., J. Arslanoglu, F. Galluzzi, C. Tokarski ,R. Snyder, *Mixing, dipping, and fixing: the experimental drawing techniques of Thomas Gainsborough.* Heritage Science, 2020. **8**(1): p. 1-14.
23. Tokarski, C., E. Martin, C. Rolando ,C. Cren-Olivé, *Identification of proteins in renaissance paintings by proteomics.* Analytical chemistry, 2006. **78**(5): p. 1494-1502.
24. Cicatiello, P., G. Ntasi, M. Rossi, G. Marino, P. Giardina ,L. Birolo, *Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings.* Analytical chemistry, 2018. **90**(17): p. 10128-10133.
25. Lluveras-Tenorio, A., R. Vinciguerra, E. Galano, C. Blaensdorf, E. Emmerling, M. Perla Colombini, L. Birolo ,I. Bonaduce, *GC/MS and proteomics to unravel the painting history of the lost Giant Buddhas of Bāmiyān (Afghanistan).* PloS one, 2017. **12**(4): p. e0172990.
26. Avrin, L., *Scribes, script, and books: the book arts from antiquity to the Renaissance*. 2010: American library association.
27. Leo, G., L. Cartechini, P. Pucci, A. Sgamellotti, G. Marino ,L. Birolo, *Proteomic strategies for the identification of proteinaceous binders in paintings.* Analytical and bioanalytical chemistry, 2009. **395**(7): p. 2269-2280.
28. Sotiropoulou, S., G. Sciutto, A.L. Tenorio, J. Mazurek, I. Bonaduce, S. Prati, R. Mazzeo, M. Schilling ,M.P. Colombini, *Advanced analytical investigation on degradation markers in wall paintings.* Microchemical Journal, 2018. **139**: p. 278-294.
29. Orsini, S., A. Yadav, M. Dilillo, L.A. McDonnell ,I. Bonaduce, *Characterization of degraded proteins in paintings using bottom-up proteomic approaches: new strategies for protein digestion and analysis of data.* Analytical chemistry, 2018. **90**(11): p. 6403-6408.
30. Bonaduce, I., M. Cito, M.P. Colombini ,A. Lluveras, *The characterisation of the organic binders.* Monuments and Sites, 2009. **19**: p. 265-280.
31. Arslanoglu, J., J. Schultz, J. Loike ,K. Peterson, *Immunology and art: using antibody-based techniques to identify proteins and gums in artworks.* Journal of biosciences, 2010. **35**(1): p. 3-10.
32. Colombini, M.P. ,F. Modugno, *Characterisation of proteinaceous binders in artistic paintings by chromatographic techniques.* Journal of Separation Science, 2004. **27**(3): p. 147-160.

33. Calvano, C.D., I.D. van der Werf, F. Palmisano ,L. Sabbatini, *Revealing the composition of organic materials in polychrome works of art: the role of mass spectrometry-based techniques.* Analytical and bioanalytical chemistry, 2016. **408**(25): p. 6957-6981.
34. Vinciguerra, R., A. Illiano, A. De Chiaro, A. Carpentieri, A. Lluveras-Tenorio, I. Bonaduce, G. Marino, P. Pucci, A. Amoresano ,L. Birolo, *Identification of proteinaceous binders in paintings: A targeted proteomic approach for cultural heritage.* Microchemical Journal, 2019. **144**: p. 319-328.
35. Leo, G., I. Bonaduce, A. Andreotti, G. Marino, P. Pucci, M.P. Colombini ,L. Birolo, *Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings.* Analytical chemistry, 2011. **83**(6): p. 2056-2064.
36. Mackie, M., P. Rüther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Peggie, H. Howard, L. Harrison ,L.J. Jensen, *Palaeoproteomic profiling of conservation layers on a 14th century Italian wall painting.* Angewandte Chemie International Edition, 2018. **57**(25): p. 7369-7374.
37. Marvasi, M., D. Cavalieri, G. Mastromei, A. Casaccia ,B. Perito, *Omics technologies for an in-depth investigation of biodeterioration of cultural heritage.* International Biodeterioration & Biodegradation, 2019. **144**: p. 104736.
38. Wilson, J., N.L. van Doorn ,M.J. Collins, *Assessing the extent of bone degradation using glutamine deamidation in collagen.* Analytical chemistry, 2012. **84**(21): p. 9041-9048.
39. Beata, G., *The use of-omics tools for assessing biodeterioration of cultural heritage: A review.* Journal of Cultural Heritage, 2020.
40. Mills, J. ,R. White, *Organic chemistry of museum objects*. 2012: Routledge.
41. van den Brink, O.F., *Molecular changes in egg tempera paint dosimeters as tools to monitor the museum environment*. 2001: FOM-Institute for Atomic and Molecular Physics.
42. Pattison, D.I., A.S. Rahmanto ,M.J. Davies, *Photo-oxidation of proteins.* Photochemical & Photobiological Sciences, 2012. **11**(1): p. 38-53.
43. Zhang, X., I.V. Berghe ,P. Wyeth, *Heat and moisture promoted deterioration of raw silk estimated by amino acid analysis.* Journal of Cultural Heritage, 2011. **12**(4): p. 408-411.
44. Nevin, A., D. Anglos, S. Cather ,A. Burnstock, *The influence of visible light and inorganic pigments on fluorescence excitation emission spectra of egg-, casein-and collagen-based painting media.* Applied Physics A, 2008. **92**(1): p. 69-76.
45. Casadio, F., C. Daher ,L. Bellot-Gurlet, *Raman spectroscopy of cultural heritage materials: overview of applications and new frontiers in instrumentation, sampling modalities, and data processing*, in *Analytical Chemistry for Cultural Heritage*. 2017, Springer. p. 161-211.
46. Bersani, D., C. Conti, P. Matousek, F. Pozzi ,P. Vandenabeele, *Methodological evolutions of Raman spectroscopy in art and archaeology.* Analytical Methods, 2016. **8**(48): p. 8395-8409.
47. Nevin, A., I. Osticioli, D. Anglos, A. Burnstock, S. Cather ,E. Castellucci, *Raman spectra of proteinaceous materials used in paintings: a multivariate analytical approach for classification and identification.* Analytical chemistry, 2007. **79**(16): p. 6143-6151.
48. Duce, C., L. Ghezzi, M. Onor, I. Bonaduce, M.P. Colombini, M.R. Tine ,E. Bramanti, *Physico-chemical characterization of protein–pigment interactions in tempera paint reconstructions: casein/cinnabar and albumin/cinnabar.* Analytical and bioanalytical chemistry, 2012. **402**(6): p. 2183-2193.

49. Pellegrini, D., C. Duce, I. Bonaduce, S. Biagi, L. Ghezzi, M.P. Colombini, M.R. Tinè ,E. Bramanti, *Fourier transform infrared spectroscopic study of rabbit glue/inorganic pigments mixtures in fresh and aged reference paint reconstructions.* Microchemical Journal, 2016. **124**: p. 31-35.
50. Duce, C., E. Bramanti, L. Ghezzi, L. Bernazzani, I. Bonaduce, M.P. Colombini, A. Spepi, S. Biagi ,M.R. Tine, *Interactions between inorganic pigments and proteinaceous binders in reference paint reconstructions.* Dalton Transactions, 2013. **42**(17): p. 5975-5984.
51. Andersen, C.K., I. Bonaduce, A. Andreotti, J. van Lanschot ,A. Vila, *Characterisation of preparation layers in nine Danish Golden Age canvas paintings by SEM–EDX, FTIR and GC–MS.* Heritage Science, 2017. **5**(1): p. 34.
52. Ghezzi, L., C. Duce, L. Bernazzani, E. Bramanti, M.P. Colombini, M.R. Tiné ,I. Bonaduce, *Interactions between inorganic pigments and rabbit skin glue in reference paint reconstructions.* Journal of Thermal Analysis and Calorimetry, 2015. **122**(1): p. 315-322.
53. Serafima, S., O. Duliu, M.M. Manea ,G. Niculescu, *FTIR, XRF and optical microscopy analysis of the painting layer of an early 19th century icon.* Romanian reports in Physics, 2016. **68**(1): p. 191-202.
54. Rosi, F., L. Cartechini, D. Sali ,C. Miliani, *Recent trends in the application of Fourier Transform Infrared (FT-IR) spectroscopy in Heritage Science: From micro-to non-invasive FT-IR.* Physical Sciences Reviews, 2019. **4**(11).
55. Pięta, E., J. Olszewska-Świetlik, C. Paluszkiewicz, A. Zając ,W.M. Kwiatek, *Application of ATR-FTIR mapping to identification and distribution of pigments, binders and degradation products in a 17th century painting.* Vibrational Spectroscopy, 2019. **103**: p. 102928.
56. Prati, S., E. Joseph, G. Sciutto ,R. Mazzeo, *New advances in the application of FTIR microscopy and spectroscopy for the characterization of artistic materials.* Accounts of Chemical Research, 2010. **43**(6): p. 792-801.
57. Perets, E., A. Indrasekara, A. Kurmis, N. Atlasevich, L. Fabris ,J. Arslanoglu, *Carboxy-terminated immuno-SERS tags overcome non-specific aggregation for the robust detection and localization of organic media in artworks.* Analyst, 2015. **140**(17): p. 5971-5980.
58. Arslanoglu, J., S. Zaleski ,J. Loike, *An improved method of protein localization in artworks through SERS nanotag-complexed antibodies.* Analytical and bioanalytical chemistry, 2011. **399**(9): p. 2997-3010.
59. Zalaffi, M.S., N. Karimian ,P. Ugo, *Electrochemical and SERS Sensors for Cultural Heritage Diagnostics and Conservation: Recent Advances and Prospects.* Journal of The Electrochemical Society, 2020. **167**(3): p. 037548.
60. Nodari, L. ,P. Ricciardi, *Non-invasive identification of paint binders in illuminated manuscripts by ER-FTIR spectroscopy: a systematic study of the influence of different pigments on the binders' characteristic spectral features.* Heritage Science, 2019. **7**(1): p. 7.
61. Meyers, R.A., *Encyclopedia of analytical chemistry*. 2006: Univerza v Novi Gorici.
62. Deyl, Z., I. Mikšık ,J. Zicha, *Multicomponent analysis by off-line combination of synchronous fluorescence spectroscopy and capillary electrophoresis of collagen glycation adducts.* Journal of Chromatography A, 1999. **836**(1): p. 161-171.
63. Nevin, A., S. Cather, D. Anglos ,C. Fotakis, *Analysis of protein-based binding media found in paintings using laser induced fluorescence spectroscopy.* Analytica chimica acta, 2006. **573**: p. 341-346.
64. Nevin, A., D. Comelli, G. Valentini, D. Anglos, A. Burnstock, S. Cather ,R. Cubeddu, *Time-resolved fluorescence spectroscopy and imaging of proteinaceous binders used in paintings.* Analytical and Bioanalytical Chemistry, 2007. **388**(8): p. 1897-1905.

65. Osticioli, I., N. Mendes, A. Nevin, A. Zoppi, C. Lofrumento, M. Becucci ,E. Castellucci, *A new compact instrument for Raman, laser-induced breakdown, and laser-induced fluorescence spectroscopy of works of art and their constituent materials.* Review of Scientific Instruments, 2009. **80**(7): p. 076109.
66. Lluveras, A., I. Bonaduce, A. Andreotti ,M.P. Colombini, *GC/MS analytical procedure for the characterization of glycerolipids, natural waxes, terpenoid resins, proteinaceous and polysaccharide materials in the same paint microsample avoiding interferences from inorganic media.* Analytical chemistry, 2010. **82**(1): p. 376-386.
67. Fico, D., E. Margapoti, A. Pennetta ,G.E. De Benedetto, *An enhanced GC/MS procedure for the identification of proteins in paint microsamples.* Journal of Analytical Methods in Chemistry, 2018. **2018**.
68. Degano, I., F. Modugno, I. Bonaduce, E. Ribechini ,M.P. Colombini, *Recent advances in analytical pyrolysis to investigate organic materials in heritage science.* Angewandte Chemie International Edition, 2018. **57**(25): p. 7313-7323.
69. Orsini, S., F. Parlanti ,I. Bonaduce, *Analytical pyrolysis of proteins in samples from artistic and archaeological objects.* Journal of Analytical and Applied Pyrolysis, 2017. **124**: p. 643-657.
70. No, A. ,A.M. Committee, *Analytical pyrolysis in cultural heritage.* Analytical Methods, 2018. **10**(46): p. 5463-5467.
71. Chiavari, G., D. Fabbri, G. Galletti ,R. Mazzeo, *Use of analytical pyrolysis to characterize Egyptian painting layers.* Chromatographia, 1995. **40**(9-10): p. 594-600.
72. Marinach, C., M.-C. Papillon ,C. Pepe, *Identification of binding media in works of art by gas chromatography–mass spectrometry.* Journal of Cultural Heritage, 2004. **5**(2): p. 231-240.
73. White, R., *The characterization of proteinaceous binders in art objects.* National Gallery Technical Bulletin, 1984. **8**: p. 5-14.
74. De la Cruz-Canizares, J., M. Doménech-Carbó, J. Gimeno-Adelantado, R. Mateo-Castro ,F. Bosch-Reig, *Suppression of pigment interference in the gas chromatographic analysis of proteinaceous binding media in paintings with EDTA.* Journal of Chromatography A, 2004. **1025**(2): p. 277-285.
75. Chiavari, G., G. Lanterna, C. Luca, M. Matteini, S. Prati ,I. Sandu, *Analysis of proteinaceous binders by in-situ pyrolysis and silylation.* Chromatographia, 2003. **57**(9-10): p. 645-648.
76. Vallance, S.L., *Critical review: Applications of chromatography in art conservation: Techniques used for the analysis and identification of proteinaceous and gum binding media.* Analyst, 1997. **122**(6): p. 75R-81R.
77. Calvano, C.D., I.D. van der Werf, L. Sabbatini ,F. Palmisano, *On plate graphite supported sample processing for simultaneous lipid and protein identification by matrix assisted laser desorption ionization mass spectrometry.* Talanta, 2015. **137**: p. 161-166.
78. Materazzi, S. ,S. Vecchio, *Evolved gas analysis by mass spectrometry.* Applied Spectroscopy Reviews, 2011. **46**(4): p. 261-340.
79. Mazel, V. ,P. Richardin, *ToF-SIMS study of organic materials in cultural heritage: Identification and chemical imaging*. 2009: Wiley.
80. Noun, M., E. Van Elslande, D. Touboul, H. Glanville, S. Bucklow, P. Walter ,A. Brunelle, *High mass and spatial resolution mass spectrometry imaging of Nicolas Poussin painting cross section by cluster TOF-SIMS.* Journal of mass spectrometry, 2016. **51**(12): p. 1196-1210.
81. Atrei, A., F. Benetti, S. Bracci, D. Magrini ,N. Marchettini, *An integrated approach to the study of a reworked painting "Madonna with child" attributed to Pietro Lorenzetti.* Journal of cultural heritage, 2014. **15**(1): p. 80-84.

82. Lee, H.Y., N. Atlasevich, C. Granzotto, J. Schultz, J. Loike ,J. Arslanoglu, *Development and application of an ELISA method for the analysis of protein-based binding media of artworks.* Analytical Methods, 2015. **7**(1): p. 187-196.
83. Li, J. ,B. Zhang, *Study of identification results of proteinous binding agents in Chinese painted cultural relics.* Journal of Cultural Heritage, 2020.
84. Sciutto, G., M. Zangheri, S. Prati, M. Guardigli, M. Mirasoli, R. Mazzeo ,A. Roda, *Immunochemical Micro Imaging Analyses for the Detection of Proteins in Artworks.* Topics in Current Chemistry, 2016. **374**(3): p. 32.
85. Cartechini, L., M. Palmieri, M. Vagnini ,L. Pitzurra, *Immunochemical methods applied to art-Historical materials: identification and localization of proteins by ELISA and IFM.* Topics in Current Chemistry, 2016. **374**(1): p. 5.
86. Bottari, F., P. Oliveri ,P. Ugo, *Electrochemical immunosensor based on ensemble of nanoelectrodes for immunoglobulin IgY detection: Application to identify hen's egg yolk in tempera paintings.* Biosensors and Bioelectronics, 2014. **52**: p. 403-410.
87. Gambino, M., F. Cappitelli, C. Cattò, A. Carpen, P. Principi, L. Ghezzi, I. Bonaduce, E. Galano, P. Pucci ,L. Birolo, *A simple and reliable methodology to detect egg white in art samples.* Journal of biosciences, 2013. **38**(2): p. 397-408.
88. Chen, R., M. Hu, H. Zheng, H. Yang, L. Zhou, Y. Zhou, Z. Peng, Z. Hu ,B. Wang, *Proteomics and Immunology Provide Insight into the Degradation Mechanism of Historic and Artificially Aged Silk.* Analytical Chemistry, 2020. **92**(3): p. 2435-2442.
89. Sciutto, G., M. Zangheri, L. Anfossi, M. Guardigli, S. Prati, M. Mirasoli, F. Di Nardo, C. Baggiani, R. Mazzeo ,A. Roda, *Miniaturized Biosensors to Preserve and Monitor Cultural Heritage: from Medical to Conservation Diagnosis.* Angewandte Chemie, 2018. **130**(25): p. 7507-7511.
90. Tokarski, C., E. Martin, C. Rolando ,C. Cren-Olivé, *Protein identification in ancient art paintings.* Biosystems Solutions, 2004. **11**: p. 38-39.
91. Tokarski, C., C. Cren-Olive, C. Rolando ,E. Martin, *Protein studies in cultural heritage.* Molecular biology and cultural heritage: Swets & Zeitlinger, 2003: p. 119-130.
92. Tokarski, C., C. Cren-Olive, E. Martin ,C. Rolando, *Proteomics approach for binding media studies in art painting.* Non-Destructive Testing and Microanalysis for the Diagnostics and Conservation of the Cultural and Environmental Heritage, Antwerp, Belgium, 2002.
93. Hynek, R., S. Kuckova, J. Hradilova ,M. Kodicek, *Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry as a tool for fast identification of protein binders in color layers of paintings.* Rapid communications in mass spectrometry, 2004. **18**(17): p. 1896-1900.
94. Fiddyment, S., B. Holsinger, C. Ruzzier, A. Devine, A. Binois, U. Albarella, R. Fischer, E. Nichols, A. Curtis ,E. Cheese, *Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting.* Proceedings of the National Academy of Sciences, 2015. **112**(49): p. 15066-15071.
95. Ramsøe, A., V. van Heekeren, P. Ponce, R. Fischer, I. Barnes, C. Speller ,M.J. Collins, *DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes.* Journal of Archaeological Science, 2020. **115**: p. 105080.
96. Kahn, P., *From genome to proteome: looking at a cell's proteins*. 1995, American Association for the Advancement of Science.
97. James, P., *Protein identification in the post-genome era: the rapid rise of proteomics.* Quarterly reviews of biophysics, 1997. **30**(4): p. 279-331.
98. Anderson, N.L. ,N.G. Anderson, *Proteome and proteomics: new technologies, new concepts, and new words.* Electrophoresis, 1998. **19**(11): p. 1853-1861.
99. Blackstock, W.P. ,M.P. Weir, *Proteomics: quantitative and physical mapping of cellular proteins.* Trends in biotechnology, 1999. **17**(3): p. 121-127.

100. Aebersold, R. ,M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
101. Han, X., A. Aslanian ,J.R. Yates III, *Mass spectrometry for proteomics.* Current opinion in chemical biology, 2008. **12**(5): p. 483-490.
102. O'Farrell, P.H., *High resolution two-dimensional electrophoresis of proteins.* Journal of biological chemistry, 1975. **250**(10): p. 4007-4021.
103. Shen, Y., S.J. Berger, G.A. Anderson ,R.D. Smith, *High-efficiency capillary isoelectric focusing of peptides.* Analytical chemistry, 2000. **72**(9): p. 2154-2159.
104. Rabilloud, T., *Solubilization of proteins for electrophoretic analyses.* Electrophoresis, 1996. **17**(5): p. 813-829.
105. Fournier, M.L., J.M. Gilmore, S.A. Martin-Brown ,M.P. Washburn, *Multidimensional separations-based shotgun proteomics.* Chemical reviews, 2007. **107**(8): p. 3654-3686.
106. Yates, J.R., C.I. Ruse ,A. Nakorchevsky, *Proteomics by mass spectrometry: approaches, advances, and applications.* Annual review of biomedical engineering, 2009. **11**: p. 49-79.
107. Liu, T., M.E. Belov, N. Jaitly, W.-J. Qian ,R.D. Smith, *Accurate mass measurements in proteomics.* Chemical reviews, 2007. **107**(8): p. 3621-3653.
108. Mann, M. ,N.L. Kelleher, *Precision proteomics: the case for high resolution and high mass accuracy.* Proceedings of the National Academy of Sciences, 2008. **105**(47): p. 18132-18138.
109. Gillet, L.C., A. Leitner ,R. Aebersold, *Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing.* Annual review of analytical chemistry, 2016. **9**: p. 449-472.
110. Zhang, Y., B.R. Fonslow, B. Shan, M.-C. Baek ,J.R. Yates III, *Protein analysis by shotgun/bottom-up proteomics.* Chemical reviews, 2013. **113**(4): p. 2343-2394.
111. Catherman, A.D., O.S. Skinner ,N.L. Kelleher, *Top down proteomics: facts and perspectives.* Biochemical and biophysical research communications, 2014. **445**(4): p. 683-693.
112. Kelleher, N.L., *Peer reviewed: Top-down proteomics*. 2004, ACS Publications.
113. Fenn, J.B., M. Mann, C.K. Meng, S.F. Wong ,C.M. Whitehouse, *Electrospray ionization–principles and practice.* Mass Spectrometry Reviews, 1990. **9**(1): p. 37-70.
114. Fenn, J.B., M. Mann, C.K. Meng, S.F. Wong ,C.M. Whitehouse, *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
115. Dole, M., L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson ,M.B. Alice, *Molecular beams of macroions.* The Journal of chemical physics, 1968. **49**(5): p. 2240-2249.
116. Iribarne, J. ,B. Thomson, *On the evaporation of small ions from charged droplets.* The Journal of chemical physics, 1976. **64**(6): p. 2287-2294.
117. Mora, J.F.d.l., G.J. Van Berkel, C.G. Enke, R.B. Cole, M. Martinez-Sanchez ,J.B. Fenn, *Electrochemical processes in electrospray ionization mass spectrometry.* Journal of Mass Spectrometry, 2000. **35**(8): p. 939-952.
118. Hoffman, E.d. ,V. Stroobant, *Mass spectrometry: principles and applications.* West Sussex: John Wiley & Sons, Bruxellas, Bélgica, 2007. **1**(2): p. 85.
119. Wilm, M. ,M. Mann, *Analytical properties of the nanoelectrospray ion source.* Analytical chemistry, 1996. **68**(1): p. 1-8.
120. Konermann, L., E. Ahadi, A.D. Rodriguez ,S. Vahidi, *Unraveling the mechanism of electrospray ionization*. 2013, ACS Publications.
121. Gibson, G.T., S.M. Mugo ,R.D. Oleschuk, *Nanoelectrospray emitters: trends and perspective.* Mass Spectrometry Reviews, 2009. **28**(6): p. 918-936.
122. Cooks, R.G., G. Glish, S.A. Mc Luckey ,R.E. Kaiser, *Ion trap mass spectrometry.* Chemical and Engineering News;(United States), 1991. **69**(12).

123. Hager, J.W., *A new linear ion trap mass spectrometer.* Rapid Communications in Mass Spectrometry, 2002. **16**(6): p. 512-526.
124. Wiley, W. ,I.H. McLaren, *Time-of-flight mass spectrometer with improved resolution.* Review of scientific instruments, 1955. **26**(12): p. 1150-1157.
125. March, R.E., *An introduction to quadrupole ion trap mass spectrometry.* Journal of mass spectrometry, 1997. **32**(4): p. 351-369.
126. Dawson, P.H., *Quadrupole mass spectrometry and its applications*. 2013: Elsevier.
127. Bogdanov, B. ,R.D. Smith, *Proteomics by FTICR mass spectrometry: top down and bottom up.* Mass spectrometry reviews, 2005. **24**(2): p. 168-200.
128. Hu, Q., R.J. Noll, H. Li, A. Makarov, M. Hardman ,R. Graham Cooks, *The Orbitrap: a new mass spectrometer.* Journal of mass spectrometry, 2005. **40**(4): p. 430-443.
129. Perry, R.H., R.G. Cooks ,R.J. Noll, *Orbitrap mass spectrometry: instrumentation, ion motion and applications.* Mass spectrometry reviews, 2008. **27**(6): p. 661-699.
130. Makarov, A., E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat ,S. Horning, *Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer.* Analytical chemistry, 2006. **78**(7): p. 2113-2120.
131. Glish, G.L. ,D.J. Burinsky, *Hybrid mass spectrometers for tandem mass spectrometry.* Journal of the American Society for Mass Spectrometry, 2008. **19**(2): p. 161-172.
132. Kingdon, K., *A method for the neutralization of electron space charge by positive ionization at very low gas pressures.* Physical Review, 1923. **21**(4): p. 408.
133. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* Analytical chemistry, 2000. **72**(6): p. 1156-1162.
134. Makarov, A.A., *Mass spectrometer arrangement with fragmentation cell and ion selection device*. 2010, Google Patents.
135. Scigelova, M. ,A. Makarov, *Orbitrap mass analyzer–overview and applications in proteomics.* Proteomics, 2006. **6**(S2): p. 16-21.
136. Knight, R., *Storage of ions from laser-produced plasmas.* Applied Physics Letters, 1981. **38**(4): p. 221-223.
137. Kharchenko, A., G. Vladimirov, R.M. Heeren ,E.N. Nikolaev, *Performance of Orbitrap mass analyzer at various space charge and non-ideal field conditions: simulation approach.* Journal of the American Society for Mass Spectrometry, 2012. **23**(5): p. 977-987.
138. Scigelova, M. ,A. Makarov, *Fundamentals and advances of Orbitrap mass spectrometry.* Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation, 2006: p. 1-36.
139. Perry, R.H., Q. Hu, G.A. Salazar, R.G. Cooks ,R.J. Noll, *Rephasing ion packets in the orbitrap mass analyzer to improve resolution and peak shape.* Journal of the American Society for Mass Spectrometry, 2009. **20**(8): p. 1397-1404.
140. Hunt, D.F., A.M. Buko, J.M. Ballard, J. Shabanowitz ,A.B. Giordani, *Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer.* Biomedical mass spectrometry, 1981. **8**(9): p. 397-408.
141. Hunt, D.F., J.R. Yates, J. Shabanowitz, S. Winston ,C.R. Hauer, *Protein sequencing by tandem mass spectrometry.* Proceedings of the National Academy of Sciences, 1986. **83**(17): p. 6233-6237.
142. Biemann, K. ,S.A. Martin, *Mass spectrometric determination of the amino acid sequence of peptides and proteins.* Mass Spectrometry Reviews, 1987. **6**(1): p. 1-75.
143. Sleno, L. ,D.A. Volmer, *Ion activation methods for tandem mass spectrometry.* Journal of mass spectrometry, 2004. **39**(10): p. 1091-1112.
144. Olsen, J.V., B. Macek, O. Lange, A. Makarov, S. Horning ,M. Mann, *Higher-energy C-trap dissociation for peptide modification analysis.* Nature methods, 2007. **4**(9): p. 709-712.

145. McAlister, G.C., D.H. Phanstiel, J. Brumbaugh, M.S. Westphall ,J.J. Coon, *Higher-energy collision-activated dissociation without a dedicated collision cell.* Molecular & Cellular Proteomics, 2011. **10**(5).
146. Zhurov, K.O., L. Fornelli, M.D. Wodrich, Ü.A. Laskay ,Y.O. Tsybin, *Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis.* Chemical Society Reviews, 2013. **42**(12): p. 5014-5030.
147. Li, X., C. Lin, L. Han, C.E. Costello ,P.B. O'connor, *Charge remote fragmentation in electron capture and electron transfer dissociation.* Journal of the American Society for Mass Spectrometry, 2011. **21**(4): p. 646-656.
148. Zubarev, R.A., N.L. Kelleher ,F.W. McLafferty, *Electron capture dissociation of multiply charged protein cations. A nonergodic process.* Journal of the American Chemical Society, 1998. **120**(13): p. 3265-3266.
149. Brodbelt, J.S., *Photodissociation mass spectrometry: new tools for characterization of biological molecules.* Chemical Society Reviews, 2014. **43**(8): p. 2757-2783.
150. Little, D.P., J.P. Speir, M.W. Senko, P.B. O'Connor ,F.W. McLafferty, *Infrared multiphoton dissociation of large multiply charged ions for biomolecule sequencing.* Analytical Chemistry, 1994. **66**(18): p. 2809-2815.
151. Wu, S., H. Ma, I. Prytkova, D. Stenoien ,L. Paša-Tolić, *Proteomics, Top-Down.* 2017.
152. Brunner, A.M., P. Lössl, F. Liu, R. Huguet, C. Mullen, M. Yamashita, V. Zabrouskov, A. Makarov, A.M. Altelaar ,A.J. Heck, *Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization.* Analytical chemistry, 2015. **87**(8): p. 4152-4158.
153. Riley, N.M., J.W. Sikora, H.S. Seckler, J.B. Greer, R.T. Fellers, R.D. LeDuc, M.S. Westphall, P.M. Thomas, N.L. Kelleher ,J.J. Coon, *The value of activated ion electron transfer dissociation for high-throughput top-down characterization of intact proteins.* Analytical chemistry, 2018. **90**(14): p. 8553-8560.
154. Michalski, A., E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Müller, R. Viner, J. Schwartz, P. Remes ,M. Belford, *Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes.* Molecular & Cellular Proteomics, 2012. **11**(3).
155. Kelstrup, C.D., C. Young, R. Lavallee, M.L. Nielsen ,J.V. Olsen, *Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer.* Journal of proteome research, 2012. **11**(6): p. 3487-3497.
156. Diedrich, J.K., A.F. Pinto ,J.R. Yates III, *Energy dependence of HCD on peptide fragmentation: stepped collisional energy finds the sweet spot.* Journal of the American Society for Mass Spectrometry, 2013. **24**(11): p. 1690-1699.
157. Rasche, F., A. Svatos, R.K. Maddula, C. Böttcher ,S. Böcker, *Computing fragmentation trees from tandem mass spectrometry data.* Analytical Chemistry, 2011. **83**(4): p. 1243-1251.
158. Penkert, M., A. Hauser, R. Harmel, D. Fiedler, C.P. Hackenberger ,E. Krause, *Electron Transfer/Higher Energy Collisional Dissociation of Doubly Charged Peptide Ions: Identification of Labile Protein Phosphorylations.* Journal of The American Society for Mass Spectrometry, 2019. **30**(9): p. 1578-1585.
159. Frese, C.K., A.M. Altelaar, H. van den Toorn, D. Nolting, J. Griep-Raming, A.J. Heck ,S. Mohammed, *Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry.* Analytical chemistry, 2012. **84**(22): p. 9668-9673.
160. Nielsen, M.L., M.M. Savitski ,R.A. Zubarev, *Improving protein identification using complementary fragmentation techniques in Fourier transform mass spectrometry.* Molecular & Cellular Proteomics, 2005. **4**(6): p. 835-845.

161. Frese, C.K., H. Zhou, T. Taus, A.M. Altelaar, K. Mechtler, A.J. Heck ,S. Mohammed, *Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD).* Journal of proteome research, 2013. **12**(3): p. 1520-1525.
162. Gomes, F.P., J.K. Diedrich, A.J. Saviola, E. Memili, A.A. Moura ,J.R. Yates III, *EThcD and 213 nm UVPD for Top-Down Analysis of Bovine Seminal Plasma Proteoforms on Electrophoretic and Chromatographic Time Frames.* Analytical Chemistry, 2020. **92**(4): p. 2979-2987.
163. Ahlf, D.R., P.D. Compton, J.C. Tran, B.P. Early, P.M. Thomas ,N.L. Kelleher, *Evaluation of the compact high-field orbitrap for top-down proteomics of human cells.* Journal of proteome research, 2012. **11**(8): p. 4308-4314.
164. Riley, N.M. ,J.J. Coon, *The role of electron transfer dissociation in modern proteomics.* Analytical chemistry, 2018. **90**(1): p. 40-64.
165. Naryzhny, S., *Inventory of proteoforms as a current challenge of proteomics: Some technical aspects.* Journal of proteomics, 2019. **191**: p. 22-28.
166. Kim, M.S., J. Zhong, K. Kandasamy, B. Delanghe ,A. Pandey, *Systematic evaluation of alternating CID and ETD fragmentation for phosphorylated peptides.* Proteomics, 2011. **11**(12): p. 2568-2572.

Chapter II Bottom-up proteomic analysis of historic and artistic samples

Abstract

In biological studies and cultural heritage investigations, bottom-up proteomics (BUP) is the mainstream strategy for protein identification and characterisation, including investigating their chemical and post-translational modifications. Since pioneering results in the 2000s, this strategy based on protein digestion has been greatly improved, leading to a more in-depth insight into artistic and historic samples. Nonetheless, research is always pursuing further improvements to decrease the sampling impact on the artwork, improving the analysis sensitivity and limiting the external interferences.

In this chapter, an optimised bottom-up methodological analysis has been discussed in application to the investigation of two different proteinaceous-based artworks: (i) wall paintings from the Nubian region (Middle Nile Valley) dated between the 6th and 14th century AD and (ii) drawings made by the English artist Thomas Gainsborough during the 18th century.

The shared aim was the achievement of clear identification of proteinaceous content despite the significant heterogeneity and the degradation conditions of the artworks and the low amount of the collected analyte. In the study of Gainsborough drawings, the samples' concentration under analysis was significantly reduced due to the implementation of two innovative minimally-invasive techniques consisting of a gentle rubbing of the surface through PVC-free erasers and fine polishing films. Furthermore, a careful examination was conducted on the protein PTMs to understand the state of conservation of the objects.

1. Bottom-up approach

In biological studies and cultural heritage investigations, the bottom-up approach (BUP) is the most common proteomic strategy for identifying proteins and the analysis of their post-translational modifications [1]. The technique, also called "*shotgun proteomics*", when performed on a proteinaceous mixture [2, 3], is based on the enzymatic or chemical digestion of these macromolecules into peptides before the mass spectrometry analysis. The wide implementation of this approach principally relies on easier separation, ionisation and fragmentation of peptides compared to intact proteins [4]. Following the protein digestion, the MS analysis can be performed using two different approaches (Figure II-1).

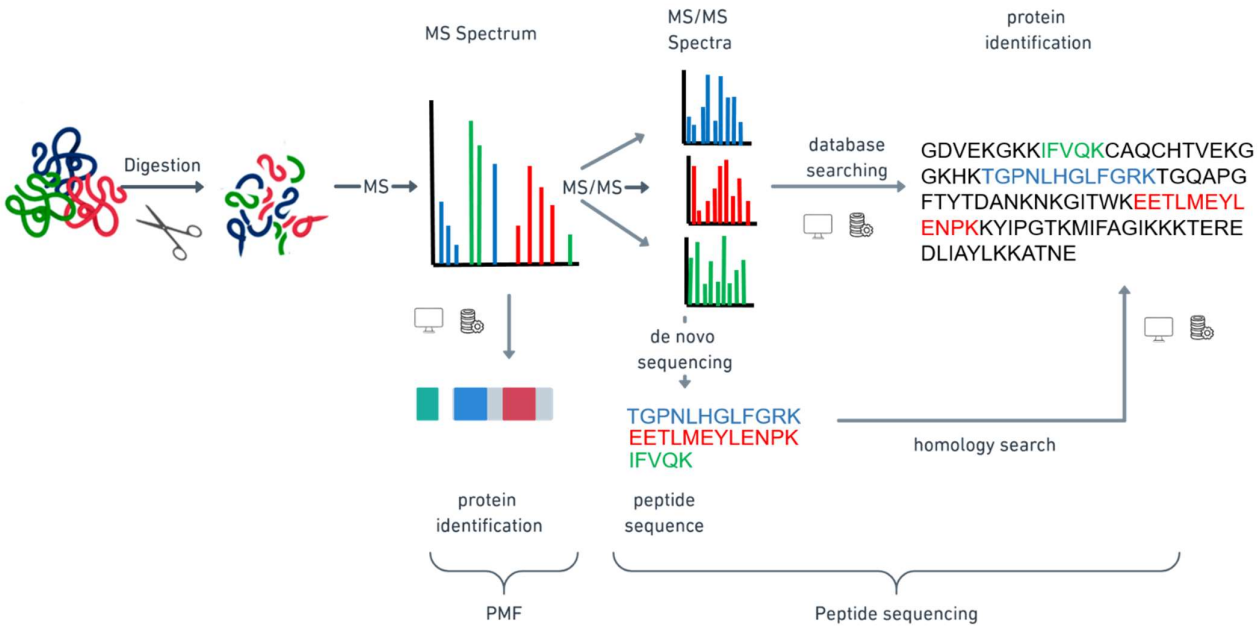


Figure II-1 Scheme of a bottom-up proteomic strategy. The protein solution is firstly digested by a proteolytic enzyme (such as trypsin), to be then analysed with mass spectrometry. PMF strategy consists of the in silico comparison of peptide masses (MS spectrum) with theoretical peptide masses from the protein identification database. In the peptide sequencing approach, individual peptides are isolated and fragmented in the mass spectrometer to produce MS/MS spectra. Protein identification is then performed using the identified sequences of the peptides using bioinformatic tools. The identification is related to the availability of the protein sequence in the protein database. De novo algorithms allow the protein attribution for unsequenced species by identifying the amino acid sequence directly from the MS/MS spectrum. A sequence homology search against the referenced databases might then infer the protein identity.

The first strategy, peptide mass fingerprinting (PMF), consists of the acquisition of intact peptide masses, matched with theoretical diagnostic peptides from the protein public repository [5]. The second approach, peptide sequencing (MS/MS or MS2), involves the isolation and fragmentation of specific peptides in the source of the mass spectrometer's analyser, followed by the reconstruction of the amino acid sequence using bioinformatic tools and protein database [6]. Proteins of species with unsequenced genomes (therefore not present in the database) can be characterised via de novo sequencing based on the direct interpretation of the amino acid sequence using the MS/MS spectrum. A cross-species homology search can follow to identify the peptide sequences [7].

Since pioneering studies in the earlier 2000s [8-10], PMF has been widely implemented as a routine methodology to shed light on the species of origin of the proteinaceous material in artistic and archaeological objects. The strategy ZooMS (Zooarchaeology by Mass Spectrometry) [11, 12] has emerged as a fast and straightforward method for the characterisation of collagen-based objects using PMF of Type I collagen via MALDI-TOF. The approach was extensively implemented in archaeology and palaeontology studies [13-18] and was shown to be effective also with samples collected with a non-invasive technique [19-21]. During the last years, the significant expansion of this technique leads to the construction of a comprehensive database, including collagens peptides diagnostic for different animal breeds, crucial for biological species identifications [21].

The first studies on cultural heritage objects implementing peptide sequencing also date back to the 2000s [8-10, 22, 23]. The strategy, applied using high-performance liquid chromatography coupled with nanoESI-based mass spectrometry, provides a confident and accurate protein identification in significantly heterogeneous and degraded samples. In the last years, numerous developments were achieved on sample preparation, instrument performance, and data processing, enhancing ancient proteins' characterisation [24-27]. Along with the progress on protein identification, increasing attention has been addressed to understand sample ageing and degradation processes through the identification and localization of protein post-translational and/or chemical modifications such as, for example, deamidations or oxidations. [1, 18, 24, 28-30].

All experiments in this doctoral project were performed using nano-ESI LC-MS/MS strategy; therefore, a principal focus will be given to the sequencing approach.

2. Bottom-up workflow

The general lines of the proteomic workflow applied to artwork analyses are shared with the classical protocol for modern biological samples. Nonetheless, specific optimisations at each step are essential to ensure the investigation of proteins extracted from a complex and often degraded matrix in reduced size samples.

The proteomic analysis through bottom-up principally consists of four main steps: (i) protein extraction, (ii) digestion into peptides, (iii) analysis and sequencement via LC/MS-MS analysis and (iv) data interpretation assisted by bioinformatics tools.

2.1 Sampling

One of the principal challenges in artwork studies is the research of a complete and appropriate comprehension of the object composition without inducing any alteration or damage [31, 32]. In the first proteomic studies applied to heritage samples, micro-samples (lower than 10μg) were used [8, 10, 22]. Recent efforts have been made to develop minimally sampling methods to collect a higher number of samples without affecting the object integrity. Examples of proposed protocols were conceived exploiting physical or mechanical actions, like a triboelectric extraction induced by the gentle rubbing of the material surface with polyvinyl chloride (PVC) polymer. The methodology firstly applied to parchment [19] was successfully expanded to other materials as leather [33] and bones[20, 21]. Recently, a further elaboration of PVC-free eraser and the use of microabrasion with fine polishing films (1–30 μm particles, 14,000–600 grit) were presented [34]. The collection of surface peptides through their digestion in situ was also proposed through a functionalised sheet of cellulose acetate modified with fungal proteins Vmh2 hydrophobin in [35] or a hydrophilic gel based on poly (2-hydroxyethyl methacrylate)/poly (vinylpyrrolidone) (pHEMA/PVP) [36], both previously loaded by a trypsin-based solution. Furthermore, other methods based on hydrogels have also been described (e.g. ethyl-vinyl acetate) [37-39].

The bases of these methodologies principally rely on proteomics' capability to identify a proteinaceous material from the identification of only a few peptides. Consequently, the protein yield (despite reduced) from the collection of proteins bound to the material’s surface is often enough to accurately identify proteins.

2.2 Sample preparation

Once the sample is collected, proteinaceous substances need to be extracted from the complex matrix in which they are embedded and then solubilised, principally by breaking their interactions (denaturation). Sample preparation is one of the challenges in the detection and identification of proteinaceous material in cultural heritage objects. The protocols have been adapted to ancient, degraded proteins with reduced solubility (for example, caused by aggregations and cross-links formation occurring during the ageing), embedded in a heterogeneous system, and low amounts. In the last years, several researches were addressed to improve the analyte solubilisation [40-42] and to facilitate the access for the protease to the protein [26]; for example, it has been proven that a deglycosylation step with Peptide-N-Glycosidase F (PNGaseF) improves the ease of investigation of egg-based binders [43].

For the extraction, solubilisation and denaturation of the proteinaceous compounds, different buffering reagents and conditions have been proposed, for example, in acidic conditions (1% TFA [22, 44, 45]), slightly basic (50 mM Ammonium bicarbonate [46]) or alkaline ones (2.5 M NH_3 in water solution [47]). Different chaotropic/denaturing agents are commonly implemented (such as urea [22], guanidinium hydrochloride [48], trifluoroethanol [49] and others), with also detergents to improve the solubilisation of the membrane proteins (like SDS [50] or CHAPS [51]). Heating was also proposed to denature proteins [52]. Furthermore, ultrasonic baths have been often applied in the analyses of artistic [47] and archaeological samples [44].

Protein disulphide bonds are then reduced with 1,4-Dithiothreitol (DTT) [53] and TCEP [54] to further enhance the protease access. A subsequent alkylation, commonly with iodoacetamide IAA, is performed to avoid the reformation of the bonds) [26].

Before proceeding with the digestion, the analyte solution could be desalted and cleaned from the solubilisation products, which might inhibit the protease action (such as urea). Some examples of desalting processes are dialysis [8] and gel filtration [47].

2.3 Protein digestion

In bottom-up proteomics, proteins are digested into peptide mixtures before MS analysis using different enzymatic proteases (more or less specific). Among them, trypsin is the enzyme most commonly used and cause a protein cleavage at the carboxy-terminal side of the lysine and arginine amino acid residues (except if proline is present on the C-terminal side).

This enzymatic protease presents several advantages, like a good specificity and the production of relatively homogeneous peptides with a length favourable for sequencing, which leads to a good quality MS2 spectra and consequently to a confident protein interpretation [55]. Procedure implementing tryptic hydrolysis directly (without protein extraction) [56] have been proposed, also achieving the simultaneous detection of proteins and lipids (through a plate process on colloidal graphite to collect multiple molecular information from one single sample) [57].

Multi-enzymes digestion is often performed to provide a better peptide length distribution with an increase of sequence coverage and the number of identified proteins. For example, trypsin can be combined with the endoprotease Lys C that acts at the same protease sites or with other enzymes targeting different hydrolysis sites, like chymotrypsin, GluC or others [55, 58]. Furthermore, controlled chemical hydrolysis with 25% of TFA assisted with microwave rays was also successfully applied to identify degraded proteins, showing a higher proteins sequence than trypsin-based digestion [29].

Before the LC-MS/MS analysis desalting of peptide solutions can be performed in different manners, including reverse-phase C18 Zip Tip pipette tip [46-48], HILIC [59-61] and SCX [59].

2.4 LC-MS/MS analysis

The peptides solutions are then separated through nano-liquid chromatography (nanoLC), directly coupled with the mass spectrometer. Considering that a mass spectrometer can resolve and analyse a limited amount of peptides species at a time, the good separation efficiency is important to increase the number of information that can be extracted from a single run and to enhance the dynamic range potentially [62].

Low-pH reversed-phase liquid chromatography (RP) implementing a non-polar sorbent material as the stationary phase is used to separate the molecules according to their hydrophobic interactions with the support of a mobile phase. Stationary phases such as C18 are the most used for the separation of complex mixtures of peptides. The peptides' elution is achieved using a solvent gradient with an increasing percentage of organic content (often acetonitrile). The peptide retention time (RT) is referenced on the total ionic chromatogram. NanoLC and nano ESI are mainly employed to analyse complex proteins mixtures from a minimal starting amount of art samples [62].

In BUP studies applied to cultural heritage, the fragmentation methods most widely used are CID and HCD [1]. Even if ETD fragmentation might provide a more comprehensive PTMs investigation, it has been rarely applied to peptide analysis since it principally detects high charge states, missing several doubly charged ions [63].

2.5 Data acquisition

The acquisition schemes of tandem mass spectrometry mostly implemented are two: data-dependent acquisition (DDA) and data-independent acquisition (DIA) [64-66]. The first one is the most used in proteomic studies of cultural heritage in the so-called "discovery-oriented study": the MS instrument firstly scans all precursors eluted in the entire m/z range (full scan) then it selects precursor ions for fragmentation based on their intensity. This technique allows the observation of peptides without knowledge *a priori* of the sample composition, but it may prevent the fragmentation of low-intensity ions. In data-independent acquisition, instead, the instrument selects and fragments all precursors in a specific m/z range avoiding the detection and the selection of individual precursor ions during the analysis [67].

2.6 Bioinformatics analysis

The interpretation of the MS and MS/MS spectra for peptide and protein identification is based on bioinformatics tools using protein database. Databanks such as UniprotKB (encompassing Swiss Prot, which is manually annotated and reviewed, and TrEMBL computationally annotated from genome database and not verified) and NCBI (assembled from different sources) are mainly used. Various algorithms have been developed to digest in-silico the protein sequences present in the databases, to define their fragmentation pattern, and subsequently match the collected information with the acquired MS2 spectra. Examples of search engines are Sequest, Mascot, Andromeda and Byonic [68-71].

Before the analysis, different parameters have to be defined, for example, the enzyme, variable and fixed PTMs, number of missed cleavages, database, precursors and fragment ion errors (in ppm or Da), and fragmentation methods. The results are displayed with a probability score that facilitates their interpretation by defining the quality of peptide-spectrum matches (PSMs). The database search integrates the calculation of false discovery rates (FDR) decoy approach or others [72].

De novo sequencing is an alternative approach, effective especially in the presence of organisms not yet sequenced; the approach leads to a definition of the peptide by a direct interpretation of the experimental MS/MS spectra, without the assistance of a databank [73]. There are different search engines to perform de novo analysis, for example, PEAKS [74]. This software also offers the possibility to expand the search to all 313 PTMs gathered in Unimod database (PEAKS PTM) and to detect peptide mutations (SPIDER) [75]. The potential to identify unsequenced species, discover unexpected modifications, and investigate possible mutations makes PEAKS a worthwhile software in studying proteinaceous compounds from ancient and artistic objects.

3. Aim of the research: implementation of bottom-up proteomics in the protein investigation from wall painting and drawing samples

The research's principal aim was to study two historic samples different in their composition and conservation state, using an optimized bottom-up proteomics approach for the study of very low sample amounts. The analytical design, based on a filter aided protocol and nanoLC-MS/MS analysis (peptide sequencing approach), was applied to:

1. Thirteen wall painting fragments from the Nubian region (Middle Nile Valley) dated between the 6th and 14th century AD;

2. Fifteen micro-samples from seven landscape drawings made by the English artist Thomas Gainsborough during the 18th century.

In both case studies, the proteomic analyses were performed together with other analytical techniques, contributing to an in-depth investigation of the artistic procedure with results otherwise unachievable. One of the BUP research's principal objectives was the precise protein identification and characterisation through the detection of proper sequence coverages despite the high heterogeneity and the degradation conditions of the artworks, as well as the low amount of protein collected.

In the study of Gainsborough's drawings, the concentration of the sample was remarkably reduced (trace analysis) because of the implementation of two minimally-invasive techniques that have been recently developed consisting of a gentle rubbing of the surface through PVC-free erasers [19] and fine polishing films [34]. Furthermore, a careful examination of the PTMs was intended to achieve a better comprehension of the protein.

4. Methodology

4.1 Sample preparation: eFasp procedure

The optimised analytical methodology implemented in bottom-up experiences was adapted from the enhanced filter-aided sample preparation (eFASP protocol) [76].

This protocol was developed to provide very clean peptide mixtures, removing all interfering substances and minimal sample loss.

The first step of the protocol was the protein extraction overnight (at 4°C under agitation) using detergent-based buffer consisting of 4% SDS, 8 M Urea, 50 mM DTT-Dithiothreitol, 0.2% deoxycholic acid (DCA) in 100 mM of Ammonium bicarbonate (pH 8.8). This lysis buffer was conceived to extract, denature and reduce the proteinaceous compound. Sodium dodecyl sulphate (SDS) is an anionic surfactant that solubilises and denatures the proteins resulting in a negatively charged SDS-protein complex. Urea is a chaotropic agent that denature proteins. [58, 77]. DTT is a reducing agent that breaks disulphide bridges, and DCA improves protein extraction and solubilisation and enhances trypsin activity. Finally, Ammonium bicarbonate buffer for protein extraction was used to maintain an alkaline pH that favours protein stability and guarantees the subsequent proteolytic digestion step. Before the incubation, the solid powder samples were crushed with a pestle in the detergent-based buffer to gather the maximum of the analyte. Simultaneously to the extraction step, 10 kDa molecular weight cut-off centrifugal filter (commercial Amicons ®) were incubated overnight in the passivation solution containing 5% (vol/vol) Tween-20 to reduce peptide loss by 300%. The analyte solution was then transferred into the Amicon, preventively rinsed to remove the Tween, and cleaned from the interfering substances with several cycles of exchange buffer (8 M Urea, 0.2% DCA in 100 mM Ammonium bicarbonate) followed by centrifugation and a discarding of the filtrate.

The alkylation was performed by incubating the analyte solution for 1 hour in the dark under agitation with a 100µl of 50 mM of Iodoacetamide in 100 mM ammonium Bicarbonate to avoid the reformation of disulphide bonds.

Before the digestion, the solution was further cleaned with 0.2% DCA in 50 mM of Ammonium Bicarbonate. Protein digestion was achieved by adding a solution of Lys-C/ Trypsin diluted in Ammonium carbonate 50 mM (pH 8.8) with an enzyme to protein mass-ratio of 1:50 followed by incubation overnight at 37°C.

The peptide mixtures collected by centrifugation was subsequently submitted to a liquid-liquid extraction using ethyl acetate with trifluoroacetic acid; three washing steps of the analyte with ethyl acetate followed. The aqueous solution was finally evaporated by a speed vacuum system and then re-suspend in Milli-Q water acidified with acid formic for the mass spectrometry analysis (Figure II-2).

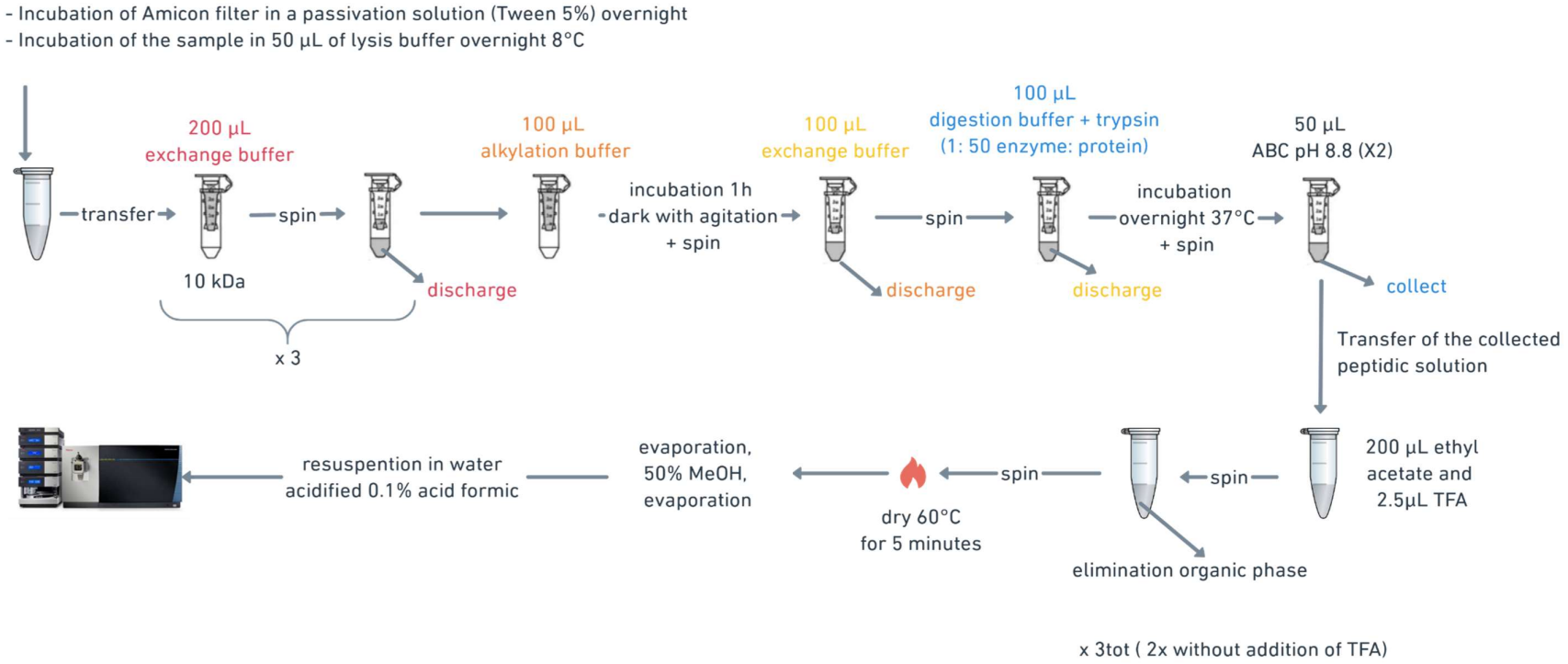


Figure II-2 Scheme of bottom-up eFasp methodology.

4.2 Nano-LC- MS/MS

Two Thermo Fisher Scientific Orbitrap mass analysers were implemented in the analyses: Q-exactivePlus and the FusionTM LumosTM TribridTM. Both systems were connected to a Thermo Fisher Scientific U3000 RSLC nanofluidic system. The samples were diluted in deionised water solution acidified with 0.1% formic acid (FA). Then, one µL of peptide mixture was loaded onto a 5-mm C18 PepMapTM trap column with 300-µm inner diameter, granulometry 5 µm (Thermo Fisher Scientific) at a flow rate of 10 µL/min with 0.1% formic acid. Peptides were then eluted from the trap column onto a reversed-phase C18 PepMapTM 50-cm column with 75-mm internal diameter granulometry 3 µm, (Thermo Fisher Scientific). Precolumn and column were both set at a temperature of 45 °C.

The solvent A was 5% acetonitrile (ACN) and 0.1% FA, and solvent B was 75% ACN and 0.1% FA. The gradient was set using the following parameters: 0–3 min 1% B, 3–6 min 5% B, 6–240 min 40% B, 240– 250 min 99% B, 250–260 min 99% B, 260–265 min 1% B, 265–285 min 1%B. The elution flow rate was set at 300 nL/min; the mass spectrometer operated in positive ion mode at a 1.9-kV needle voltage. The settings used for mass spectrometry (MS) analysis were: resolution 70,000; AGC target 1e6; Maximum IT 100 ms for MS and 160 ms for MS/ MS; scan range m/z 350–1600. Data were acquired using Xcalibur 4.1 software in a data-dependent mode. MS scans were recorded in the m/z 375-1500 range at a resolution of R=120,000 (at m/z 200), while MS/MS scans were collected in the orbitrap at a resolution R=30,000 (at m/z 200).

Ions between 2^+ to 7^+ charges were selected for high-energy collisional dissociation (HCD) fragmentation. Blanks and positive and negative controls were injected and tested alongside the historical samples to prevent contamination and carryover.

4.3 MS- data processing

Raw data files generated from nano-LC-ESI-MS/MS spectral acquisition were searched using PEAKS 8.5 software, firstly against the general and updated UniProtKB/Swiss-Prot database followed by a second search against a restricted database on the base of identified proteins to improve the number and the reliability of peptides. General parameters for data analysis were set as follows: 3 missed cleavages maximum; precursor ion error tolerance 10.0 ppm; fragment ion error tolerance 0.02 Da.

Carbamidomethylation was set as a fixed modification, while oxidation (methionine, proline), deamidation (asparagine, glutamine) were selected as variable modifications. For the investigation of milk proteins in Gainsborough drawings, phosphorylation (serine, threonine) and lactosylation (lysine, arginine) were also added. Protein false discovery rate (FDR) was set to 0.1% with a -10lgP ≥18 score threshold for peptides.

A manual filter of at least two different non-overlapping peptides was fixed for protein identification. Data resulting from the injection of positive and negative controls and blanks were carefully examined before injecting the historical samples to verify the possible presence of contaminants or residues from carryover. Results from PEAKS PTMs and PEAKS SPIDER's unbiased research were inquired to investigate unexpected modifications and amino acids substitutions. MS/MS spectra of each peptide of interest were then manually inspected.

Finally, the peptides characteristic of a specific biological species were investigated through alignment of experimental sequences against the whole nrNCBI database via Basic Local Alignment Search Tool (BLAST).

5. Analysis of wall painting samples from the Nubian region

A mass spectrometry-based proteomic study via a bottom-up approach was performed on thirteen samples collected from Nubian wall paintings manufactured in different churches and building of the Middle Nile Valley between the 6th and 14th century AD. The fragments were collected from different museum collections (including the Sudan National Museum in Khartoum, National Museum in Warsaw, Humboldt University in Berlin, Polish Centre of Mediterranean Archaeology, University of Warsaw) and supplied by Dr Dobrochna Zielinska (Institute of Archaeology of the University of Warsaw).

The proteomic analysis contributed to a more extensive study composed of different analytical methodologies (principally XRF and spectroscopic analyses) aimed to better understand the painting technique and materials (plaster, pigments, binders and lime washes) during the Nubian mediaeval kingdom, and consequently to learn more about the economics, patronage and commerce of the time.

Protein investigation was conducted to verify the presence of secco ("dry") painting technique consisting in the application of a mixture of organic material and pigment on the dry plaster [78].

The analyses' results have been discussed in chapter 5 of the published book "Medieval Nubian Wall Paintings – Techniques and Conservation" [79].

5.1 Historic samples

Thirteen wall painting fragments were investigated for protein analysis. The museum curators carefully selected the wall paintings not contaminated from previous restoration treatments to ensure the study of the original painting technique. This condition resulted in a collection of sample mainly from the kingdom of Makuria, which corresponds to modern-day Northern Sudan and Southern Egypt. The painting production ranges from the beginnings of the kingdom (before the Christianization) around the 5-6th century until the 12th century. One single sample was coming from the Kingdom of Alwa, actual central and southern Sudan, manufactured around the 9th and the 11th century. Several tens of micrograms (estimated by the sample size assuming a density of 1) of the powdered samples were collected from each fragment by carefully scraping the mural-painted surface with a scalpel. Two samplings were performed in two different coloured areas of four fragments to investigate if there were differences in the colour application; consequently, the proteomic analysis was performed on a total of seventeen samples. The nomenclature, the origin and the sampling location are outlined in Annexe i.

5.2 Analysis and discussion of the results

The analysis evidenced the presence of several proteins in all samples. A considerable percentage of the signal derived from the contamination of exogenous proteins related to the handling activity prior and throughout the samplings, consisting of human skin and hair (e.g. keratins) and sweat proteins (e.g. protein S-100, serpin). More interestingly, egg white proteins were identified in all seventeen samples analysed; collagen proteins were also observed in some samples (Table II-1)

5.2.1 Egg proteins

Two major egg white proteins, ovalbumin and lysozyme (which constitute 54% and 3.4% of protein content in the chicken egg) [80], were identified with remarkable sequence coverages through the confident fragment assignments of several peptides. Ovalbumin (45 kDa protein) was found in the majority of samples with a sequence average of 40%; for instance, in the sample D_BV_6 (red pigment) from the Old Dongola, Church of Archangel Raphael (9^{th} century), 46% of ovalbumin protein was successfully detected with 17 peptides and 12 unique peptides.

Figure II-3-A reports the MS/MS spectrum of a triply charged ion at m/z 761.0650 (Δm= -0.7 ppm) detected in this sample, which provided the determination of the ovalbumin peptide ^{86}DILNQITKPNDVYSFSLASR^{105} by its y (from y_1^+ to y_{18}^+) and b (from b_2^+ to b_8^+ and b_{11}^+, b_{12}^+) ion fragments. Lysozyme (14 kDa) was also detected in all samples with a protein coverage averaging 60%. Spectra showed great signal to noise ratios and clear fragmentation pattern with a comprehensive y- and b- ions series from the protein peptides, leading to specific sequence assignments. In the spectrum of the doubly charged ion m/z 878.4036 (Δm= 1.2 ppm) (Figure II-3-B), for instance, the identification of the fragment ions from y_1^+ to y_{14}^+ and b_2^+ to b_{12}^+ provided its attribution to the lysozyme peptide ^{46}nTDGSTFYGILQInSR^{61} presenting two deamidations on the asparagine residues N46 and N59.

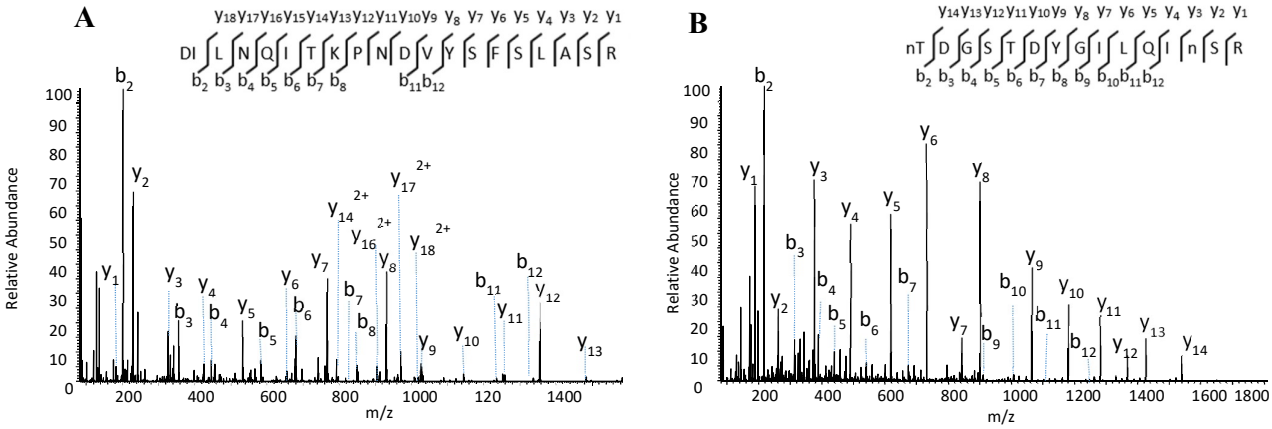


Figure II-3 MS/MS spectrum of **(A)** the triply charged ion at m/z 761.0650 (Δm= -0.7 ppm) belonging to the ovalbumin in the sample Gh_01 (Ghazali Monastery, North Church dated second half of the 7th century). The sequence 86 -105 was identified using y_1^+ to y_{18}^+ and b_2^+ to b_8^+; **(B)** the doubly charged ion at m/z 878.4036 (Δm= 1.2 ppm) related to lysozyme peptide 46- 61 in sample D_BV_6. The fragmentation pattern of y_1^+ to y_{14}^+ and b_2^+ and b_{12}^+ provided an exact sequence assignment and the localisation of two deamidations on N(46) and N(59).

A significant amount of deamidations and photo-oxidative modifications were determined in the detected proteins, as highlighted in the lysozyme sequence coverage from the sample Gh_01 (Ghazali Monastery, North Church dated second half of the 7th century) in Figure II-4. Deamidations (Δ= +0.98 Da) were noticed in almost all asparagine and glutamine residues, while all methionine residues were identified with oxidations (Δ= +15.99 Da). All tryptophan residues were also detected with oxidations and di-hydroxylation (Δ= +31.99 Da).

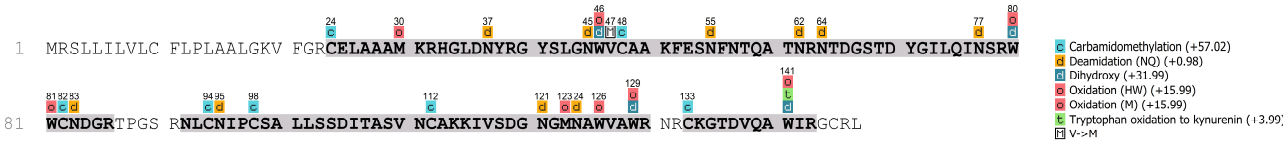
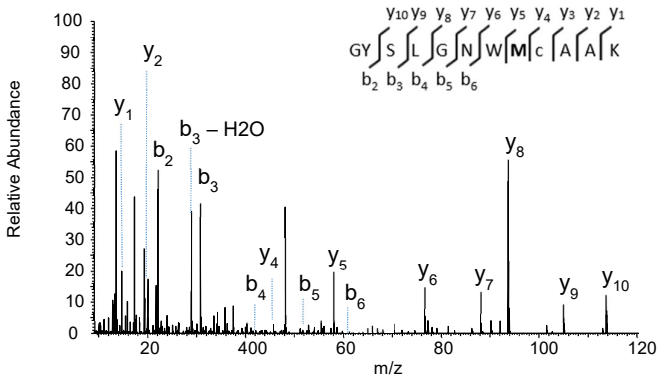


Figure II-4 Lysozyme sequence coverage from the sample Gh_01 (red pigment area). The sample is from the Ghazali Monastery, North Church, in the second half of the 7th-century. The protein was characterised with 77% of coverage through the identification of 32 peptides.

The significant presence of deamidation modifications was consistent with ancient proteins since the amino acid residues Asn and Gln naturally deamidate over time [24, 81]. Instead, the oxidation of tryptophan could be characteristic evidence of UV damage, as already proved in biological [82, 83], archaeological [84] and artistic [48] materials. This amino residue is photosensitive [85, 86]. Its oxidative modifications might be caused by several factors such as ageing, light processes or other pathways [87, 88]. The sample preparation may also induce these oxidations, and the underlying causes are often hardly distinguishable [89].

The unbiased research of modifications and substitutions performed by the PEAKS SPIDER algorithm highlighted the presence of specific peptides not assignable to *Gallus gallus* or other species sequenced in the public database. For example, the lysozyme peptide ^{22}GYSLGNWVCAAK^{33} was detected with substitution of the residue valine (V) by the methionine (M) in three samples manufactured between the end of the 7th century and the 9th century (Gh_01, D_H_A_18, D_BV_6). The localisation of the substitution in the lysozyme sequence is illustrated in Figure II-4, while Figure II-5 shows the MS/MS spectrum of the peptide sequence GYSLGNW**M**cAAK.

In the MS2 spectrum of the doubly charged precursor ion at m/z 663.3191 (Δm= 0.4 ppm) in Figure II-5, the amino acid substitution (V->M) was highlighted by the presence of the full y- fragmentation pattern as well as b_2^+ - b_6^+ fragment ions. The alignment of this experimental peptide sequence with all sequenced species present in the NCBI database did not show any species correspondence, with only a partial alignment of 91.67% for *Gallus gallus*. Even though all mural paintings were identified as having been formulated with an egg-based binder, eggs from oviparous animals not referenced in the current database were used in the formulation.



Sequences producing significant alignments:

Description	Max Score	Total Score	Query cover	E Value	Per. Ident	Accession
lysozyme [Gallus gallus]	38.8	38.8	100%	0.047	91.67	ACL81571.1
lysozyme C [Meleagris gallopavo]	38.8	38.8	100%	0.047	91.67	XP_003202118.1
lysozyme [Francolinus pondicerianus interpositus]	38.8	38.8	100%	0.047	91.67	ACL81753.1
LYSC protein [Eudromia elegans]	38.8	38.8	100%	0.047	91.67	NXA33732.1
lysozyme [Bambusicola thoracicus]	38.8	38.8	100%	0.047	91.67	ACL81751.1
LYSC protein [Crypturellus soui]	38.8	38.8	100%	0.047	91.67	NWI10981.1
LYSC protein [Odontophorus gujanensis]	38.8	38.8	100%	0.047	91.67	NXJ04694.1

Figure II-5 MS/MS of doubly charged precursor ion at m/z 663.3191 (Δm= 0.4 ppm) related to the peptide 40-51 of the lysozyme from sample D_H_A_18 (Old Dongola, House A 8th century). The peptide sequence was identified through the full y- fragmentation pattern and b_2^+ - b_6^+ fragment ions. The sequence presents a homology with *Gallus Gallus* species a part of an amino acid substitution (V28 to M28). The experimental peptide sequence alignment with all sequenced species present in the NCBI database did not show any correspondence with sequenced species.

Along with ovalbumin and lysozyme, another significant egg white protein, ovotransferrin (77 kDa), was detected in the sample US. 22_8c (church on Us Island, 10th/11th century). Even if only two peptides were observed for a total of 3% sequence coverage, the good peptidic fragmentation (Figure II-6–A) highly suggested the validity of their presence, which in turn strengthened the hypothesis of egg white proteins.

Furthermore, two egg yolk proteins, vitellogenin 1 and 2, were detected in the sample Se_5_2 (Selib, end of 6th/7th century). Both proteins of very high molecular weights (205 and 210 kDa, respectively) were identified with low sequence coverage (1% and 3%). Nevertheless, their actual presence was strongly supported by clear sequence assignments via MS2 spectra characterised by clear fragmentation patterns with almost full y and b ion series (Figure II-6-B).

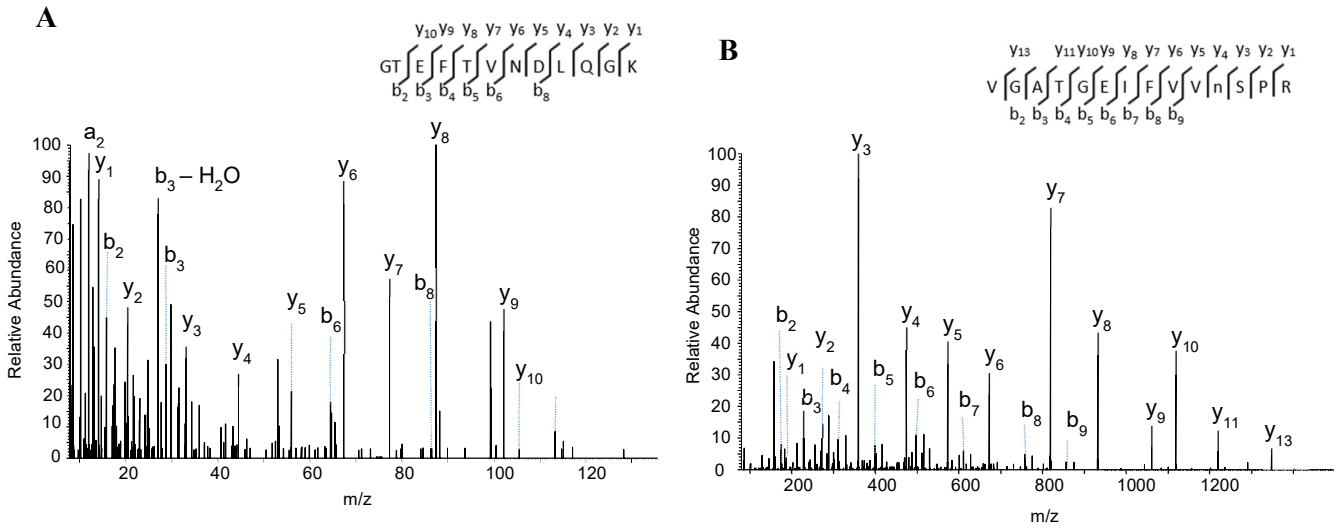


Figure II-6 MS/MS of **(A)** doubly charged ion at m/z 654.8245 (Δm= -1.0 ppm) attributed to ovotransferrin in the sample US. 22_8c (church on Us Island, 10th/11th century). The sequence 120 -131 was identified using y_1^+ to y_{10}^+ and b_2^+ to b_8^+ and b_8^+, b_{12}^+ fragments; **(B)** doubly charged ion m/z 723.8826 (Δm= 0.1 ppm) attributed to the peptide ^{642}VGATGEIFVVnSPR^{655} of vitellogenin 1 from the sample Se_5_2 (Selib, end of 6th/7th century). A clear and almost full fragmentation pattern localised a deamidation on asparagine (N11).

The application of egg with nut oil as a binder was observed in Nubian monuments (6th/7th-century Cruciform Building murals) [90] or other Egyptian monuments, such as the tomb of Nefertari [91], but its presence is problematic interpretation. A possible interpretation of egg white presence provided by the archaeologists was as an external varnishing layer (i.e., white glaze) [91-94].

5.2.2 Collagen proteins

Trace amounts of collagens α1 type I and α2 type I were detected in four samples under investigation (D_BV_6, US. 22_8c, D_D_1 and D_H_A_18), all dated to the period from the 8th to the 11th century.

These high molecular weight proteins (140 kDa and 130 kDa) are commonly present in bones, tendons, or skin of vertebrates implemented in animal glue preparation [95]. As an example, in the specimen US. 22_8c (white) α1 type I and α2 type I collagens were identified with sequence coverage of 18% and 16%.

Figure II-7 shows the MS2 spectrum of the doubly charged ion at m/z 443.7227 (Δm= 0.6 ppm); the observation of a full y-ions pattern and b_2^+ - b_4^+ fragment ions provided its exact assignment to the peptide ^{361}GSEGPQGVR^{369} of α1 type I collagen.

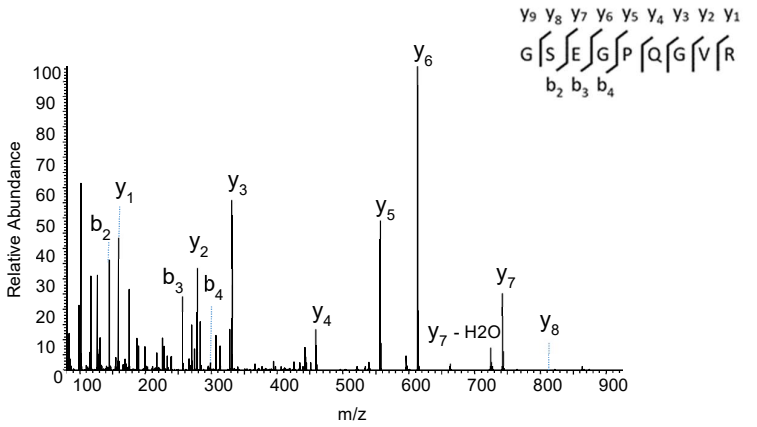


Figure II-7 MS/MS spectrum of the doubly charged ion at m/z 443.7227 (Δm= 0.6 ppm) from the sample US. 22_8c white (church on Us Island, 10th/11th century) presenting the full fragmentation pattern of y- and three b- ions fragments. The ion was assigned to the peptide ^{361}GSEGPQGVR^{369} of the α1 type I collagen.

In the mural fragments D_BV_6, US.22_8c, and D_D_1, collagen proteins were observed in each different coloured area (grey and red areas; brown and white areas; light brown and the grey regions, respectively). These observations led to assume that a collagen-based glue was applied regardless of the pigment implemented.

Collagen-based animal glue was a known adhesive in Egypt, and its application was suggested as glue either for ground or paint layer in the production of several ancient Egyptian mural paintings [96-101]; its presence might be ascribed as a binding medium in the paint layer or lower, in the ground (preparatory) layer. If the latter, it could be possible that both the paint layer and some ground beneath were collected during the sampling.

The results achieved suggest that the application of egg-based compounds was used in all mural paintings under investigation, while the animal glue was identified only in some mural paintings dated between the 8th and 11th centuries (Figure II-8).

The archaeologists have suggested that the binder's homogeneity may be related to the strong position of Old Dongola as a centre of authority and artistic centre. Its influential role was also observed in other aspects of art, such as pigments' choice. Unfortunately, due to the lack of comparative analyses, these are still suppositions [79].

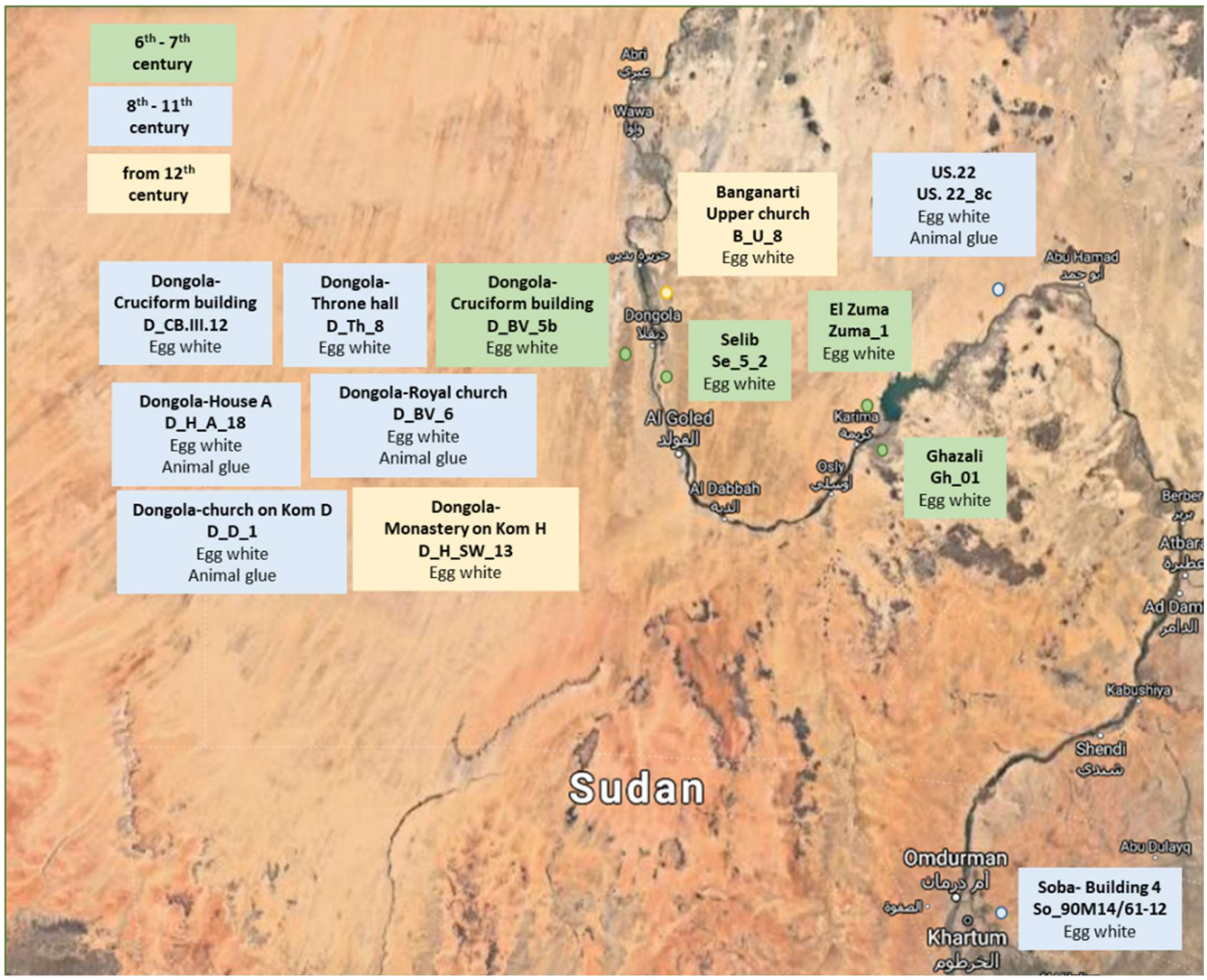


Figure II-8 Distribution of samples under investigations and the detected proteinaceous compounds

Table II-1 Selection of proteins identified in the samples from the Nubian wall paintings analysed.

Sample (colour)	Sample origin	Proteins	Accession	Coverage (%)	Score (-10lgP)	Peptide	Unique peptides	MW (Da)
Zuma-1 (ochre)	El-Zuma	Lysozyme	P00698	65	169.74	20	8	16239
		Ovalbumin	P0112	38	149.86	10	10	42881
Se_5_2 (brick red)	Selib	Lysozyme	P00698	41	142.11	8	5	16239
		Vitellogenin-1	P87498	1	26.37	1	1	210629
		Vitellogenin-2	P02845	3	99.96	6	6	204807
D_CB.111_12 (red-orange)	Cruciform Building Dongola	Lysozyme	P00698	44	116.13	6	3	16239
Gh-01 (red)	Katholikon, Ghazali	Lysozyme	P00698	77	184.03	32	3	16239
		Ovalbumin	P0112	47	157.91	14	11	42881
Gh-01 (light brown)	Katholikon, Ghazali	Lysozyme	P00698	63	155.73	18	11	16239
		Ovalbumin	P0112	39	152.63	11	11	42881
D-H-A-18 (beige-brown)	House A, Dongola	Collagen α 1(I)	P02453	13	143.99	15	1	138939
		Collagen α 2(I)	P02465	16	153.94	16	9	129064
		Lysozyme	P00698	63	172.15	18	9	16239
		Ovalbumin	P0112	32	150.05	9	9	42881
D-BV-6 (grey)	Royal Church, Dongola	Collagen α 1(I)	C0HJN7	11	119.52	10	3	85634
		Collagen α 2 (I)	P02465	14	145.02	15	3	129064
		Lysozyme	P00698	69	219.27	29	4	16239
		Ovalbumin	P0112	39	177.24	14	14	42881
D-BV-6 (red)	Royal Church, Dongola	Collagen α 1 (I)	P02453	21	166.44	22	1	138939
		Collagen α 2 (I)	P02465	16	153.81	17	4	129064
		Lysozyme	P00698	76	248.34	43	20	16239
		Ovalbumin	P0112	46	180.85	17	12	42881
D-BV-5b (green)	Royal Church Dongola	Lysozyme	P00698	41	127.27	7	4	16239
D-D-1 (light brown)	Church on Kom D, Dongola	Collagen α 1 (I)	P02453	23	150.67	17	0	138939
		Collagen α 2(I)	P02465	17	147.48	19	1	129064
		Lysozyme	P00698	61	174.92	22	4	16239
		Ovalbumin	P0112	44	159.89	13	13	42881
D-D-1 (grey)	Church on Kom D, Dongola	Collagen α 1 (I)	P02453	*3*	*79.38*	*4*	*4*	*138939*
		Collagen α 2(I)	P02465	*5*	*104.51*	*5*	*5*	*129064*
		Lysozyme	P00698	61	181.50	19	3	16239
		Ovalbumin	P0112	38	161.43	11	11	42881
D-Th-8 (beige)	Throne Hall, Dongola	Lysozyme	P00698	67	223.16	30	8	16239
		Ovalbumin	P0112	39	153.73	12	12	42881
So-90M14/61-12 (light grey)	Soba, Building 4 (Church)	Lysozyme	P00698	61	181.36	22	9	16239
		Ovalbumin	P0112	47	178.94	15	15	42881
US. 22_8c (brown)	Us Island, church	Collagen α 1 (I)	P02453	13	146.28	14	2	138939
		Collagen α 2 (I)	P02465	11	139.07	11	11	129064
		Lysozyme	P00698	45	179.11	16	11	16239
		Ovalbumin	P0112	36	165.43	12	12	42881
		Ovotransferrin	P02789	3	45.57	2	1	77777
US. 22_8c (white)	Us Island, church	Collagen α 1 (I)	P02453	18	151.63	19	2	138939
		Collagen α 2 (I)	P02465	16	159.70	19	8	129064
		Lysozyme	P00698	59	197.81	23	6	16239
		Ovalbumin	P0112	42	175.67	13	13	42881
D-H-SW-13 (light brown)	Monastery on Kom H, SW Annexe, Dongola	Lysozyme	P00698	50	126.76	8	3	16239
		Ovalbumin	P0112	18	75.38	5	5	42881
B_U_8 (red)	Upper Church, Banganarti, Chapel 3	Lysozyme	P00698	54	180.30	12	3	16239
		Ovalbumin	P0112	44	198.88	13	13	42881

6. Analysis of Thomas Gainsborough drawings collected with the micro-invasive sampling methods

6.1 Introduction

Bottom-up based proteomic analysis was applied to the investigation of seven drawings painted by the English artist Thomas Gainsborough (1727-1788) during the 18th- century. The study was investigated in partnership with Dr Julie Arslanoglu and Dr Federica Pozzi (*Department of Scientific Research, The Metropolitan Museum of Art, New York*) and Reba Snyder (*Thaw Conservation Center, The Morgan Library & Museum*).

Thomas Gainsborough is a central figure in 18th-century art, who was highly renowned for his portraits and landscape paintings. Furthermore, the artists produced over 900 drawings, experimenting with various materials on paper, such as graphite, chalk, oil, watercolours, pastel, and protein-based compounds to obtain original visual effects. Over the last 20 years, conservators and art historians have proven the innovation in Gainsborough's technique, which was characterised by unique drawing and painting methods. The artist accomplished real "experiments on paper" with the disruption of the traditional drawing rules: new visual effects were each time achieved through the refinement of original styles. A significant source of information about Gainsborough techniques is the letter that the artist wrote to his friend William Jackson in January 1773 (currently stored at Yale Center for British Art). In this letter, he provided a brief description of his complicated recipe for making *"oils on paper",* implementing unconventional techniques, which have subverted the classical principle of the art contemporary to him.

The artist in the letter documented the application of dry lead white pigment on the paper surface and the subsequent immersion of the entire drawing in milk multiple times to enhance the attachment of pigments and prevent discolouring. The wet drawing adhered to a temporary wooden frame or stretcher lined with paper. Auxiliary light and dark layers were added, followed by a second immersion in milk. Colours were then applied together with an Arabic Gum glazing. The final steps comprised the application of three varnish layers (natural resin) on both sides (to keep the paper flat) and removing the drawing from the strainer and its mounting. [102, 103]. It is noteworthy that Gainsborough specifically asked his friend to keep the recipe secret: *"Swear now never to impart my secret to anyone living"* [104].

The proteomic research contributed to an in-depth scientific study technical study of Gainsborough's works housed at The Morgan Library & Museum.

The research aimed to get an insight into the artistic technique, focusing on possible surface coatings and underdrawings (magnification under transmitted and raking light) and on the characterisation of the different white pigments and potential fixatives used by the artist (XRF and Raman spectroscopy). Proteomic analysis particularly allowed the detection and characterisation of milk-based fixatives potentially applied by the artist as pigment fixative in some of his drawings [103, 105, 106].

6.2 Milk proteins

Milk proteins can be subdivided into (i) caseins, (ii) whey proteins, and (iii) milk fat globule membrane proteins (MFGM). The four caseins (αS1-, αS1-, β-, κ- caseins) are low molecular weight phospho-proteins of 20-25 kDa, which together cover the 78% (w/w) of the total protein milk content; their singular proportions in cow's milk are 38%, 10%, 35%, and 12%, respectively [107]. The caseins are characterised by several allelic variations principally consequent to an amino acid substitution: 12 in β-CN, 13 in κ-CN, 9 in α_{S1}-CN, 4 in α_{S2}-CN [108-110]. A large extent of post-translational modifications was widely identified, like variable extents of serine or threonine phosphorylations detected ranging from 4P to 5P in β-CN, 8P to 9P in α_{S1}-CN, 10P to 13P in α_{S2}-CN, 1P to 3P in κ-CN. The caseins exploit these phosphorylation sites to assemble themselves into micelles (colloidal structures) and, in turn, to stabilise their structure.

Furthermore, glycosylation has been frequently localised in threonine and serine residues of the caseins, principally in the case of the second half of κ-CN, named casein macropeptide (CMP). This polypeptide, formed due to cleavage at the F105-M106 peptide bond by chymosin or pepsin, presents eight potential residues that are often glycosylated T142-T152- S153- T154- T157- T163- S170- T186. The kappa casein structure also has a disulfide bond formed between C32-C109 [108, 111-113]. Furthermore, proteolysis of caseins in shorter polypeptides caused by indigenous milk enzymes has been widely documented [114, 115].

In bovine milk, whey proteins constitute 17% of protein content. The most abundant proteins are β-lactoglobulin (β-LG), α-lactalbumin (α-LA), bovine serum albumin (BSA), immunoglobulins (IG), representing 60%, 20% 10%, and 10% of total whey proteins [116]. Bovine lactoferrin (BLF), serotransferrin, and lactoperoxidase (LP) constitute other minor whey proteins.

As in the case of caseins, also β -lactoglobulin and α -lactalbumin present genetic variants (11 β-LG and 3 α-LA). Disulphide bonds have been observed in β-lactoglobulin (82-176; 122-137 alternated with 122-135) and α-lactalbumin (25-139; 47-130; 80-96; 92-110) [116].

Milk fat globule membrane (MFGM) account for 1-2% of the protein content and consist of lipid (40%), protein (60%), and cholesterol [117, 118]. Precisely, the triacylglycerol core is prior coated by a phospholipid-monolayer connected by a second outer phospholipid bilayer through a proteinaceous coat. Several MFGM proteins, including butyrophilin and lactadherin (milk fat globule-EGF factor 8), are glycoproteins located in the external membrane surface and xanthine dehydrogenase/oxidase, which is situated among the outer glycerolphospholipid bilayer and the inner glycerolphospholipid monolayer (Figure II-9).

Perilipins are phosphorylated proteins (the so-called PAT family) that, similarly to adipophilin (ADPH), are capable of binding with the surface of lipid droplets present in the cytoplasm, stabilising the latter from the enzymatic lysis [119, 120].

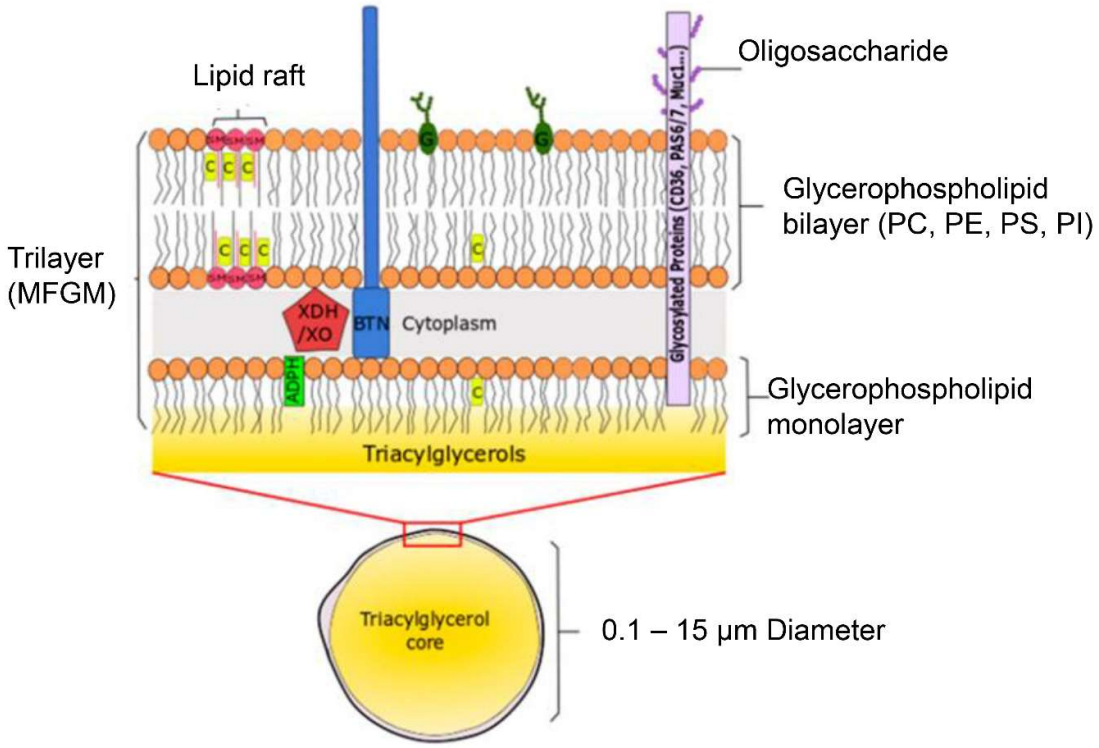


Figure II-9 Milk fat globule membrane. The triacylglycerol core is prior coated by a phospholipid-monolayer connected by a second outer phospholipid bilayer through a proteinaceous coat. Butyrophilin and lactadherin are on the external membrane surface, while xanthine dehydrogenase/oxidase is located among the two layers. Figure extracted from Lordan et al., 2017.

6.3 Samples

6.3.1 Sampling method

During this proteomic study, the use of two micro-sampling techniques was proposed. One protocol consisted of a rubbing action on the material surface with small cubes (approximately 3-5 mm) of PVC-free eraser (Staedtler) to obtain a triboelectric extraction of proteins [19].

The other protocol involved a micro superficial abrasion through small discs (Ø 5-6 mm) of aluminium oxide Fiber Optic Polishing Film Discs (PFP) of 6- and 15-µm grits purchased by Precision Fiber Products, Inc., formerly attached to a polystyrene rod enclosed in a nickel Starette Pin Vise (Grainger) [34] (Figure II-10).

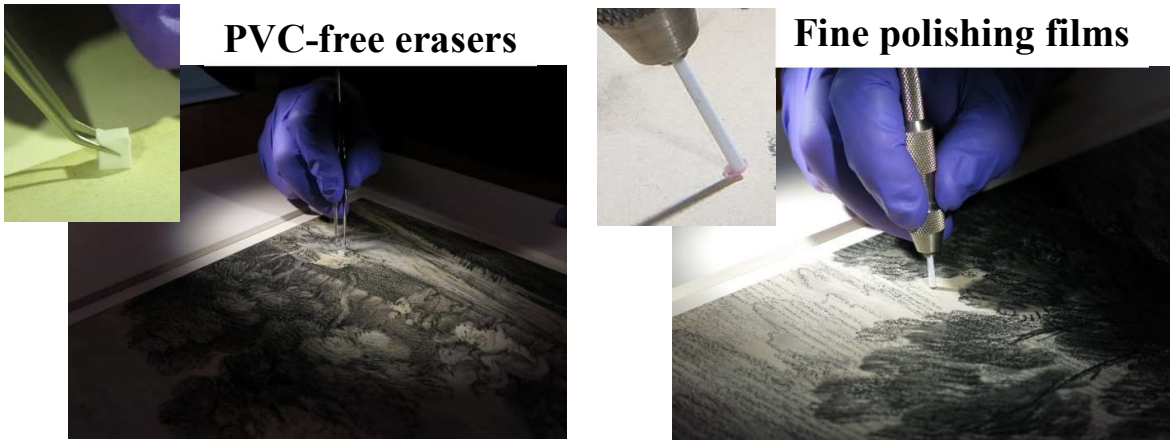


Figure II-10 Sampling techniques using PVC –free erasers and fine polishing films. Gainsborough, Thomas (1727-1788) Hill Landscape with Cows on the Road. ca. 1780. III, 62 Photo taken by Reba F. Snyder, Paper Conservator, Thaw Conservation Center and Lindsey Tyne, Associate Paper Conservator, Thaw Conservation Center, The Morgan Library & Museum.

6.3.2 .Standard samples

Reference samples were collected from glass and paper supports with small PVC-free erasers, 6- and 15-µm grits papers to compare the different sampling procedures. Mock-up preparation and micro-sampling were performed under a microscope by Julie Arslanoglu, *Department of Scientific Research, The Metropolitan Museum of Art, New York*.

6.3.3 Gainsborough's drawings

Seven of Gainsborough's drawings from The Morgan Library & Museum were investigated. Fifteen samples were collected from these historical artworks by Julie Arslanoglu, *Department of Scientific Research, The Metropolitan Museum of Art, New York*, using minimally invasive techniques. Both PVC free erasers and 6- and 15-µm film discs were transferred in Eppendorf of 1.5 ml and entirely covered with lysis buffer for the extraction before the eFasp digestion. A list of the studies artworks is provided in Annexe ii.

6.4 Analysis and discussion of the results

6.4.1 Standard samples

Reference samples of whole and skim milk on glass and paper supports (sampled with small PVC eraser, 6- and 15-µm grits) were investigated. The principal objectives were the assessment of the analytical protocol and a comparison of the different sampling techniques. Even though the analytes were at trace levels (dust almost invisible), the investigation succeeded in identifying a high number of milk proteins belonging to all three protein groups, caseins, whey proteins and MFGM (Table II-2). Significant similarities were observed among the results of the different micro-sampling protocols: caseins were detected in all samples with high coverage (except for sample J, which probably encountered sampling issues) independently of the sampling, type of surface or milk investigated.

For example, casein α_{S1}-CN in whole milk was identified with 89%, 72%, and 76% of sequence coverage in sample A (small PVC eraser), B (6 µm grit) and C (15 µm grit), respectively.

Spectra with good signal to noise ratios and abundant fragmentation patterns ensured a straightforward characterisation of peptide sequences, as in the case of the peptide 119-134 of α_{S1}-CN casein, attributed from the interpretation of the doubly charged ion at m/z 976.4818 (Δm=2.0 ppm) using fragments ions from y_1^+ to y_{15}^+ and from b_2^+ to b_9^+, b_{11}^+. The peptide sequence presented phosphorylation on serine S130 (Figure II-11–A).

All major whey proteins were also characterised in almost all samples; for instance, beta-lactoglobulin was characterised with high coverage in the majority of samples (i.e., 92% in sample A and sample C). Figure II-11– B reports as a demonstration of the almost full sequence of the beta-lactoglobulin peptide (141 -154) detected in sample A.

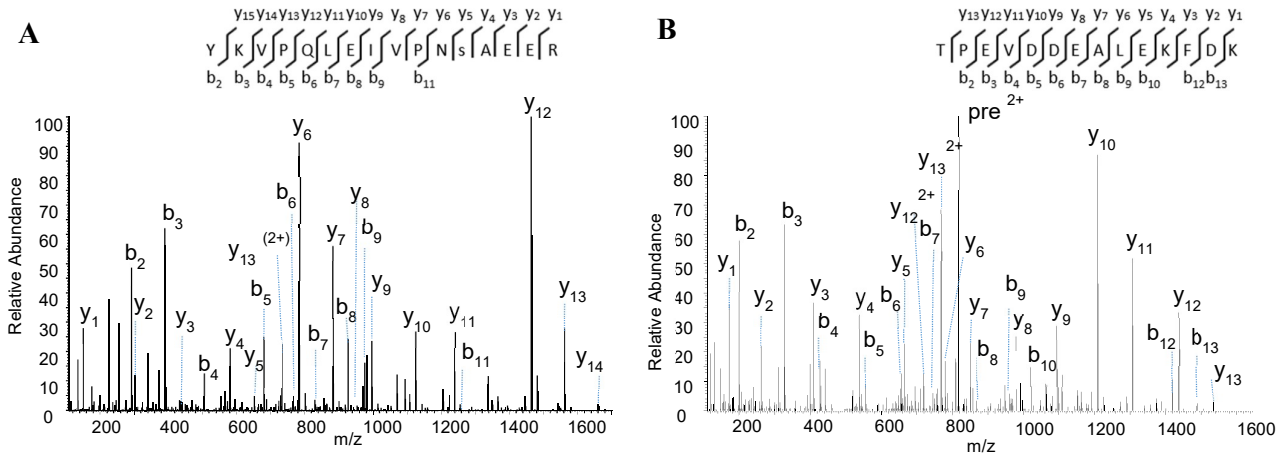


Figure II-11 MS/MS spectra from the sample A (mock-up standard, PVC eraser of whole milk) **(A)** of the doubly charged ion at m/z 976.4818 (Δm=2.0 ppm) . The $_{\alpha S1}$-CN (119 – 134) peptide was identified using y_1^+ to y_{15}^+ and b_2^+ - b_9^+ and b_{11}^+ fragment ions. The peptide sequence presents a phosphorylation on serine S130**; (B)** of the doubly charged ion at m/z 818.3922 (Δm=1.4 ppm) . The beta- lactoglobulin (141 -154) peptide was identified using y_1^+ to y_{13}^+ and b_2^+ - b_{10}^+ and b_{12}^+, b_{13}^+ fragment ions.

Nonetheless, this protein remained undetected in two samples (sample L and sample BB). An analogous situation was pointed out for α-lactalbumin, which was detected in all samples of whole milk but not in samples J and K (both skim milk). Another whey protein, lactoperoxidase (a milk enzyme), was detected only in two samples (11% in sample C and 4% in sample BB). Further experiments (experiment repetitions) would have been necessary to complement these results.

Finally, four low abundance MFGM proteins were identified in the samples, assessing the methodology's potential also for low abundant analytes. Butyrophilin was detected in all samples with a significant coverage ranging around 30% in whole milk samples and 15% in skim milk samples, while xanthine dehydrogenase/oxidase and lactadherin were observed in almost all samples with variable proteins coverage. Perilipin, another MFGM protein, was detected in only one sample but with a relatively good sequence coverage of 20% with 6 peptides.

Both micro-sampling techniques allow confident identification of the 3 main types of milk proteins, caseins, whey proteins and MFGM. Up to 93% sequence coverage was observed for the most abundant milk proteins (e.g. b-casein); low abundant proteins were also identified successfully with low sequence coverages but good spectral qualities that enabled the unambiguous attribution of proteins' characteristic peptides.

Table II-2 List of the milk proteins identified in the standard samples prepared with skim and whole milk on glass and paper surface and then sampled with two micro-invasive methods (PVC-free eraser and µm grit papers).

Proteins identified	Accession	Samples	Glass slide dipped in whole milk 1X			Glass slide dipped in skim milk 1X			Paper surface in skim milk 1X
			Small PVC eraser	6 µm grit	15 µm grit	Small PVC eraser	6 µm grit	15 µm grit	Small PVC eraser
			A	B	C	J	K	L	BB
Alpha S1 casein	P02662	Casein proteins	89	72	76	46	68	77	80
Alpha S2 casein	P02663		59	37	59	11	35	37	57
Beta casein	P02666		92	92	93	43	91	84	73
Kappa casein	P02668		58	24	71	24	28	31	66
Beta-lactoglobulin	P02754	Whey proteins	92	70	92	46	51	-	-
Alpha-lactalbumin	P00711		37	20	50	-	-	11	18
Serum albumin	P02769		14	9	34	27	7	2	26
Lactotransferrin	P24627		7	8	30	3	10	6	25
Serotransferrin	Q29443		2	-	3	-	-	-	-
Poly-immunoglobulin	P81265		20	8	34	2	10	11	20
Ig-like domain-containing protein	F1MLW7		30	22	46	17	28	12	19
Lactoperoxidase	P80025		-	-	11	-	-	-	4
Lactadherin	Q95114	MFGM proteins	26	4	26	-	9	-	23
Xanthine dehydrogenase /oxidase	P80457		5	5	18	-	8	6	5
Butyrophilin subfamily 1, member A1	P18892		36	23	37	5	12	17	13
Peripilin	A6QLL0		-	-	20	-	-	-	-

6.4.2 Gainsborough's drawing samples

The same methodologies were applied to the study of the historic samples. Despite the low amount of samples investigated, remarkably high protein sequence coverages were obtained, providing the unambiguous characterisation of several milk proteins in all fifteen samples analysed collected from the seven drawings. The results are resumed in Table II-3.

As a representative example, the data achieved for sample S1C2 from *"Landscape with Horse and Cart Descending a Hill"* drawing (2014.32) will be detailed and discussed. All four caseins were identified with a significant protein coverage, precisely 72% for α_{S1}-CN (20 peptides); 51% for α_{S2}-CN (24 peptides); 30% for β-CN (11 peptides) and 71% for κ-CN (24 peptides). Similarly to reference standards, spectra show excellent signal to noise ratios and clearly defined MS2 patterns with an almost full y and b ions series, providing a precise peptide characterisation. Figure II-12-A shows, for instance, the MS2 of the doubly-charged ion m/z 999.0391 characteristics of ^{90}SPAQILQwQVLSnTVPAK^{107} peptide of kappa casein (Δm= 0.4 ppm). The sequence attribution was achieved by y_1^+ to y_{13}^+ and b_2^+ to b_7^+, b_9^+ fragments ions, which allowed the localization of an asparagine deamidation on N103 and oxidation on the tryptophan residue W98.

Several of the major whey proteins were also observed with good sequence coverage. In the sample S1C2 (2014.32), the whey proteins detected were: β-lactoglobulin (56% 18 peptides), bovine serum albumin (11% with 9 peptides) and lactotransferrin (7% with 5 peptides), and immunoglobulins (6% with 4 peptides). The lack or low detection of α-lactalbumin in all samples was unexpected, considering that it is the second most abundant whey protein accounting for 20% of the total content [121]. Nonetheless, this protein was already found in the reference samples with lower sequence coverage (between 30-40%) compared to what was expected based on the scores obtained for other milk proteins identified. The ability of this metal-binding protein to establish interactions with other organic compounds such as lipids is well established [122], although the reasons for its absence compared to other milk protein with similar physicochemical properties remains unclear and in need of further investigations [123, 124].

The observation of some peptides of MFGM proteins can be pointed out; for example, lactadherin was identified in almost all samples with sequence coverage ≤ 15%. In the sample S1C2 (2014.32), despite the relatively low coverage value of 13% (compared to its molecular weight of about 50%), the optimal MS2 spectral quality led to the unambiguous attribution of 5 peptides.

In Figure II-12-B is shown the sequence of y^+ fragments (y_1^+,y_2^+ and y_5^+- y_{12}^+) and b^+ (b_2^+ -b_3^+) that provided the attribution of the doubly-charged ion at m/z 854.9502 to the lactadherin peptide ^{233}INLFDTPLETQYVR^{246} with Δm= 1.5 ppm.

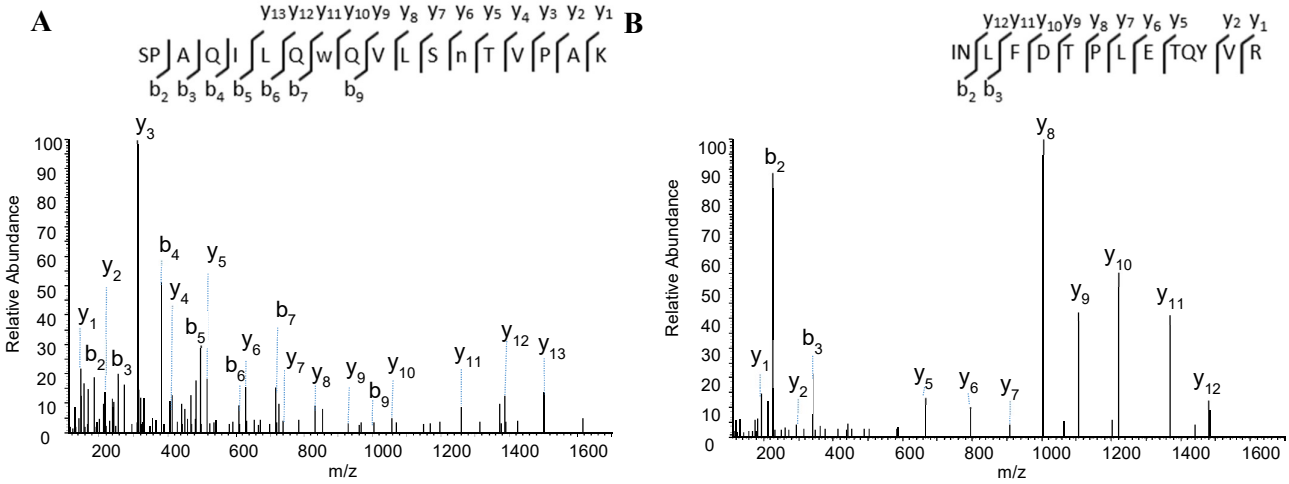


Figure II-12 MS/MS spectra from the sample S1C2 (2014.32): **(A)** of the di-charged charged ion at m/z 999.0391 (Δm=1.4 ppm) presenting y and b fragments of kappa casein 90-107 peptide. The peptide sequence presents oxidation on tryptophan W98 deamidation on asparagine N103; **(B)** of the di-charged charged ion at m/z 854.9502 (Δm=1.5 ppm) presenting y and b fragments of the lactadherin peptide ^{233}INLFDTPLETQYVR^{246}.

Few peptides of xanthine dehydrogenase/oxidase and butyrophilin were also observed in two samples (2 peptides) and five samples (2 peptides). MFGM proteins in these samples are complicated to interpret if related to Gainsborough's 1773 letter that refers explicitly to skim milk. Thanks to industrialisation and process advancements, nowadays skim milk is undoubtedly more efficiently defatted than milk in 18th century England; therefore, the ancient skim milk might have preserved a higher amount of MFGM than current skim milk [125]. Nevertheless, milk treatment, including heat treatment, centrifugation, and cooling, can also cause the discharge of specific membrane-associated proteins from MFGM into the aqueous phase, such as skim milk [126]. Finally, MFGM size can cause differences in protein composition and extent [109, 127-129]. Further investigations implementing milk mock-ups are currently ongoing to better understand the protein composition in the 18th-century skimmed milk.

Table II-3 List of the main milk proteins identified in the Gainsborough drawings for selected locations. Values presented in the table are protein sequence coverages expressed as %, and the obtained -10logP protein score is above 40 for all the proteins listed.

Proteins identified	Accession	Drawings	2005.82		2017.89		2014.32		III, 61		III, 62		III, 63a		III, 63		
		Samples	B	S1C2	S2B1	S2C2	S1B2	S1C2	S2B2	S2C2	S1C2	S3B2	S1C	S4B	S3B2	S2C2	S1A
		Sampling method	6-µm paper	PVC-free eraser	6-µm paper	PVC-free eraser	6-µm paper	PVC-free eraser	6-µm paper	PVC-free eraser	PVC-free eraser	6-µm paper	PVC-free eraser	6-µm paper	6-µm paper	PVC-free eraser	15-µm paper
Alpha S1 casein	P02662	Casein proteins	64	67	50	71	63	72	43	31	39	63	71	67	28	41	36
Alpha S2 casein	P02663		57	48	45	48	53	51	49	36	50	49	44	49	26	42	32
Beta casein	P02666		25	25	31	34	25	30	25	6	31	35	29	29	31	56	31
Kappa casein	P02668		67	67	64	69	70	71	69	33	38	62	65	71	15	30	19
Beta-lactoglobulin	P02754	Whey proteins*	60	50	60	56	56	56	55	39	43	53	50	65	18	40	29
Serum albumin	P02769		11	25	14	13	12	11	12	6	5	12	11	13	---	---	4
Lactotransferrin	P24627		10	9	16	---	9	11	10	3	---	13	6	12	---	---	---
Lactoferrin	B9VPZ5		13	8	12	16	6	7	10	---	---	11	6	9	---	---	---
Serotransferrin	Q29443		7	3	---	---	11	---	---	---	---	1	---	---	---	---	---
Poly-immunoglobulin	P81265		3	8	4	7	5	6	5	---	2	5	6	7	---	---	---
Lactadherin	Q95114	MFGM proteins	14	10	12	12	10	13	15	---	2	13	13	13	---	4	2
Xanthine dehydrogenase /oxidase	P80457		---	---	1	---	---	---	---	---	---	---	---	---	---	1	---
Butyrophilin subfamily 1, member A1	P18892		---	---	4	4	---	---	---	---	2	---	---	---	---	2	2

**Alpha-lactalbumin (accession number P00711) is not referenced in this table due to absence/very poor detection in the studied sample

Numerous protein modifications were detected in milk protein sequences, such as deamidations, oxidations and phosphorylations. However, no lactosylations (a non-enzymatic modification principally resulting from milk heating treatments) were pointed out in the investigated samples [130-132]. To better highlight the absence of lactosylation modifications, identical sequences of α_{S1}-CN were compared with reference samples of heated milk. In the historical samples, the peptide ^{121}VPQLEIVPnsAEER^{134} was detected with a deamidation on N129 and phosphorylation on s S130 (Figure II-13-A). The peptidic sequence was also observed with two miscleavages at K118 and K121 (^{118}KYKVPQ LEIVPNsAEER^{134}), presenting the phosphorylation on serine but the lysine unmodified (Figure II-13-B).

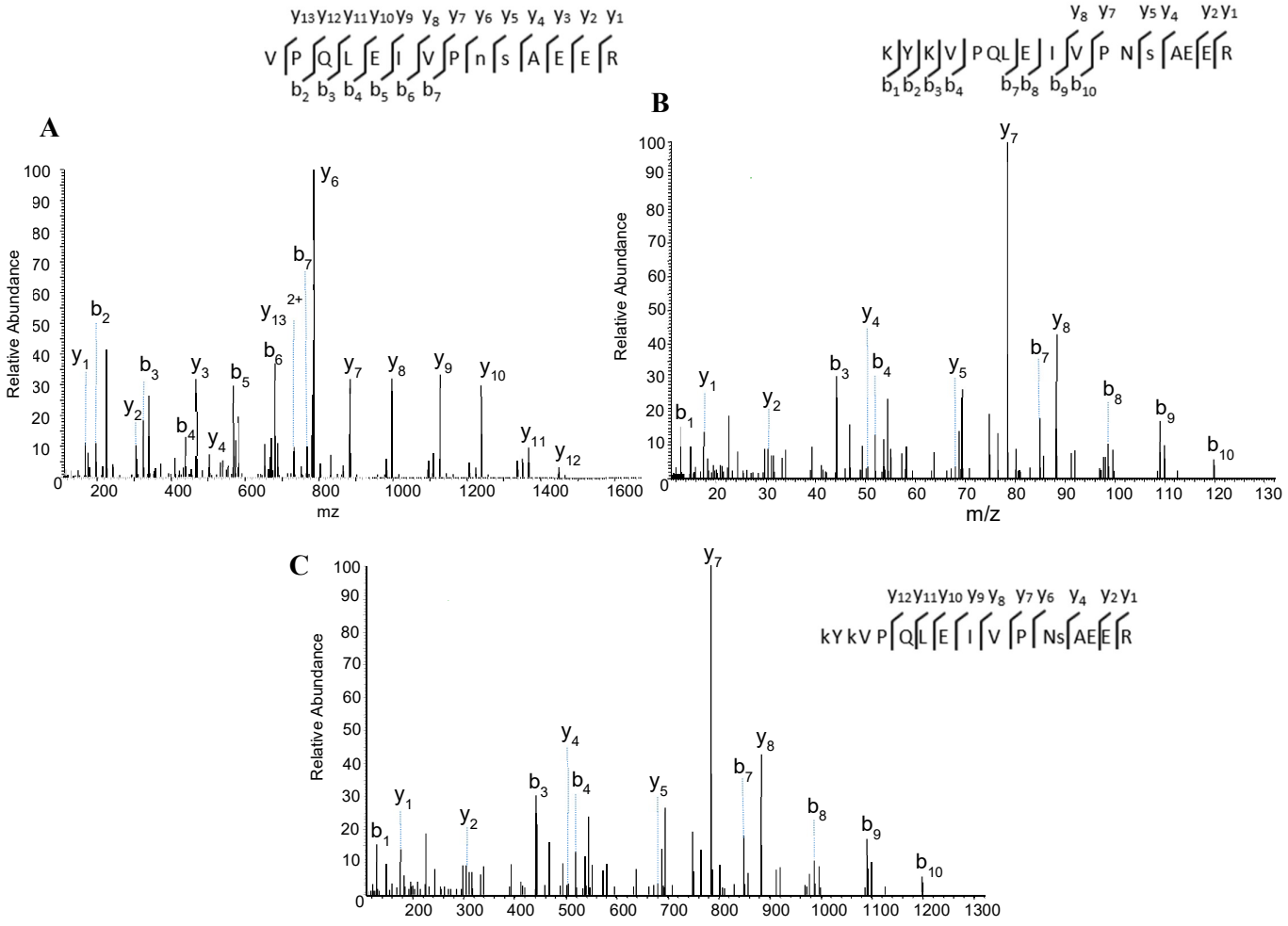


Figure II-13 MS/MS spectra of peptides identified in sample S2C2 from Coastal Scene with Shipping, Figures, and Cows (ca. 1780; 2017.89). **(A)** The doubly charged ion at m/z 831.3934 (Δm=0.8 ppm) enabled the identification of the 121-134 peptide of alpha S1 casein, which includes one phosphorylation (S130) and one deamidation (N129); **(B)** The triply charged ion at m/z 694.0209 (Δm=0.5 ppm) enabled the identification of the 118-134 peptide of alpha S1 casein, which includes one phosphorylation (S130); **(C)** MS/MS spectrum from the reference sample of modern milk subject to heating. The spectrum of the triply charged ion at m/z 910.0919 (Δm=0.9 ppm) led to the identification of the 118–134 peptide of alpha S1 casein, which includes two lactosylations (K118 and K120) and one phosphorylation (S130).

Contrarily, in the heated reference samples, the peptide ^{118}KYKVPQLEIVPNsAEER^{134} was detected with phosphorylation on S130 and lactosylations in two lysine residues K128 and K120 (Figure II-13-C).

The identification of several peptide markers led to the assumption of an application of milk from a Bovinae subfamily (*Bos taurus, Bos mutus, Bos indicus, Bos grunniens, Bison bison bison*). The peptide ^{90}SPAqILQwQVLSNTVPAK^{107} of κ-CN, depicted in Figure II-12–A, for example, is unique for this subfamily; other discriminant peptides from α_{S1}-CN, α_{S2}-CN, κ-CN, and β-lactoglobulin are illustrated in Table II-4.

According to the literature references on Gainsborough, milk-based fixative was most likely derived from *Bos taurus* (domesticated cattle). The other species belonging to the Bovinae subfamily seemed improbable due to his geographical context: *Bos mutus* (wild yak), *Bos indicus* (domestic cattle native to South Asia), *Bos grunniens* (domestic yak), or *Bison bison bison* (American bison). The discovery that Gainsborough applied milk from cattle provenance rather than from other animals, like goat or sheep, helped to get an essential insight into the urban context, specifically in the cities of Bath (1759-1774) and London (1774-1788), where the artist worked on his drawings.

The results achieved through the bottom-up proteomics have significantly contributed to understanding the complexity of Gainsborough's materials and innovative techniques. Several milk proteins specific to the Bovinae subfamily were characterised in all samples investigated, independently on the sample location (front or back, painted figure or paint background) and the drawing date. The application of milk fixative was most likely implemented as early as 1750 for an extended period, until at least 1780-1785. Complementary analytical data achieved from different techniques led to the assumption that milk sizing was applied in drawings having multiple layers (i.e. III, 61; III, 62; III, 63a) as well as purely graphite works (2005.82) and simple watercolours made with lead whitewash (2014.32). Therefore, contrary to what was initially assumed by the experts, regardless of the white pigment (calcite or lead white) or the medium (graphite, chalk, watercolour, or oil paint), milk was implemented. These findings have opened numerous questions on the extent of Gainsborough's use of milk fixative in his drawings, which might prompt further analyses. Furthermore, this research would also encourage new studies to probe the use of casein-based fixatives in other drawings of artists such as Degas and Van Gogh. The results obtained from these analyses have been discussed in the article "Mixing, dipping, and fixing: the experimental drawing techniques of Thomas Gainsborough" [133].

Table II-4 List of the peptides indicative of the Bovinae subfamily. Lower case letters in peptide sequences indicate the following modifications: n = asparagine deamidation; m = methionine oxidation; s = serine phosphorylation; c = cysteine carbamidomethylation (induced by sample treatment); w = tryptophan oxidation; q = N-terminal glutamate conversion to pyroglutamate.

Proteins	Accession	Peptides	Positions	m/z (exp)	z	ppm	-10lgP	RT	Species
Alpha S1 casein	P02662	HPIKHQGLPQEVLNENLLR	19-37	559.5649	4	1.0	50.55	99.7	*Bos taurus, Bos mutus, Bos indicus, Bos grunniens, Bison bison bison*
		YYVPLGTQYTDAPSFSDIPNPIGSENSEK	180-208	1063.836	3	1.1	47.31	184.3	*Bos taurus*
		YKVPQLEIVPNsAEER	119-134	651.3231	3	1.2	42.45	128.3	*Bos taurus, Bos mutus, Bos indicus, Bos grunniens, Bison bison bison*
Alpha S2 casein	P02663	nAVPITPTLNR	130-140	598.8367	2	2.4	36.27	111.8	*Bos taurus, Bos indicus x Bos taurus, Bubalus bubalis, Bos mutus, Bos grunniens, Bison bison bison*
		AmKPWIQPK	204-212	557.808	2	0.9	31.44	46.2	*Bos taurus, Bos indicus x Bos taurus, Bos mutus, Bison bison bison*
		NmAINPsKENLcSTFcK	40-56	709.2982	3	2.2	24.82	88.4	*Bos taurus, Bos indicus x Bos taurus, Bos grunniens, Bison bison bison*
		TKVIPYVR	213-220	488.3035	2	1.2	22.53	58.3	*Bos taurus, Bubalus bubalis, Bos mutus, Bos grunniens, Bison bison bison*
Kappa casein	P02668	SPAQILQwQVLSNTVPAK	90-107	998.5488	2	2.1	64.79	161.5	*Bos taurus, Bos mutus, Bos indicus, Bos grunniens, Bison bison bison, Bos indicus x Bos taurus, Bos gaurus,Bison bison athabascae,Boselaphus tragocamelus*
		qEQNQEQPIRcEKDER	22-37	690.6498	3	2.2	39.68	39.9	*Bos taurus, Bos mutus, bos grunniens, Bos indicus, Bison bison bison*
		QWQVLSNTVPAK	96-107	685.8767	2	2.6	35.62	107.3	*Bos taurus, Bos mutus, Bos indicus, Bison bison bison, Bos grunniens, Bos indicus x Bos taurus, Bos frontalis, Bison bonasus, Boselaphus tragocamelus, Tetracerus quadricornis, Bos javanicus, Bison bison athabascae, Boselaphus tragocamelus*
		QVLSnTVPAK	98-107	529.2989	2	1.6	25.62	58.0	*Bos taurus, Bos mutus, Bos indicus, Bison bison bison, Bos grunniens, Bos indicus x Bos taurus, Bos frontalis, Bison bonasus, Boselaphus tragocamelus, Tetracerus quadricornis, Bos javanicus, Bison bison athabascae, Boselaphus tragocamelus*
Beta-lactoglobulin	P0275	TPEVDDEALEKFDK	141-154	818.3918	2	1.0	43.07	99.8	*Bos taurus, Bos mutus, Bubalus bubalis, Bos indicus, Bison bison bison*
		LIVTQTmK	17-24	475.2734	2	0.8	22.55	42.9	*Bos taurus, Bos mutus, Bos indicus, Bison bison bison*

7. Conclusions

The bottom-up strategy implemented in the laboratory for the investigation of cultural heritage samples was further optimised, guaranteeing the successful identification of proteinaceous content in two different artworks, also at trace level: (i) wall paintings from the Nubian region (Middle Nile Valley) dated between the 6th and 14th century AD and (ii) landscape drawings created by Thomas Gainsborough between 1750-1780. The applied protocol was proven successful in protein identification, providing an excellent protein coverage with a high number of peptides identified, regardless of the heterogeneity and degradation of the artwork and the low amount of sample collected. Especially in the Gainsborough study, the methodology was successfully performed on micro-samples collected with a minimally invasive technique. In both studies, protein detection unveiled new aspects of the artworks. In Nubian wall paintings, the results revealed the presence of proteinaceous compounds in artworks, often exclusively inorganic, while the investigation on Gainsborough drawings led to the disclosure of the artist secret recipe comprised of milk-based sizing.

The identification of species-specific peptides, together with the accurate study of amino acid substitutions, provided a more comprehensive vision of the proteinaceous compounds regarding other analytical techniques, offering better insights into the manufacturing context. Finally, the detection of several post-translational modifications (PTMs), principally deamidation and oxidations, enabled us to go beyond the protein identification and better understand the enhanced state of degradation of all samples caused by the ageing as well as the storage conditions.

Further improvements can always be pursued, for instance, by increasing the sensitivity of the instrument, limiting the interferences of contaminants with targeted analyses [134], and through the sequencing of more ancient species not yet in the database [25]. Increasing attention is also addressed to the investigation of modifications, encompassing the structural changes (as in the case of cross-linking studies, better detailed in chapters IV and V) to gain an insight on the sample ageing, degradation processes and the effect of protein environment (either internal like pigments or external like pollution or restorations treatments).

Reference

1. Dallongeville, S., N. Garnier, C. Rolando ,C. Tokarski, *Proteins in art, archaeology, and paleontology: from detection to identification.* Chemical reviews, 2016. **116**(1): p. 2-79.
2. Wolters, D.A., M.P. Washburn ,J.R. Yates, *An automated multidimensional protein identification technology for shotgun proteomics.* Analytical chemistry, 2001. **73**(23): p. 5683-5690.
3. Yates III, J.R., *Mass spectral analysis in proteomics.* Annu. Rev. Biophys. Biomol. Struct., 2004. **33**: p. 297-316.
4. Zhang, Y., B.R. Fonslow, B. Shan, M.-C. Baek ,J.R. Yates III, *Protein analysis by shotgun/bottom-up proteomics.* Chemical reviews, 2013. **113**(4): p. 2343-2394.
5. Pappin, D.J., P. Hojrup ,A.J. Bleasby, *Rapid identification of proteins by peptide-mass fingerprinting.* Current biology, 1993. **3**(6): p. 327-332.
6. Biemann, K. ,S.A. Martin, *Mass spectrometric determination of the amino acid sequence of peptides and proteins.* Mass Spectrometry Reviews, 1987. **6**(1): p. 1-75.
7. Valcu, C.-M. ,B. Kempenaers, *Proteomics in behavioral ecology.* Behavioral Ecology, 2015. **26**(1): p. 1-15.
8. Tokarski, C., C. Cren-Olive, C. Rolando ,E. Martin, *Protein studies in cultural heritage.* Molecular biology and cultural heritage: Swets & Zeitlinger, 2003: p. 119-130.
9. Tokarski, C., C. Cren-Olive, E. Martin ,C. Rolando. *Proteomics methodology for protein identification in ancient art and archeological samples*. in *ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY*. 2003. AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA.
10. Tokarski, C., C. Cren-Olive, E. Martin ,C. Rolando, *Proteomics approach for binding media studies in art painting.* Non-Destructive Testing and Microanalysis for the Diagnostics and Conservation of the Cultural and Environmental Heritage, Antwerp, Belgium, 2002.
11. Buckley, M., M. Collins, J. Thomas-Oates ,J.C. Wilson, *Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry, 2009. **23**(23): p. 3843-3854.
12. Collins, M., M. Buckley, H.H. Grundy, J. Thomas-Oates, J. Wilsona ,N. van Doorna, *ZooMS: the collagen barcode and fingerprints.* group (Saarland University, Germany), 2010. **1**: p. 2.
13. Stewart, J.R., R.B. Allen, A.K. Jones, K.E. Penkman ,M.J. Collins, *ZooMS: making eggshell visible in the archaeological record.* Journal of Archaeological Science, 2013. **40**(4): p. 1797-1804.
14. Buckley, M., N.D. Melton ,J. Montgomery, *Proteomics analysis of ancient food vessel stitching reveals> 4000-year-old milk protein.* Rapid Communications in Mass Spectrometry, 2013. **27**(4): p. 531-538.
15. Buckley, M., S. Fraser, J. Herman, N.D. Melton, J. Mulville ,A.H. Pálsdóttir, *Species identification of archaeological marine mammals using collagen fingerprinting.* Journal of Archaeological Science, 2014. **41**: p. 631-641.
16. Welker, F., M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo ,A. Kramarz, *Ancient proteins resolve the evolutionary history of Darwin's South American ungulates.* Nature, 2015. **522**(7554): p. 81-84.

17. Welker, F., M. Hajdinjak, S. Talamo, K. Jaouen, M. Dannemann, F. David, M. Julien, M. Meyer, J. Kelso ,I. Barnes, *Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne.* Proceedings of the National Academy of Sciences, 2016. **113**(40): p. 11162-11167.
18. Welker, F., M.A. Soressi, M. Roussel, I. van Riemsdijk, J.-J. Hublin ,M.J. Collins, *Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen.* STAR: Science & Technology of Archaeological Research, 2017. **3**(1): p. 15-27.
19. Fiddyment, S., B. Holsinger, C. Ruzzier, A. Devine, A. Binois, U. Albarella, R. Fischer, E. Nichols, A. Curtis ,E. Cheese, *Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting.* Proceedings of the National Academy of Sciences, 2015. **112**(49): p. 15066-15071.
20. Brandt, L.Ø., K. Haase ,M.J. Collins, *Species identification using ZooMS, with reference to the exploitation of animal resources in the medieval town of Odense.* Danish Journal of Archaeology, 2018. **7**(2): p. 139-153.
21. McGrath, K., K. Rowsell, C.G. St-Pierre, A. Tedder, G. Foody, C. Roberts, C. Speller ,M. Collins, *Identifying archaeological bone via non-destructive ZooMS and the materiality of symbolic expression: examples from Iroquoian bone points.* Scientific reports, 2019. **9**(1): p. 1-10.
22. Tokarski, C., E. Martin, C. Rolando ,C. Cren-Olivé, *Identification of proteins in renaissance paintings by proteomics.* Analytical chemistry, 2006. **78**(5): p. 1494-1502.
23. Tokarski, C., E. Martin, C. Rolando ,C. Cren-Olivé, *Protein identification in ancient art paintings.* Biosystems Solutions, 2004. **11**: p. 38-39.
24. Ramsøe, A., V. van Heekeren, P. Ponce, R. Fischer, I. Barnes, C. Speller ,M.J. Collins, *DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes.* Journal of Archaeological Science, 2020. **115**: p. 105080.
25. Carletti, M.S., A.M. Monzon, E. Garcia-Rios, G. Benitez, L. Hirsh, M.S. Fornasari ,G. Parisi, *Revenant: a database of resurrected proteins.* Database, 2020. **2020**.
26. Vinciguerra, R., A. De Chiaro, P. Pucci, G. Marino ,L. Birolo, *Proteomic strategies for cultural heritage: From bones to paintings.* Microchemical Journal, 2016. **126**: p. 341-348.
27. Linn, R., I. Bonaduce, G. Ntasi, L. Birolo, A. Yasur-Landau, E.H. Cline, A. Nevin ,A. Lluveras-Tenorio, *Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings.* Angewandte Chemie, 2018. **130**(40): p. 13441-13444.
28. Leo, G., I. Bonaduce, A. Andreotti, G. Marino, P. Pucci, M.P. Colombini ,L. Birolo, *Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings.* Analytical chemistry, 2011. **83**(6): p. 2056-2064.
29. Orsini, S., A. Yadav, M. Dilillo, L.A. McDonnell ,I. Bonaduce, *Characterization of degraded proteins in paintings using bottom-up proteomic approaches: new strategies for protein digestion and analysis of data.* Analytical chemistry, 2018. **90**(11): p. 6403-6408.
30. Solazzo, C., J. Wilson, J.M. Dyer, S. Clerens, J.E. Plowman, I. Von Holstein, P. Walton Rogers, E.E. Peacock ,M.J. Collins, *Modeling deamidation in sheep α-keratin peptides and application to archeological wool textiles.* Analytical chemistry, 2014. **86**(1): p. 567-575.
31. Artioli, G. ,I. Angelini, *Scientific methods and cultural heritage: an introduction to the application of materials science to archaeometry and conservation science*. 2010: Oxford University Press.
32. Hendy, J., F. Welker, B. Demarchi, C. Speller, C. Warinner ,M.J. Collins, *A guide to ancient protein studies.* Nature ecology & evolution, 2018. **2**(5): p. 791-799.

33. Ebsen, J.A., K. Haase, R. Larsen, D.V.P. Sommer ,L.Ø. Brandt, *Identifying archaeological leather–discussing the potential of grain pattern analysis and zooarchaeology by mass spectrometry (ZooMS) through a case study involving medieval shoe parts from Denmark.* Journal of Cultural Heritage, 2019. **39**: p. 21-31.
34. Kirby, D.P., A. Manick ,R. Newman, *Minimally Invasive Sampling of Surface Coatings for Protein Identification by Peptide Mass Fingerprinting: A Case Study with Photographs.* Journal of the American Institute for Conservation, 2019: p. 1-11.
35. Cicatiello, P., G. Ntasi, M. Rossi, G. Marino, P. Giardina ,L. Birolo, *Minimally Invasive and Portable Method for the Identification of Proteins in Ancient Paintings.* Analytical chemistry, 2018. **90**(17): p. 10128-10133.
36. Calvano, C., E. Rigante, R. Picca, T. Cataldi ,L. Sabbatini, *An easily transferable protocol for in-situ quasi-non-invasive analysis of protein binders in works of art.* Talanta, 2020: p. 120882.
37. Zilberstein, G., U. Maor, E. Baskin, A. D'Amato ,P.G. Righetti, *Unearthing Bulgakov's trace proteome from the Master i Margarita manuscript.* Journal of proteomics, 2017. **152**: p. 102-108.
38. Barberis, E., M. Manfredi, E. Marengo, G. Zilberstein, S. Zilberstein, A. Kossolapov ,P.G. Righetti, *Leonardo's donna nuda unveiled.* Journal of proteomics, 2019. **207**: p. 103450.
39. Demarchi, B., R. Boano, A. Ceron, F. Dal Bello, S.E. Favero-Longo, S. Fiddyment, E.F. Marochetti, G. Mangiapane, M. Mattonai ,C. Pennacini, *Never boring: Non-invasive palaeoproteomics of mummified human skin.* Journal of Archaeological Science, 2020. **119**: p. 105145.
40. He, J., H. Yan ,C. Fan, *Optimization of ultrasound-assisted extraction of protein from egg white using response surface methodology (RSM) and its proteomic study by MALDI-TOF-MS.* RSC Advances, 2014. **4**(80): p. 42608-42616.
41. Chambery, A., A. Di Maro, C. Sanges, V. Severino, M. Tarantino, A. Lamberti, A. Parente ,P. Arcari, *Improved procedure for protein binder analysis in mural painting by LC-ESI/Qq-TOF mass spectrometry: detection of different milk species by casein proteotypic peptides.* Analytical and bioanalytical chemistry, 2009. **395**(7): p. 2281-2291.
42. Barker, A., B. Venables, S.M. Stevens, K.W. Seeley, P. Wang ,S. Wolverton, *An optimized approach for protein residue extraction and identification from ceramics after cooking.* Journal of Archaeological Method and Theory, 2012. **19**(3): p. 407-439.
43. Vinciguerra, R., E. Galano, F. Vallone, G. Greco, A. Vergara, I. Bonaduce, G. Marino, P. Pucci, A. Amoresano ,L. Birolo, *Deglycosylation step to improve the identification of egg proteins in art samples.* Analytical chemistry, 2015. **87**(20): p. 10178-10182.
44. Solazzo, C., W.W. Fitzhugh, C. Rolando ,C. Tokarski, *Identification of protein remains in archaeological potsherds by proteomics.* Analytical Chemistry, 2008. **80**(12): p. 4590-4597.
45. Zhu, Z., P. Tan, J. Yang, H. Ge ,L. Liu, *Mass Spectrometric Identification of Adhesive Utilized in a Tian-tsui Tiara of the mid-Qing Dynasty (1776–1839 CE) in the Collection of the Tang Clan Folk Museum.* Studies in Conservation, 2019. **64**(4): p. 187-192.
46. Kirby, D.P., M. Buckley, E. Promise, S.A. Trauger ,T.R. Holdcraft, *Identification of collagen-based materials in cultural heritage.* Analyst, 2013. **138**(17): p. 4849-4858.
47. Leo, G., L. Cartechini, P. Pucci, A. Sgamellotti, G. Marino ,L. Birolo, *Proteomic strategies for the identification of proteinaceous binders in paintings.* Analytical and bioanalytical chemistry, 2009. **395**(7): p. 2269-2280.
48. Mackie, M., P. Rüther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Peggie, H. Howard, L. Harrison ,L.J. Jensen, *Palaeoproteomic profiling of conservation layers on a 14th century Italian wall painting.* Angewandte Chemie International Edition, 2018. **57**(25): p. 7369-7374.

49. Dallongeville, S., M. Richter, S. Schäfer, M. Kühlenthal, N. Garnier, C. Rolando ,C. Tokarski, *Proteomics applied to the authentication of fish glue: application to a 17 th century artwork sample.* Analyst, 2013. **138**(18): p. 5357-5364.
50. Chen, R., M. Hu, H. Zheng, H. Yang, L. Zhou, Y. Zhou, Z. Peng, Z. Hu ,B. Wang, *Proteomics and Immunology Provide Insight into the Degradation Mechanism of Historic and Artificially Aged Silk.* Analytical Chemistry, 2020. **92**(3): p. 2435-2442.
51. Dallongeville, S., N. Garnier, D.B. Casasola, M. Bonifay, C. Rolando ,C. Tokarski, *Dealing with the identification of protein species in ancient amphorae.* Analytical and bioanalytical chemistry, 2011. **399**(9): p. 3053-3063.
52. Mazurek, J., M. Svoboda, J. Maish, K. Kawahara, S. Fukakusa, T. Nakazawa ,Y. Taniguchi, *Characterization of binding media in Egyptian Romano portraits using enzyme-linked immunosorbant assay and mass spectrometry.* e-Preserv Sci, 2014. **11**: p. 76-83.
53. Grieken, R.v., *Art 2002: 7th international conference on non-destructive testing and microanalysis for the diagnostics and conservation of the cultural and environmental heritage, 2-6 June 2002: proceedings.* 2002.
54. Kirby, D.P., N. Khandekar, J. Arslanoglu ,K. Sutherland. *Protein identification in artworks by peptide mass fingerprinting*. in *Preprints of the ICOM Committee for Conservation 16th Triennial Conference, Lisbon*. 2011.
55. Tsiatsiani, L. ,A.J. Heck, *Proteomics beyond trypsin.* The FEBS journal, 2015. **282**(14): p. 2612-2626.
56. Calvano, C.D., I.D. van der Werf, F. Palmisano ,L. Sabbatini, *Identification of lipid- and protein-based binders in paintings by direct on-plate wet chemistry and matrix-assisted laser desorption ionization mass spectrometry.* Analytical and bioanalytical chemistry, 2015. **407**(3): p. 1015-1022.
57. Calvano, C.D., I.D. van der Werf, L. Sabbatini ,F. Palmisano, *On plate graphite supported sample processing for simultaneous lipid and protein identification by matrix assisted laser desorption ionization mass spectrometry.* Talanta, 2015. **137**: p. 161-166.
58. Betancourt, L.H., A. Sanchez, I. Pla, M. Kuras, Q. Zhou, R. Andersson ,G. Marko-Varga, *Quantitative Assessment of Urea In-Solution Lys-C/Trypsin Digestions Reveals Superior Performance at Room Temperature over Traditional Proteolysis at 37° C.* Journal of proteome research, 2018. **17**(7): p. 2556-2561.
59. Fremout, W., M. Dhaenens, S. Saverwyns, J. Sanyova, P. Vandenabeele, D. Deforce ,L. Moens, *Tryptic peptide analysis of protein binders in works of art by liquid chromatography–tandem mass spectrometry.* Analytica chimica acta, 2010. **658**(2): p. 156-162.
60. Witkowski, B., M. Biesaga ,T. Gierczak, *Proteinaceous binders identification in the works of art using ion-pairing free reversed-phase liquid chromatography coupled with tandem mass spectrometry.* Analytical Methods, 2012. **4**(5): p. 1221-1228.
61. Witkowski, B., A. Duchnowicz, M. Ganeczko, A. Laudy, T. Gierczak ,M. Biesaga, *Identification of proteins, drying oils, waxes and resins in the works of art micro-samples by chromatographic and mass spectrometric techniques.* Journal of separation science, 2018. **41**(3): p. 630-638.
62. Šesták, J., D. Moravcová ,V. Kahle, *Instrument platforms for nano liquid chromatography.* Journal of Chromatography A, 2015. **1421**: p. 2-17.
63. Kim, M.S., J. Zhong, K. Kandasamy, B. Delanghe ,A. Pandey, *Systematic evaluation of alternating CID and ETD fragmentation for phosphorylated peptides.* Proteomics, 2011. **11**(12): p. 2568-2572.
64. Michalski, A., J. Cox ,M. Mann, *More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC–MS/MS.* Journal of proteome research, 2011. **10**(4): p. 1785-1793.

65. Hu, A., W.S. Noble ,A. Wolf-Yadlin, *Technical advances in proteomics: new developments in data-independent acquisition.* F1000Research, 2016. **5**.
66. Egertson, J.D., A. Kuehn, G.E. Merrihew, N.W. Bateman, B.X. MacLean, Y.S. Ting, J.D. Canterbury, D.M. Marsh, M. Kellmann ,V. Zabrouskov, *Multiplexed MS/MS for improved data-independent acquisition.* Nature methods, 2013. **10**(8): p. 744-746.
67. Bilbao, A., E. Varesio, J. Luban, C. Strambio-De-Castillia, G. Hopfgartner, M. Müller ,F. Lisacek, *Processing strategies and software solutions for data-independent acquisition in mass spectrometry.* Proteomics, 2015. **15**(5-6): p. 964-980.
68. MacCoss, M.J., C.C. Wu ,J.R. Yates, *Probability-based validation of protein identifications using a modified SEQUEST algorithm.* Analytical chemistry, 2002. **74**(21): p. 5593-5599.
69. Brosch, M., L. Yu, T. Hubbard ,J. Choudhary, *Accurate and sensitive peptide identification with Mascot Percolator.* Journal of proteome research, 2009. **8**(6): p. 3176-3181.
70. Cox, J., N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen ,M. Mann, *Andromeda: a peptide search engine integrated into the MaxQuant environment.* Journal of proteome research, 2011. **10**(4): p. 1794-1805.
71. Bern, M., Y.J. Kil ,C. Becker, *Byonic: advanced peptide and protein identification software.* Current protocols in bioinformatics, 2012. **40**(1): p. 13.20. 1-13.20. 14.
72. Wang, G., W.W. Wu, Z. Zhang, S. Masilamani ,R.-F. Shen, *Decoy methods for assessing false positives and false discovery rates in shotgun proteomics.* Analytical chemistry, 2009. **81**(1): p. 146-159.
73. Tran, N.H., R. Qiao, L. Xin, X. Chen, C. Liu, X. Zhang, B. Shan, A. Ghodsi ,M. Li, *Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry.* Nature methods, 2019. **16**(1): p. 63-66.
74. Ma, B., K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby ,G. Lajoie, *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.* Rapid communications in mass spectrometry, 2003. **17**(20): p. 2337-2342.
75. Han, X., L. He, L. Xin, B. Shan ,B. Ma, *PeaksPTM: mass spectrometry-based identification of peptides with unspecified modifications.* Journal of proteome research, 2011. **10**(7): p. 2930-2936.
76. Erde, J., R.R.O. Loo ,J.A. Loo, *Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments.* Journal of proteome research, 2014. **13**(4): p. 1885-1895.
77. Sun, S., J.-Y. Zhou, W. Yang ,H. Zhang, *Inhibition of protein carbamylation in urea solution using ammonium-containing buffers.* Analytical biochemistry, 2014. **446**: p. 76-81.
78. Cennini, C., C.D.A. Cennini ,G. De Beer, *The craftsman's handbook*. Vol. 2. 1954: Courier Corporation.
79. Zielinska, D., *Medieval Nubian Wall Paintings: Techniques and Conservation*. 2019: Archetype Publications.
80. Stadelman, W.J., D. Newkirk ,L. Newby, *Egg science and technology*. 2017: CRC Press.
81. Robinson, N.E. ,A.B. Robinson, *Molecular clocks.* Proceedings of the National Academy of Sciences, 2001. **98**(3): p. 944-949.
82. Kerkaert, B., F.d.r. Mestdagh, M. Obando, T. Cucu ,B. De Meulenaer, *Identification of modified lysozyme peptides upon photo-oxidation by LC-TOF-MS.* Journal of agricultural and food chemistry, 2013. **61**(51): p. 12727-12736.
83. Wilson, J., N.L. van Doorn ,M.J. Collins, *Assessing the extent of bone degradation using glutamine deamidation in collagen.* Analytical chemistry, 2012. **84**(21): p. 9041-9048.

84. Solazzo, C., S. Clerens, J.E. Plowman, J. Wilson, E.E. Peacock ,J.M. Dyer, *Application of redox proteomics to the study of oxidative degradation products in archaeological wool.* Journal of Cultural Heritage, 2015. **16**(6): p. 896-903.
85. Davies, M.J., *Protein oxidation and peroxidation.* Biochem J, 2016. **473**(7): p. 805-25.
86. Davies, M.J., *Reactive species formed on proteins exposed to singlet oxygen.* Photochemical & Photobiological Sciences, 2004. **3**(1): p. 17-25.
87. Pattison, D.I., A.S. Rahmanto ,M.J. Davies, *Photo-oxidation of proteins.* Photochemical & Photobiological Sciences, 2012. **11**(1): p. 38-53.
88. Elert, K. ,C. Cardell, *Weathering behavior of cinnabar-based tempera paints upon natural and accelerated aging.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019. **216**: p. 236-248.
89. Duce, C., L. Ghezzi, M. Onor, I. Bonaduce, M.P. Colombini, M.R. Tine ,E. Bramanti, *Physico-chemical characterization of protein–pigment interactions in tempera paint reconstructions: casein/cinnabar and albumin/cinnabar.* Analytical and bioanalytical chemistry, 2012. **402**(6): p. 2183-2193.
90. Jakobielski, S. ,P.O. Scholz, *Dongola-Studien: 35 Jahre polnischer Forschungen im Zentrum des makuritischen Reiches*. 2001: Zaś Pan.
91. Palet, A. ,E. Porta. *Chemical analysis of pigments and media in the mural paintings of the tomb of Nefertari*. in *VIII congress of conservation of cultural property. Valencia*. 1990.
92. Newman, R. ,S.M. Halpine, *The binding media of ancient Egyptian painting*, in *Colour and painting in ancient Egypt*. 2001. p. 22-32.
93. Imbrogno, J., A. Nayak ,G. Belfort, *Egg white varnishes on ancient paintings: A molecular connection to amyloid proteins.* Angewandte Chemie, 2014. **126**(27): p. 7134-7137.
94. Woudhuysen-Keller, R. ,P. Woudhuysen-Keller, *The History of eggwhite varnishes.* Hamilton Kerr Institute bulletin, 1994(2): p. 90-141.
95. Brodsky, B. ,J.A. Ramshaw, *The collagen triple-helix structure.* Matrix Biology, 1997. **15**(8-9): p. 545-554.
96. Lloyd, A.B., *A companion to ancient Egypt*. Vol. 52. 2010: John Wiley & Sons.
97. Chiavari, G., D. Fabbri, G. Galletti ,R. Mazzeo, *Use of analytical pyrolysis to characterize Egyptian painting layers.* Chromatographia, 1995. **40**(9-10): p. 594-600.
98. Brøns, C., K.L. Rasmussen, M.M. Di Crescenzo, R. Stacey ,A. Lluveras-Tenorio, *Painting the Palace of Apries I: ancient binding media and coatings of the reliefs from the Palace of Apries, Lower Egypt.* Heritage Science, 2018. **6**(1): p. 6.
99. LeFur, D., *La conservation des peintures murales des temples de Karnak*. 1994: Ed. Recherche sur les civilisations.
100. Shedrinsky, A.M., T.P. Wampler, N. Indictor ,N.S. Baer, *Application of analytical pyrolysis to problems in art and archaeology: a review.* Journal of Analytical and Applied Pyrolysis, 1989. **15**: p. 393-412.
101. Scott, D.A., S. Warmlander, J. Mazurek ,S. Quirke, *Examination of some pigments, grounds and media from Egyptian cartonnage fragments in the Petrie Museum, University College London.* Journal of Archaeological Science, 2009. **36**(3): p. 923-932.
102. Derow, J.P., *Gainsborough's Varnished Watercolor Technique.* Master Drawings, 1988: p. 259-308.
103. Doucette, C. ,C. Wilda, *A Technical Examination and Art Historical Investigation of The Woodman, after Thomas Gainsborough.*
104. Hayes, J., *The Letters of Thomas Gainsborough*. 2001.
105. Hayes, J., *Gainsborough drawings: a supplement to the catalogue raisonné.* Master Drawings, 1983: p. 367-391.

106. Weir, J., *Eggs, milk and fish: a practical investigation of artists' use of fixatives'.* The Book and Paper Group Annual. **26**: p. 134.
107. Singh, H. ,J. Flanagan, *Milk proteins.* FOOD SCIENCE AND TECHNOLOGY-NEW YORK-MARCEL DEKKER-, 2006. **149**(1): p. 26.
108. Farrell, H., *Milk proteins| casein nomenclature, structure, and association.* 2011.
109. Vincent, D., V. Ezernieks, A. Elkins, N. Nguyen, P.J. Moate, B.G. Cocks ,S. Rochfort, *Milk bottom-up proteomics: Method optimization.* Frontiers in genetics, 2016. **6**: p. 360.
110. Ng-Kwai-Hang, K., *Milk Proteins| Heterogeneity, Fractionation, and Isolation.* 2011.
111. Farrell Jr, H., R. Jimenez-Flores, G. Bleck, E. Brown, J. Butler, L. Creamer, C. Hicks, C. Hollar, K. Ng-Kwai-Hang ,H. Swaisgood, *Nomenclature of the proteins of cows' milk—Sixth revision.* Journal of dairy science, 2004. **87**(6): p. 1641-1674.
112. Bijl, E., H. van Valenberg, T. Huppertz, A. van Hooijdonk ,H. Bovenhuis, *Phosphorylation of αS1-casein is regulated by different genes.* Journal of dairy science, 2014. **97**(11): p. 7240-7246.
113. Caroli, A., S. Chessa ,G. Erhardt, *Milk protein genetic variation in cattle: Impact on animal breeding and human nutrition.* Journal of Dairy Science, 2009. **92**: p. 5335-5352.
114. Kelly, A., F. O'Flaherty ,P. Fox, *Indigenous proteolytic enzymes in milk: A brief overview of the present state of knowledge.* International dairy journal, 2006. **16**(6): p. 563-572.
115. Law, J. ,A. Haandrikman, *Proteolytic enzymes of lactic acid bacteria.* International Dairy Journal, 1997. **7**(1): p. 1-11.
116. Boland, M. ,H. Singh, *Milk proteins: from expression to food*. 2019: Academic Press.
117. Mather, I., *Milk Lipids| Milk Fat Globule Membrane.* 2011.
118. Lordan, R., A. Tsoupras ,I. Zabetakis, *Phospholipids of animal and marine origin: Structure, function, and anti-inflammatory properties.* Molecules, 2017. **22**(11): p. 1964.
119. Sztalryd, C. ,D.L. Brasaemle, *The perilipin family of lipid droplet proteins: Gatekeepers of intracellular lipolysis.* Biochimica et biophysica acta (bba)-molecular and cell biology of lipids, 2017. **1862**(10): p. 1221-1232.
120. Mather, I.H., *A review and proposed nomenclature for major proteins of the milk-fat globule membrane.* Journal of dairy science, 2000. **83**(2): p. 203-247.
121. Layman, D.K., B. Lönnerdal ,J.D. Fernstrom, *Applications for α-lactalbumin in human nutrition.* Nutrition reviews, 2018. **76**(6): p. 444-460.
122. Barbana, C. ,M.D. Pérez, *Interaction of α-lactalbumin with lipids and possible implications for its emulsifying properties–A review.* International dairy journal, 2011. **21**(10): p. 727-741.
123. Vegarud, G.E., T. Langsrud ,C. Svenning, *Mineral-binding milk proteins and peptides; occurrence, biochemical and technological characteristics.* British Journal of Nutrition, 2000. **84**(S1): p. 91-98.
124. Singh, H., *Milk protein products| functional properties of milk proteins.* 2011.
125. Atkins, P.J., *The pasteurization of England: the science, culture and health implications of milk processing, 1900-1950.* Food, science, policy and regulation in the twentieth century: international and comparative perspectives London: Routledge, 2000: p. 37-51.
126. Park, Y.W. ,G.F. Haenlein, *Milk and dairy products in human nutrition: production, composition and health*. 2013: John Wiley & Sons.
127. Fox, P., *The major constituents of milk*. 2003: CRC Press and Woodhead Publishing Limited. Boca Raton.
128. Zheng, H., R. Jiménez-Flores ,D.W. Everett, *Bovine milk fat globule membrane proteins are affected by centrifugal washing processes.* Journal of agricultural and food chemistry, 2013. **61**(35): p. 8403-8411.

129. Haddadian, Z., G.T. Eyres, A. Carne, D.W. Everett ,P. Bremer, *Impact of different milk fat globule membrane preparations on protein composition, xanthine oxidase activity, and redox potential.* International Dairy Journal, 2017. **64**: p. 14-21.
130. Deeth, H.C. ,M.J. Lewis, *High temperature processing of milk and milk products*. 2017: John Wiley & Sons.
131. Gasparini, A., S. Buhler, A. Faccini, S. Sforza ,T. Tedeschi, *Thermally-Induced Lactosylation of Whey Proteins: Identification and Synthesis of Lactosylated β-lactoglobulin Epitope.* Molecules, 2020. **25**(6): p. 1294.
132. Czerwenka, C., I. Maier, F. Pittner ,W. Lindner, *Investigation of the lactosylation of whey proteins by liquid chromatography− mass spectrometry.* Journal of agricultural and food chemistry, 2006. **54**(23): p. 8874-8882.
133. Pozzi, F., J. Arslanoglu, F. Galluzzi, C. Tokarski ,R. Snyder, *Mixing, dipping, and fixing: the experimental drawing techniques of Thomas Gainsborough.* Heritage Science, 2020. **8**(1): p. 1-14.
134. Vinciguerra, R., A. Illiano, A. De Chiaro, A. Carpentieri, A. Lluveras-Tenorio, I. Bonaduce, G. Marino, P. Pucci, A. Amoresano ,L. Birolo, *Identification of proteinaceous binders in paintings: A targeted proteomic approach for cultural heritage.* Microchemical Journal, 2019. **144**: p. 319-328.

Chapter III Top-down proteomic analysis of historical and artistic samples

Abstract

This chapter describes a new methodology based on top-down proteomics to identify proteoforms at trace level from historical artworks. The presented protocol, established with commercially available caseins standard (alpha and beta caseins), consists of a filter-based sample preparation optimised to remove interfering substances. Analyses were performed on-line with nanoLC using both MS mode and MS/MS analysis using EThcD fragmentation. The strategy was also evaluated on older casein-based glue samples (dated 1900-1950), leading to the detection of various truncated casein proteoforms characterised by modifications characteristic of ageing and degradation processes, such as oxidation and deamidation. Finally, the protocol was applied to a historic sample from Gainsborough's drawing "Landscape with Horse and Cart Descending a Hill", collected with a minimally-invasive technique. Despite the incredibly minute amount of sample, several caseins proteoforms, both truncated and highly modified, were identified. The results suggested the application of milk as pigment fixative, supporting the artist writings and the bottom-up data on the same drawing. Besides, new insights were achieved about the biological specie of milk and the sample degradation, proving the complementarity of bottom-up and top-down approaches and showing that TDP might provide a more in-depth investigation of proteins PTMs in artistic artefacts.

1. Top-down approach

The bottom-up approach might face some limitations, such as the lack of detection of a variable extent of peptides with a consequent achievement of a limited protein sequence coverage, the loss of labile post-translational modifications (PTMs), or conversely, the formation of artificial alterations during the bottom-up sample preparation, such as the digestion step [1-4].

Top-down proteomics (TDP) has arisen as a valid alternative methodology that, through the elimination of the protein digestion, allows the identification and characterisation of intact proteins in their endogenous form by direct fragmentation. TDP enables the achievement of the protein complete sequence coverage and the preservation and determination of valuable information about the molecular networks, alternative splice variants, isoforms, proteolytic processing and PTMs [5, 6] (Figure III-1).

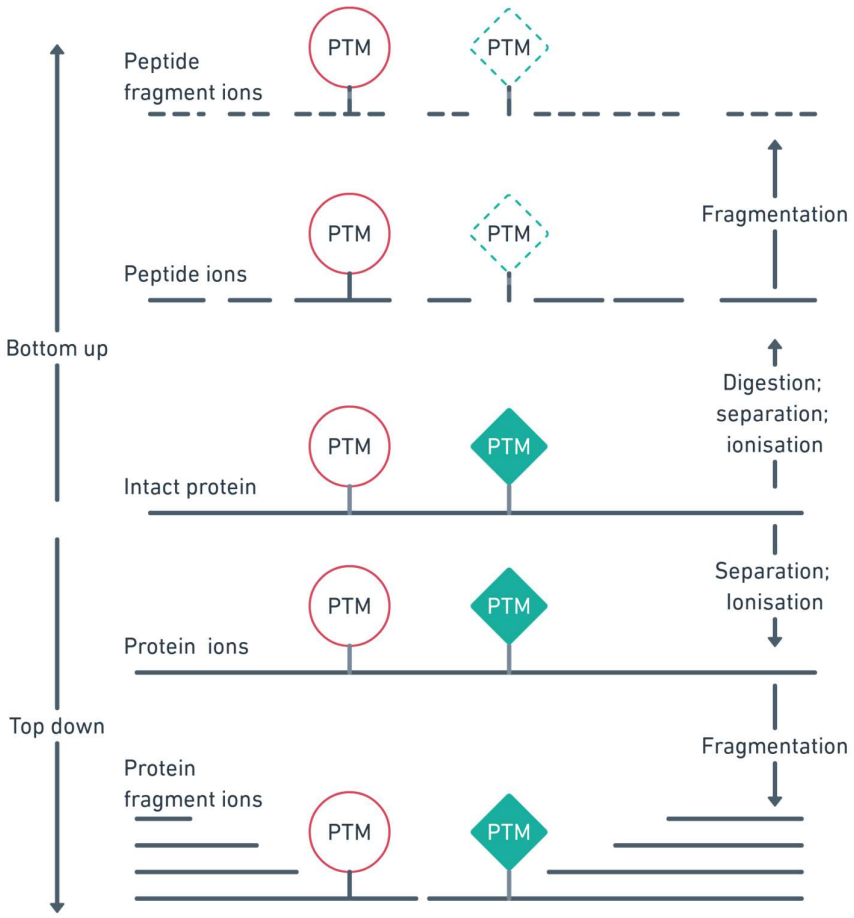


Figure III-1 Schematic representation of top-down vs bottom-up approaches. During the protein digestion into peptides in the BUP, some PTMs might get lost. Contrarily, top-down proteomics guarantees the preservation and identification of all PTMs present in the protein. Figure elaborated from Catherman et al., 2014.

The technique, hence, is the most reliable procedure for the investigation and characterisation of the *proteoforms*, a term established to denote the complex assembly of molecular protein forms of one single gene consequent to various natural and genetic variations, alternative splicing, allelic alterations and multiple combinations of PTMs [7],[8].

Even if top-down proteomics presents great potentialities and considerable advantages over the bottom-up approach, it is less implemented. The workflow has to face technical limitations and difficulties, such as the reduced solubility of proteins and the different physicochemical properties of one proteoform (i.e., size, charge, hydrophobicity) and its wide dynamic range that make the separation considerably challenging [1, 9]. TDP also required the use of high-resolution instruments to resolve the high molecular weight ion patterns.

The MS analyses of intact proteins result in complex data and fragmentation spectra that are still not entirely automatically interpreted by the available bioinformatics tool. Nonetheless, in the last years, significant progress has been achieved in reducing the sample complexity prior to MS and the improvement of separation methods, MS instrumentation, and computational solutions to support high-throughput data [10, 11],[12].

A combination of the BUP and TDP approaches is also proposed and named middle-down proteomics [13]. The method is based on a limited proteolytic cleavage that produces middle-range polypeptides from 3.0 kDa to 10 kDa, facilitating the analyte investigation.

2. Top-down workflow

The top-down workflow is typically divided into three main steps: (i) extraction and fractionation of complex mixtures, (ii) MS and MS/MS analysis and (iii) data processing. Theoretically, this approach differs from BUP mainly in the lack of protein digestion; however, the significant variances between the two analytes compel the entire TDP procedure to be adapted.

2.1 Extraction and fractionation

Likewise to the other approach, the analyte has to be extracted from its surrounding matrix to assure the injection of only proteoforms. Large proteins require detergents, such as Sodium dodecyl sulfate (SDS), to tackle their limited solubility. However, if implemented, SDS should be carefully cleaned due to its incompatibility with chromatographic separation and MS analysis [14].

A commonly implemented strategy compatible with TDP is the precipitation, for example, with heptane mixture [15], ethanol[14], chloroform-methanol water, and acetone [16-18]. Alternatively, the filtration (to disrupt the SDS-protein bonds) is also used through cellulose spin MWCO filters [19], FASP [20], or transmembrane electrophoresis (TME) devices [21]. Fractionation before MS analysis is critical to reducing the intact proteoform's complexity, thus enriching the signal and ensuring the detection of less abundant proteoforms [10, 11]. Different separation approaches, both chromatographic and electrophoretic, have been designed for intact proteins; the most common fractionations have been resumed in Table III-1. The combination of several separation techniques can be exploited to resolve the whole proteoform and guarantee full coverage [22, 23].

Chromatographic (online)		Electrophoresis (off-line)	
Ion exchange chromatography (IEX)	Exploitation of charge differences for the analyte separation by increasing the ionic strength of the mobile phase that flows through a charged stationary phase [24].	2D-gel SDS-PAGE	Probably the most common strategy in BUP but rarely applied in TDP because of the difficulties in the extraction and the consequent poor protein recovery.
Hydrophylic interaction liquid chromatography (HILIC)	Polar stationary phase and gradients of increasing water content. Hydrophobic molecules elute first. Recent development of capillary-based HILIC-MS helped with the problem of low volume loading [25].	Continuous-elution gel electrophoresis (mini prep-cell by Bio-Rad Laboratories)	Continuous elution of the sample into a cylindrical gel column, with a separation of the single factions and their collection at the end of the tube. SDS replacement with an acid-labile surfactant [26].
Reversed-phase chromatography (RPLC)	Implementation of non-polar stationary phase and a polar mobile phase. Hydrophilic analytes elute first. Long columns (>1 m) and stationary phases composed by short alkali-chains (C2-C5) or monolithic phases provides improved protein recovery and highest peak capacity [27].	Gel-eluted liquid fraction entrapment electrophoresis (GELFrEE by Expedeon)	It differs from prep cell by the addition of a sample collection chamber for the manual collection of the fractions. Reduction of HMWM dilution and dispersion into many fractions. [28].
		Isoelectric solution focusing (IEF)	Sample separation based on differences in isoelectric points. [3].
		Capillary electrophoresis (CE)	Capillary zone electrophoresis (CZE): sample separation on electrophoretic mobility diversities in analytes inside an open narrow capillary. Capillary isoelectric focusing (CIEF): a polyacrylamide coating in the capillary reduce the protein adsorption, increasing in turn the sensitivity, resolution and speed [29].

Table III-1 Protein mixture fractionation in the top-down workflow.

Among the group of capillary columns for chromatography online, monolithic stationary phases are highly exploited in top-down analyses [30-34] because of their excellent features, such as high resolution, remarkable selectivity and ability to resolving peaks. These columns consist of a unique bloc of the porous medium (called silica-rod), which fills its volume, composed by a cross-linked network of small mesopores (which involved in the retention and separation of the analytes) interlinked by large flow-through pores where the molecules of the solute migrate [35]. The resulting high permeability and the lack of void volumes in the column reduce the pressure during the fractionation [36] and, in turn, enable the use of short gradients with great flow-rate.

2.2 Mass spectrometry analysis

2.2.1 Ionisation methods

The ionisation of intact proteins is prevalently achieved through electrospray ionisation (ESI) because, producing multiple charge state ions, it provides high resolving power and facilitates the fragmentation with a consequent improvement of protein and PTMs characterisation [11]. Nonetheless, ESI also presents some inconveniences, like the production of signals with low S/N and the co-ionisation of various proteins with the induction of overlapping peaks and output of complex interpretation data [37]. In top-down proteomics, the MALDI source is mostly applied to rapidly screen multiple proteins and imaging studies [38, 39].

2.2.2 Mass analysers

The investigation and characterisation of complex proteoforms require high sensitivity, high mass accuracy (5 Hz), high mass resolving power (>50,000 FWHM) to detect the protein precursor ions as well as fragmentation ions [10]·[40]. An adequate sensitivity is also critical in the investigation since high molecular weight (HMW) proteins have a broader isotopic distribution than low molecular weight proteins (i.e., the signal of a single protein is distributed in multiplied charged state ions peaks) [41].

The high-resolution MS instruments most implemented in top-down approaches are quadrupole time-of-flight (Q-TOFs), (Bruker and Waters) characterised by a theoretically unlimited dynamic range but with a low mass resolving power [10], Fourier transform ion cyclotron resonance (FTICR) (Bruker) [42], and Orbitrap (Thermo) [43].

Orbitrap analyser assures high resolving power (up to 1,000,000 for m/z 200) and mass accuracy (<3 ppm external), even if the FTICR is characterised the highest resolving power achieved so far (> 10,000,000) [44].

Multiple hybrid architectures constituted by different mass analysers (for instance, the benchtop quadrupole–Orbitrap [45-47], Q h-FTICRs [48, 49]) have already been successfully applied to numerous top-down experiments for their improved features.

2.2.3 Fragmentations

The isotopic distribution in the MS protein spectra already provides a complex but accurate set of data. However, the production of a sufficient number of specific fragment ions is critical for a comprehensive overview of the intact protein sequence and the identification and localisation of its post-translational modifications. Among the various MS2 fragmentation techniques, higher-energy collisional dissociation (HCD) [50-52], electron-based dissociations (both ECD [53-55] and ETD [56, 57]), sustained off-resonance irradiation collisionally activated dissociation (SORI-CAD) [58], Infrared multiple photon dissociation (IRMPD) [59] and UVPD [51, 60] have been implemented in TDP experiments. The combination of infrared photoactivation with ETD in the hybrid activated ion-ETD dissociation (AI-ETD), provided the enhanced characterisation of proteins [61-63].

2.3 Data elaboration

Contrary to the bottom-up approach, only a few software solutions are available for the MS and MS/MS interpretation of intact proteins, which often require a manual inspection for validation and supplementation [64]. The lack of bioinformatics is related to several challenges that computational tools have to face. For instance, protein spectra present a great complexity related to the large protein size and the formation of multiple charge states that distribute the ion signal over a broader range, leading to a reduction of the signal-to-noise ratio [65].

The data interpretation often starts with the spectra's deconvolution to reduce the MS2 complexity; at this stage, various fragments ions with the same chemical formula and charge state are gathered into isotopomer envelopes [40]. The principal algorithms for the automated deconvolution are THRASH [66], Xtract [5] and MS-Deconv [67]. With the definition of both charge state and monoisotopic mass of each envelope, the theoretical isotopic distribution can be predicted.

However, an exact assignment of proteoforms might be prevented by overlapping or wrongly-attributed peaks. Therefore the high resolving power is crucial for the correct elaboration [1].

The final step of the data analysis is comparing the observed envelope with theoretical isotopic distributions and the consequent achievement of a list of potential proteoform attributions with a defined score, called PSM (Protein–Spectrum-Match). The software main implemented in top-down proteomics are listed in Table III-2.

Table III-2 Principal software implemented in top-down proteomics.

Software	Characteristic	Reference
ProsightPC	First and more flexible tool developed for top-down identification of protein and PTMs. It implements multiple search modes to analyse Thermo Scientific raw data generated by various fragmentation techniques, including EThcD. It allows the exploitation of proteome warehouses containing the information present in UniProt (such as allelic variations, SNPs, and several PTMs). Supports of also middle-down experiments. ProSightPD is a node for Proteome Discoverer software.	[68], [69, 70]
Prosight Lite	Simplified but free Windows version of ProsightPC for proteoform characterisation and localisation of PTMs. The application enables the matching between theoretical and experimental data of a single intact protein and the visualisation of its corresponding fragment ions and PTMs sites.	[71, 72]
Mascot TD (Big Mascot)	An additional version of Mascot software (Available from Mascot 2.2 onwards) that allows the investigation of proteins with precursor mass between16 kDa and 110 kDa.	[73]
MS-Align+	Analysis tool based on the spectral alignment that leads to the identification of proteoforms with unexpected PTM and mutations. It provides a statistical evaluation of PrSMs with high confidence. It also enables the efficient filtering of candidate proteins when searching against an extensive protein database.	[74]

ToPIC	ToPIC (Top-down mass spectrometry-based Proteoform Identification and Characterization) is the updated version of MS-Align+ that provides increased speed, sensitivity, and accuracy to identify proteoforms with unexpected PTMs efficiently.	[75] [76]
PIITA	PIITA engine (Precursor Ion Independent Top-down Algorithm) is a precursor ion independent method that exploits only fragment ions for protein identification. It enables identifying proteoforms with unpredicted PTMs and provides FIT scores and Δ scores rather than statistical significance estimates.	[77]
MS-Deconv	Software that implements a combinatorial algorithm for top-down spectral deconvolution. Contrarily to other tools, the algorithm scores groups rather than single isotopomer envelopes for a spectrum.	[67]
MASH Suite	Freely available software interface for the processing of top-down, middle-down and bottom-up MS data. Provides a tool to visualise complex raw data and manually validate the results achieved.	[78]
MSPathFinder	Database search engine. In a similar manner to BUP tools, it provides the identification of proteoforms with combinations of PTMs and truncations and the following elaboration of statistical proteoform spectrum matches (PrSMs) and scores.	[64, 79]

3. Top-down proteomics applied to cultural heritage studies

Bottom-up proteomics is currently the mainstream technique for identifying proteins in historical and artistic samples, as discussed in the previous chapter. Nonetheless, the increasing interest in an in-depth understanding of the proteinaceous compounds and their molecular/structural modifications prompts the researchers to employ alternative strategies, such as the top-down approach, to overcome the limitations of the peptide-based analysis [80]. The top-down approach can provide a better insight into protein ageing, degradation processes, and environmental impact by inspecting various proteoforms, including allelic variations, PTMs, and polypeptide chains that underwent partial truncation or cross-linking. This knowledge may assist, in turn, the definition of more adequate protocols for appropriate restoration and conservation treatments. Furthermore, notwithstanding a few unique peptides which can already identify the biological origin of proteins, top-down proteomics enables more direct discrimination of species with considerable sequence resemblance, as in the case of collagen α2(I) that present two amino acids of difference (located in the same tryptic peptide) between the breeds of *Capra hircus* and *Ovis aries*.

Although, besides the several challenges shared with biological samples analysis [81-83], one of the main issues that have curbed the investigation of proteins in artworks with top-down is the low sensibility offered for large analytes, which commonly compel an increase in the material analysed, an option that is to disregard in this field.

Pioneering studies of ancient proteins in their integrity were already conducted in the late 1980s on paleontological samples with the successful identification of intact proteins with electrophoresis gel [84-86], followed by some investigations with MALDI-TOF around the 2000s [87-91]. More recently, former digestion-free analyses have been conducted on high-resolution LC-MS on paleontological remains [92, 93] in which undigested fragments were investigated with the support of bottom-up instrumental set up. First studies of intact proteins in artificially aged painting mock-ups date from the early 2000s [80, 94, 95]. They lead to a more in-depth insight into protein degradation, identifying structural (partial breakdown and cross-linking) and chemical (oxidations) modifications.

4. Aim of the research

The principal aim of this research was to develop a robust top-down method, consisting of sample preparation and instrumental set-up, for the investigation of intact proteins and their proteoforms in artistic and historical artworks from very low amounts of sample (i.e., trace level analysis of samples collected implementing a minimally invasive technique).

5. Materials

5.1 Standard samples implemented for the protocol development

The methodology was developed using a mixture of commercially available casein standards from bovine milk purchased from Sigma-Aldrich: beta-casein (β-CN), >98% and alpha S-casein mixture of alpha casein S1 and S2 (αs_1-CN and αs_2-CN), ≥ 70% prepared at a concentration of 1 mg/mL in deionized water (stock solution). The stock solution was then diluted to optimise both sample preparation and instrumental settings (definition of the most appropriate chromatographic column to use, an optimized gradient and the most efficient fragmentation method). The stock solution was stored at −20 °C for up to 4 weeks.

5.2 Casco industrial glue

The protocol was then assessed for older and more degraded samples by investigating a few tens of μg of Casco industrial glue (1900-1950) (Figure III-2). The samples were provided by Dr Julie Arslanoglu, *Department of Scientific Research, The Metropolitan Museum of Art, New York*.

Figure III-2 Casco Powdered Casein Glue.

This casein-based glue, sold by the Casein Company of America, was resistant to moisture and heat, and it was suitable for different materials, including timber and plywood. Formerly implemented for aeroplanes during World War 1, Casco glue was later made available for general use and promoted for several purposes (including sizing canvas for paintings, tempering plaster and bookbinder).

The preparation process principally consisted of a division of the curd (crude casein) from the skim milk's whey followed by other steps, among which the pressurisation, drying, grounding of the casein into dry powder and blending with various chemical additives [96].

5.3 Artistic and historical sample: Thomas Gainsborough drawing

The top-down proteomic approach was finally conducted on a micro-sample (III,63_S1B2) collected from the artwork "*Landscape with Horse and Cart Descending a Hill*" made by the English artist Thomas Gainsborough around ca. 1780 **(**Figure III-3).

The study was investigated in partnership with Dr Julie Arslanoglu and Dr Federica Pozzi (*Department of Scientific Research, The Metropolitan Museum of Art, New York*) and Reba Snyder (*Thaw Conservation Center, The Morgan Library & Museum*)**.**

Figure III-3 Landscape with Horse and Cart Descending a Hill. The sample collected for the top-down analysis is the III, 63_S1B2 (6 μm grit). Photo from the web site of The Morgan Library&Museum.

The specimen was collected together with the samples for bottom-up investigation (Chapter II2.1 on page 30), implementing a minimally-invasive sampling. Specifically, a small disc (Ø 5-6 mm) of aluminium oxide Fiber Optic Polishing Film Discs (PFP) of 6 μm grit was lightly scraped on the paper surface [97].

6. Methodology

6.1 Development of sample preparation based on eFASP protocol

The methodology was developed based on the eFasp protocol [98] optimised for the bottom-up analysis of a few µg of ancient and degraded samples (chapter II). A 200 µL of a mixture of urea 6 M and 0.2% DCA in 100 mM ABC at pH 8.8 was added to few tens of µg of dry powder (standard proteins and Casco sample) or directly to the small disc of 6µm grids (micro-sample from Gainsborough drawing) until fully covered. The solution was then incubated overnight at 8 °C under agitation.

Concomitantly, the Amicons® Ultra 0.5 ml – 3 kDa cutoff units were incubated in a passivation solution of Tween 5%. The next day, the protein solution was transferred to the filter and subjected to three cycles of buffer addition (urea 6 M, 0.2% DCA in 100 mM ABC at pH 8.8) and centrifugation (12000 g). If required, the protein can be subjected to reduction and alkylation of the disulphide bonds: addition of 200 µL of reduction buffer (urea 6 M, 20 mM DTT in 100mM ABC, pH 8.8) and incubation for 1h at room temperature with shaking (followed by a spin to discharge) and incubation of the solution with 200 µL of alkylation buffer (urea 6 M, 50Mm IAA in 100mM ABC, pH 8.8) in the dark and shaking for 1h at 37°C. After the alkylation buffer discharge, the protein solution was once cleaned with buffer (urea 6 M, 0.2% DCA in 100 mM ABC at pH 8.8), then three times with 50 mM ammonium bicarbonate pH 8.8 and finally three times in pure water to remove interfering substances.

After the recovery of the filter solution, a further liquid-liquid extraction was performed with 800 µL of ethyl acetate mixed to 2.5 µL of pure TFA and repeated twice with only ethyl acetate. Following the complete evaporation of the solution through a speed vacuum system, the protein mixture was resuspended in a minimal volume of deionised water (4µl) (Figure III-4).

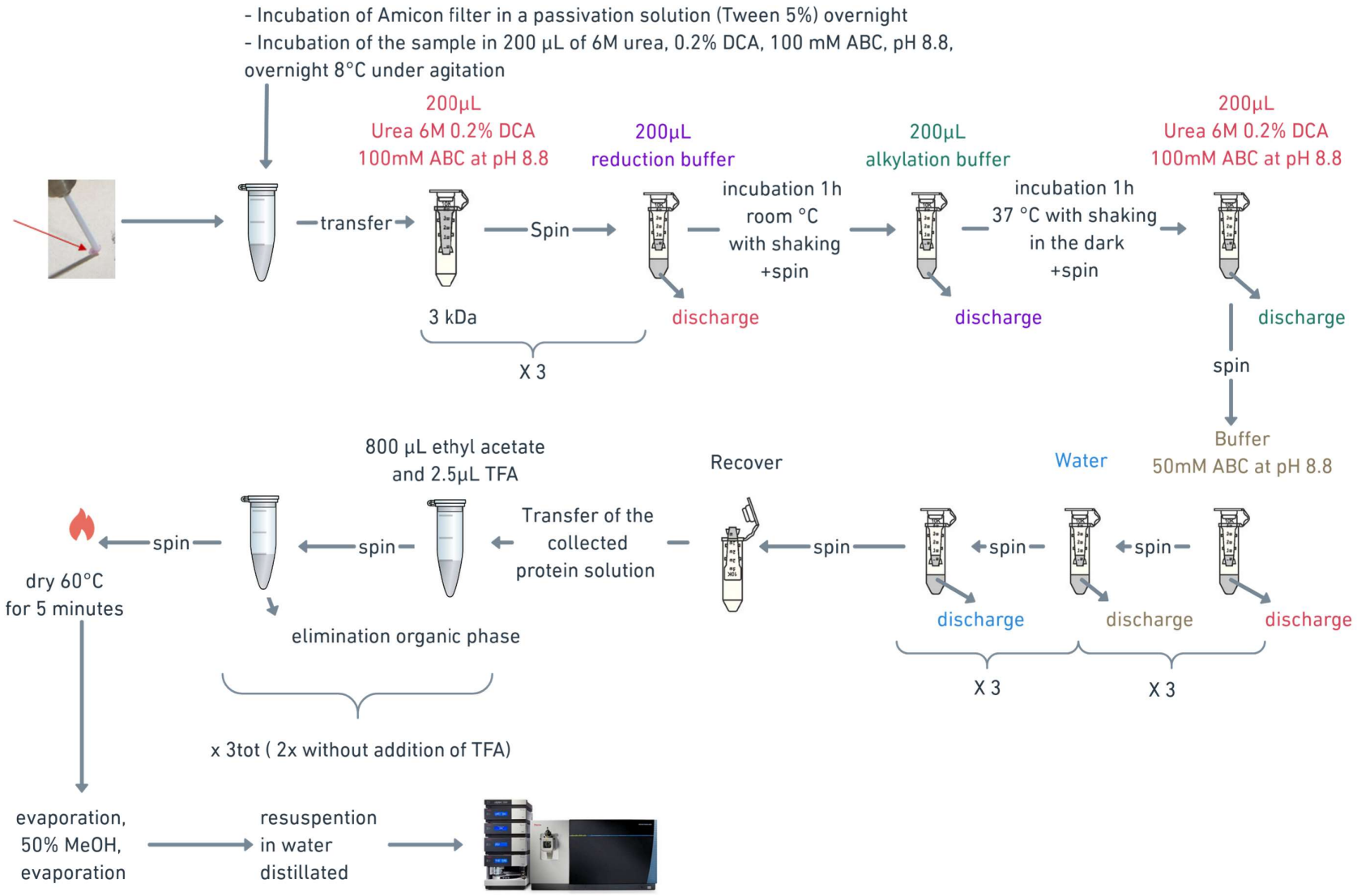
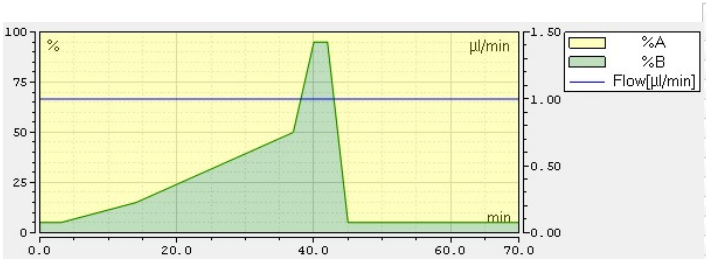


Figure III-4 Scheme of top-down eFasp methodology.

6.2 Fractionation (Liquid Chromatography)

The column used was the EASY-Spray PepSwift Monolithic (PV-DVB) column at 200 µm x 250 mm at 30 ºC with 0.875 µL/min flow rate. Monolithic columns were already implemented to analyse intact milk proteins, evidencing a better resolving power and higher resolution than the C4-C5 columns [99].

A pre-column C4 PepMap300 at 300 µm ID x 5 mm, 5 µm pore was implemented with a flow rate of 10 µL/min flow rate. Buffer A was 95% H2O, 5% Acetonitrile, 0.2% Formic acid, while Buffer B was 95% Acetonitrile, 5% H2O, 0.2% Formic acid. The gradient established was of 70-minutes: 5% B for 0-3 minutes, 15% B at 15 minutes, 50% B at 37 minutes, 95% B for 40-42 minutes, and 5% B for 45-50 minutes (Figure III-5). Column equilibration and sampler washing were performed in the last 20 minutes before proceeding to the next sample in the queue. The sample injection was performed implementing the µLPickup mode that allowed the injection of volume as low as 1 µL.



No	Time	Flow [µl/min]	%B	Curve
1	0.000	Run		
2	0.000	1.000	5.0	5
3	3.000	1.000	5.0	5
4	14.000	1.000	15.0	5
5	37.000	1.000	50.0	5
6	40.000	1.000	95.0	5
7	42.000	1.000	95.0	5
8	45.000	1.000	5.0	5
9	50.000	1.000	5.0	5
10	*New Row*			
11	70.000	Stop Run		

Figure III-5 70-minute gradient curve and flow rate.

6.3 MS and MS/MS Methods

The samples were analysed with the Orbitrap Fusion Lumos (Thermo Fisher Scientific, San Jose, CA, USA) in positive Intact Protein Mode with a pressure of 2 mTorr. MS parameters were 120 k resolution, AGC target of 5.0e5 and a scan range of 600-2400 m/z using Xcalibur 4.2 (Thermo Fisher Scientific, San Jose, CA, USA). The implemented Data Dependent mode workflow, defined as 3 second Cycle Time. The Monoisotopic Peak Determination (MIPS) was set to intact protein. This feature, available in Orbitrap instruments, allowed the monoisotopic ion's identification and selection within an isotopic cluster for data-dependent scans. Independently of the most abundant isotope intensity, the monoisotopic mass can be introduced in the scan header providing the exclusion of all other ions in the same isotopic cluster.

The charged state range was set among 5-24 to filter out the signals from interfering substances and noise, but the detection of 25^+ charge state and higher was enabled.

For MS/MS analysis, the quadrupole isolation was applied to remove ions with mass-to-charge ratios outside the scan range before entering the Orbitrap mass analyser. The isolation window was set to 3 m/z with the RF Lens at 30%. The fragmentation was performed with electron transfer dissociation with supplemental higher-energy collisional dissociation (EThcD) to provide b, c, y, and z ions. The ETD reaction time was set at 10 ms. The collision energy of HCD was at 15% to reduce excessive fragmentation and maximum injection time at 50 milliseconds. The other settings were: 3 m/z isolation window, resolution at 120,000, AGC target of 5.0e5 and m/z 600-2400.

6.4 Data processing and analysis

The extracted data were processed using Proteome Discoverer 2.4, complemented with ProSight PC 4.0 (Thermo Fisher Scientific, San Jose, CA, USA) (ProSight PD 1.3). The software implements the deconvolution algorithm Xtract™ for intact ESI spectra and creates proteome warehouses from the UniProt database by gathering proteins sequence and information about PTM, sequence variants, and single nucleotide polymorphisms (SNPs). The results were searched against *Bos taurus* (Proteome ID UP000009136) top-down database with 1.05 Da precursor tolerance and 10 ppm fragment mass tolerance.

The workflow followed for the analysis was ProSightPD HI/HI Three Tier Search consisting of (i) an initial absolute mass search against all intact proteoforms of proteins in the database, followed by (ii) a biomarker search with a narrow precursor and fragment ion mass tolerance and by (iii) second absolute mass investigation as an error-tolerant search. The absolute mass search implements the precursor ion mass to create a subgroup of the proteome database in which the theoretical fragments ions are compared with the experimental fragment ions. The assignments were reported under the P-Score, based on the number of fragment ions matching with the theoretical masses.

The biomarker search is instead comparable to a bottom-up data analysis, without the selection of the enzyme: it is a two-step process consisting of the initial detection of a potential theoretical mass matching with the observed one, followed by the calculation of the possible theoretical fragment ions and their comparison with the experimental fragment ions of the candidate.

For the two absolute mass investigations, the precursor ion tolerance was set at 1.05 Da, and the fragment tolerance was set at 10 ppm, while for the biomarker search, the mass tolerance was 10 ppm for both precursor and fragment masses. Multiple runs have been conducted to examine different parameters, particularly various modifications not incorporated in the protein database were inspected, such as deamidations and oxidations. Due to the complexity of the data set, a manual inspection of the MS spectra resulted crucial in the identification of several proteoforms not detected by the software. The data obtained on intact proteins of the historic sample were compared with results previously acquired on the bottom-up approach on different micro-samples collected from the same Gainsborough drawing.

7. Analysis and discussion of the results

7.1 Standard samples implemented for the protocol development

Formerly, the development of the top-down investigation was conducted on a mixture of two commercially available casein standards (αs-CN and β-CN). The principal objective was to achieve a proper separation and characterisation of the casein proteoforms, starting from the little amount of material injected down to 500 fmol (for successful identification). A clear protein mixture separation is illustrated in the chromatographic peaks profile in Figure III-6: a first elution resulted around 31.70 min (blue point) followed by a second one at 32.04 min (orange point).

The high-quality resolution achieved in the MS spectrum ensured the detection of various αs-CN and β-CN proteoforms. For example, three protein isoforms of αs-CN with different variants and post-translational modifications were detected at the first elution interval. The predominant massive of multiple charges from 29^+ to 11^+ (Figure III-6 left_blue point) was ascribed to α_{S1}-CN_B-8P (i.e., variant B with 8 phosphorylations) with an observed monoisotopic mass of 23600.2 Da and Δm= 1.1 ppm (theoretical 23600.2 Da, $C_{1035}H_{1595}N_{265}O_{341}S_5P_8$).

A second low-intensity isotopic pattern with a shift of +16 Da was attributed to the oxidised form of α_{S1}-CN_B-8P. Another pattern at 23528.3 Da was assigned to the natural variant C of α_{S1}-CN (substitution of Glu_{207}->Gly_{207}), presenting 8 phosphorylations ($C_{1032}H_{1591}N_{265}O_{339}S_5P_8$, Δm =4.0 ppm).

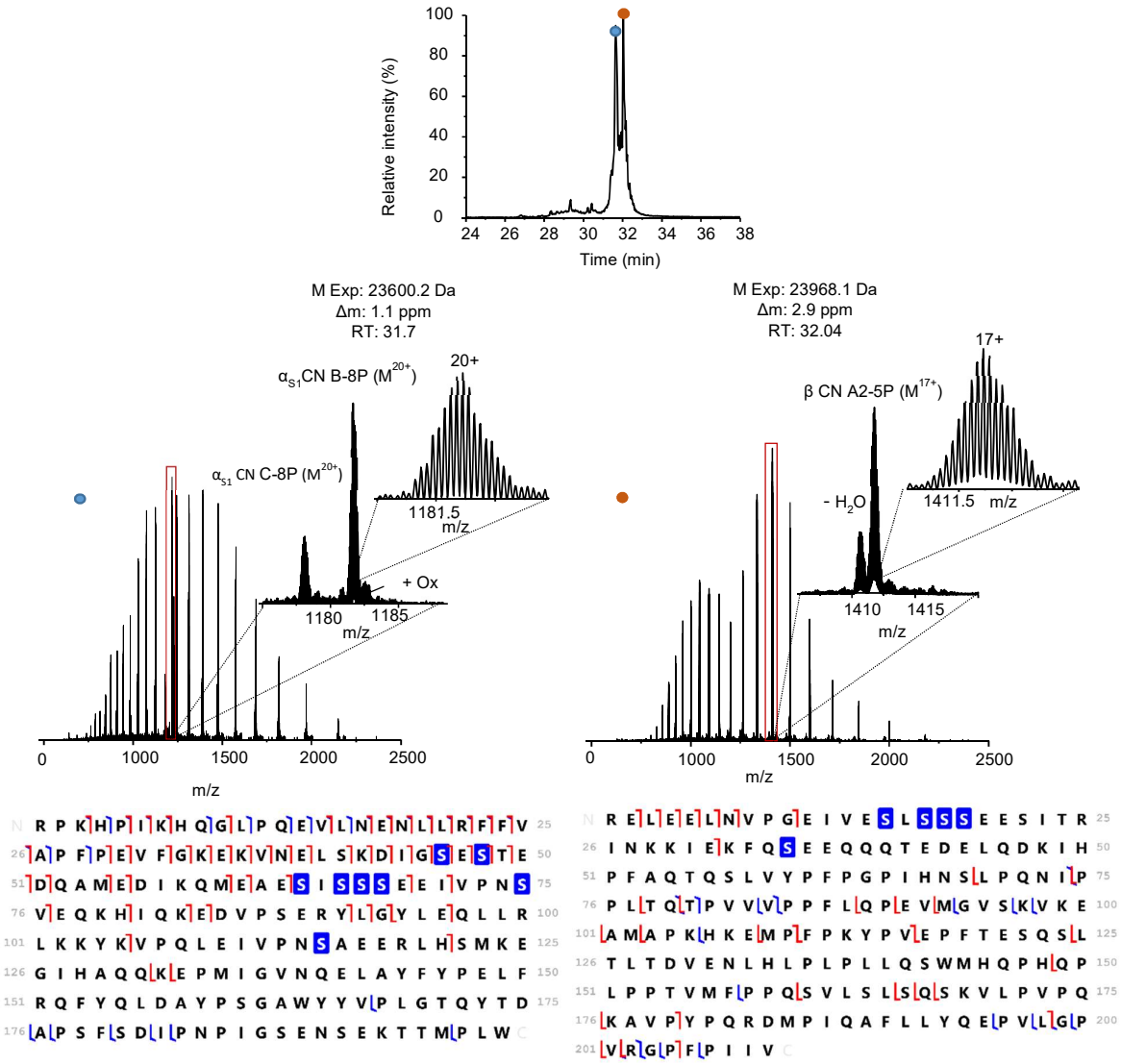


Figure III-6 NanoLC nano ESI-Orbitrap MS spectrum of commercially available casein standards investigation. On the top of the figure, in the TIC chromatogram are highlighted the two peaks corresponding to the elution of α_{S1}-CN (blue point) and β-CN (red point). Mass spectra for casein proteins observed with multiple charge state distributions, first zoom on the principal proteoforms followed by a second zoom on the isotopic pattern of the two proteoforms α_{S1}CN B-8P (20^+, m/z: 1181.017) and β-CN A2-5P (17^+ m/z 1411.732), respectively. The fragment map of the two selected ions reports the matching fragment ions and the O-phospho-L-serine modifications' localisation.

The EThcD fragmentation of the 20^+ charged ion m/z 1181.017 provided the confident characterisation of the intact protein α_{S1}-CN_B-8P and the accurate localisation of the 8 O-phospho-L-serine (S61-S63-S79-S81-S82-S83-S90-S130) via the detection of 94 matching fragments.

A more phosphorylated proteoform, α_{S1}-CN_B-9P with 9 phosphorylations, was also detected at a slight higher RT (31.84-31.90 min); the observed mass was 23680.2 Da, Δm =3.1 ppm ($C_{1035}H_{1596}N_{265}O_{344}S_5P_9$).

The predominant pattern eluted in the second elution range (RT 32.04-32.24 min) illustrated in Figure III-6 right_orange point shows multiple charges from 30^+ to 11^+. The high isotopic resolution ensured a confidential attribution of the main isotopic pattern with a monoisotopic mass of 23968.2 Da to the most common beta-casein variant in dairy cattle, β-CN_A2-5P, namely variant A2 with 5 phosphorylations ($C_{1080}H_{1697}N_{268}O_{325}S_6P_5$, Δm=2.9 ppm). A second lower isotopic pattern corresponding to the loss of a water molecule was also detected (monoisotopic mass 23950.2 Da).

As in the case of α_{S1}-CN, the MS observations were further confirmed by a significant fragmentation of the proteoform β-CN_A2-5P: the attribution of several fragments (c/z and y/b) in the MS2 of the 17^+ charged ion m/z 1411.732 provided the accurate protein characterisation and the localisation of the five 5 O-phospho-L-serine (S30-S32-S33-S34-S50). The second most frequent variant A1 of the β casein with 5 phosphorylations was also identified at a higher RT (32.30-32.35 minutes) with an MW 24008.2 Da, Δm=3.6 ppm, $C_{1081}H_{1697}N_{270}O_{325}S_6P_5$.

Similarly to other publications [99], no intact α_{S2}-CN variants were observed in the casein mix standard. However, two truncated forms of this casein eluted at RT 20-23 min were identified through fragmentation patterns: one fragment (MW 4290.95 Da Δm=3.3 ppm, $C_{172}H_{290}N_{47}O_{72}S_1P_3$) was attributed to α_{S2}-CN sequence (130-165) with three O-phospho-L-serine at S146-S150-S158, while 11^+ charged ion m/z 642.0036 correspondings to MW 7050.95 Da was assigned to an α_{S2}-CN truncated form (166-222) without modifications ($C_{330}H_{525}N_{87}O_{82}S_1$, Δm= 3.2 ppm,). All the proteoforms characterised have been summarised in Table III-3.

Table III-3 List of the casein proteoform identified in standard casein

Protein	Accession	Position	Theoretical MW (Da)	Observed MW (Da)	Error (ppm)	RT	Natural variant	Modifications
α_{S1}-CN	P02662	16-214	23600.2	23600.2	1.1	31.70	B	8 phospho (S61-S63-S79-S81-S82-S83-S90-S130)
			23616.2	23616.2	1.2	31.73		8 phospho + 1ox
			23680.2	23680.2	3.1	31.84		9 phospho (S56- S61-S63-S79-S81-S82-S83-S90-S130)
			23528.2	23528.3	4.0	31.77	C	8 phospho
β-CN	P02666	16-224	23968.2	23968.2	2.9	32.04	A2	5 phospho (S30-S32-S33-S34-S50)
			23950.1	23950.2	2.9	32.04		5 phospho – H2O
			24008.1	24008.2	3.6	32.31	A1	5 phospho
α_{S2}-CN	P02663	130-165	4290.95	4290.95	3.3	20.11		3 phospho (S146-S150-S158)
		166-222	7050.93	7050.95	3.2	23.66		

7.2 Casco industrial glue

Few tens of µg of Casco powdered casein glue (commercialised in the 20th century) were analysed to assess the protocol's potentiality in older and more degraded samples. The chromatographic profile of the protein solution presented a significant complexity defined by multiple signals eluted at shorter RT times and within a broader range compared to the modern standard mixture (Figure III-7-A).

Despite the very low amount of sample injected, different proteoforms of the four casein proteins were detected, mainly as truncation and modified products (Table III-4). For example, α_{S1}-CN, more than 60% of the protein sequence was identified and characterised by the detection of different truncated forms. For instance, the truncated proteoform 119-145 with one O-phospho-L-serine in the residue S130 was distinguished from the experimental monoisotopic mass 3182.56 Da (5^+ m/z 637.519, Δm= 1.2 ppm) eluted at RT 19.25min (Figure III-7-B-red point). A great extent of casein proteoforms with oxidations and deamidations was observed, presumably due to the older age of the standard sample. For example, the same truncated sequence 119-145 was detected oxidised (MW 3198.56 Da, Δm= 2.7 ppm), with the modification localised on the methionine residue M138 (17.05 min).

A cleaved form of α_{S1}-CN (25-60) presenting a deamidation was eluted around 29.46 min (Figure III-7 B-purple point). The straightforward characterisation of the proteoform was achieved through the interpretation of the MS/MS spectrum of 5^+ charged ion m/z 815.643 (Δm= 6.0 ppm). The generation of various dual ion series via EThcD fragmentation provided an accurate sequence reconstruction and the localisation of the deamidation on the asparagine residue N32 (Figure III-8).

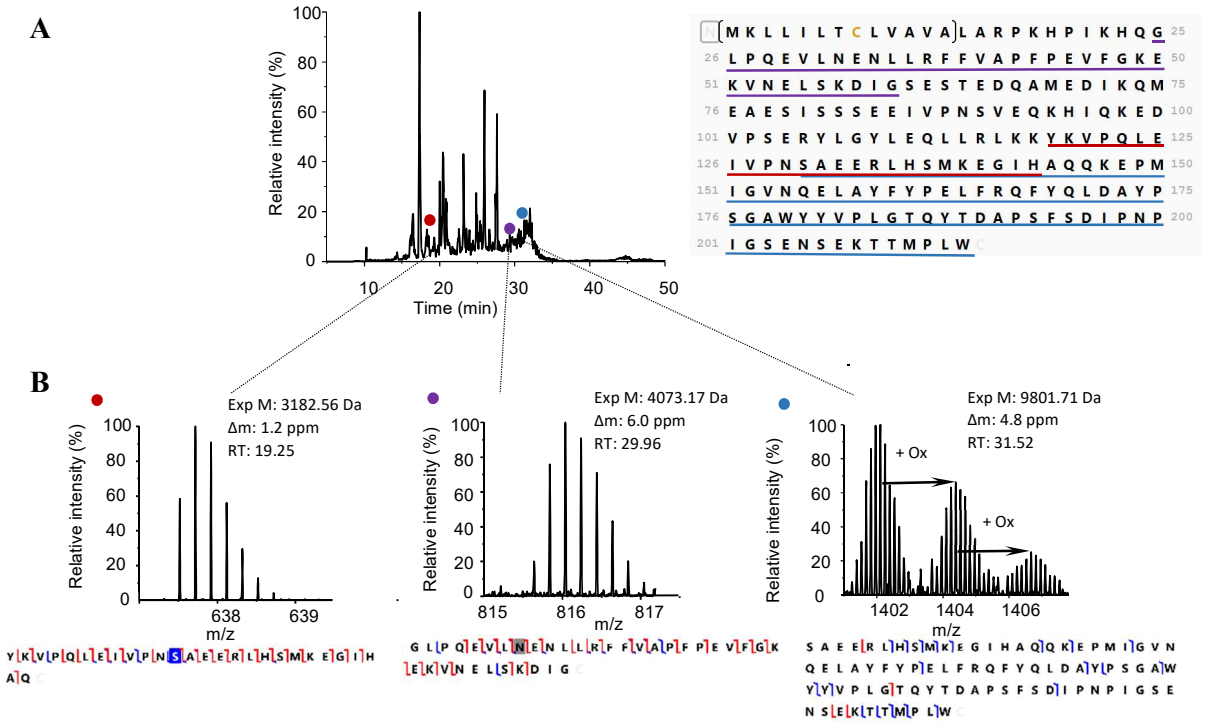


Figure III-7 NanoLC nanoESI-Orbitrap spectra related to Casco glue sample. **(A)** TIC chromatogram with a focus on casein S1. Three examples of truncated proteoforms of αS1CN are shown, reporting their RT with coloured points. The same colours have been used to underline the proteoforms in the protein sequence and illustrate **(B)** the charge state distribution selected for the MS2 (5^+ m/z 637.519, 5^+ m/z 815.643 and 7^+ m/z 1401.244, respectively). At the bottom, the fragment maps are reported with the matching fragment ions and the modifications' localisation. (blue for the phosphorylation and grey for the deamidation).

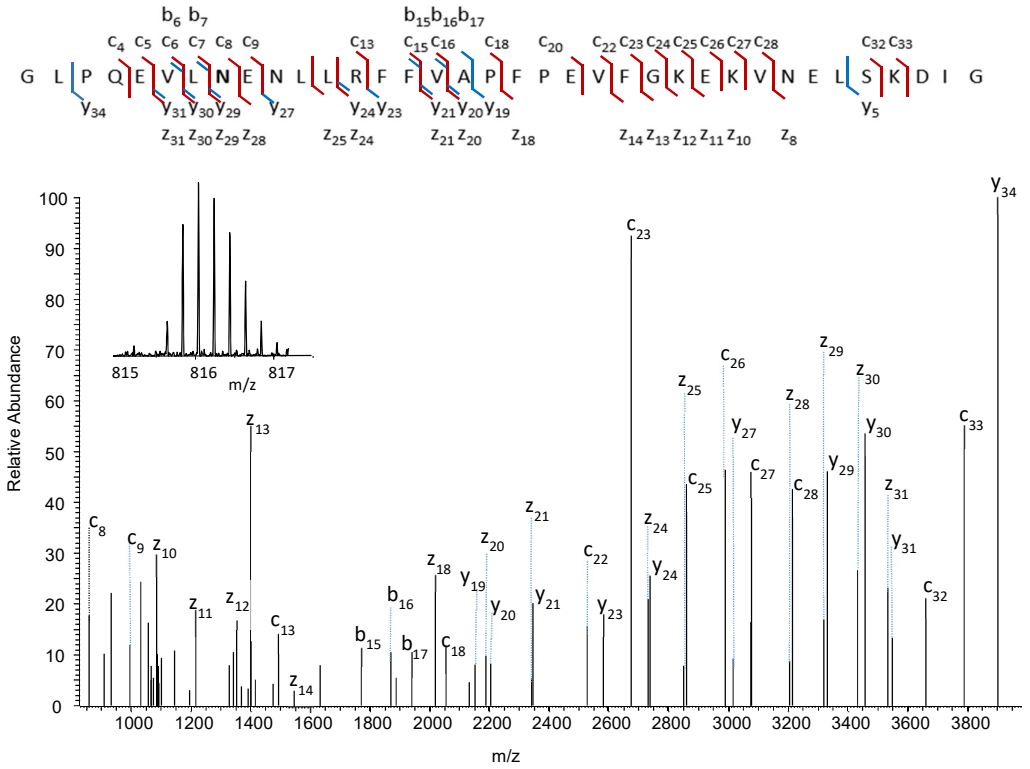


Figure III-8 EThcD spectra of the 5^+ charged ion m/z 815.643 from the Casco glue, presenting dual ion series (z/c and y/b) of the proteoform α_{S1}CN (25-60) with a deamidation localised on the asparagine residue (N32).

The last cleaved segment α_{S1}-CN (130-214) was identified through the MS and MS2 interpretation of the 7^+ charged ion m/z 1401.244 (Δm= 4.8 ppm) the amino acid Glu_{207} confirmed its attribution to the most common B variant of the casein (Figure III-7-B- blue point). As illustrated in the MS of the figure, two other protein forms were eluted with a sequential shift of +16 Da each, identifying two oxidation forms. In this case, the absence of fragmentation prevented an accurate definition and localisation of the modifications; the sequence presents several potential oxidation sites (such as three methionine, two histidine, two tryptophan and seven tyrosine residues).

The other caseins, β-casein, κ-casein and α_{S2}-CN, were also identified by detecting nine, two, and two truncation forms, respectively (Table III-4). Likewise to α_{S1}-CN, high heterogeneity was observed in all three caseins, characterised by natural variants (A1 and A2 variants of β-casein), various PTMs (phosphorylation, deamidation and oxidations) and different sites of truncation. Notably, cleavage patterns differing on the truncation site for one or a few amino acids were noted for β-casein.

Table III-4 List of the principal casein proteoforms identified in Casco powdered industrial casein-glue dated 1900-1950

Protein	Accession	Position	Theorical MW (Da)	Observed MW (Da)	Error (ppm)	RT	Modifications	Specie
α_{S1}-CN	P02662	25 - 60	4073.15	4073.17	6.0	29.96	Q (28)	*B.taurus, B.indicus, B.grunniens, B.mutus, B.b.bison*
		119-145	3182.56	3182.56	1.2	19.25	phospho (S130)	*B.taurus, B.indicus, B.mutus, B.b.bison*
			3198.55	3198.56	2.7	17.05	phospho (S130), 1 ox (M138)	
		130-214	9801.66	9801.67	1.2	31.52	-	*B. taurus*
			9817.65	9817.65	-0.3	31.52	1 oxidation	
			9833.65	9833.64	-0.9	31.52	2 oxidation	
β-CN	P02666	40-71	3992.93	3992.93	1.5	16	phospho (S50)	*B.taurus, B.indicus, B.mutus, B.b.bison*
		41 - 66	3261.56	3261.55	-1.2	13.97		
		41-71	3836.83	3836.84	4.0	16.97		*B.taurus, B.mutus, B.indicus, B.indicus x B. taurus, B.grunniens, B.b.bison, B. bubalis*
		44-71	3401.64	3401.65	3.5	16.73	-	*B.taurus, B.mutus, B.indicus, B.b.bison, B. bubalis*
		44-83	4803.29	4803.30	3.1	23.66	phospho (S50), P67->H67 (variant A1)	*B.taurus, B. indicus*
		121-156	4187.17	4187.19	6.7	28.43	-	*B.taurus, B.mutus, B.indicus, B.indicus x B. taurus, B.grunniens, B.b.bison, B.bubalis*
			4203.16	4203.19	6.9	28.06	1 oxidation M(124)	
		115-156	4814.47	4814.48	2.7	27.84	-	*B.taurus, B.indicus, B.indicus x B.taurus, B.grunniens, B.b.bison, B. bubalis*
		112-140	3372.73	3372.73	0.9	18.99	-	*B.taurus, B.mutus, B.indicus, B.indicus x B.taurus, B.grunniens, B.b.bison, B.bubalis*
		84-121	4103.23	4103.25	4.1	26.57	-	*B.taurus, B.mutus, B.indicus, B.indicus x B.taurus, B.grunniens, B.b.bison, B.bubalis*
κ-CN	P02668	127-190	6783.33	6783.34	1.3	25.34	phospho (S170)	*B.taurus, B.indicus, B.indicus x B.taurus*
		127-187	6512.18	6512.18	-0.3	24.78		
α_{S2}-CN	P02663	66-148	9872.68	9872.69	0.7	31.95	3 phospho (S71-S72-S146)	*B.taurus, B.mutus, B. indicus x B.taurus, B.b.bison, B.bubalis*
		144-198	6783.31	6783.27	-5.2	25.34	3 phospho (S146-S150-S158), 3 deam. (N149-N174-N177)	*B.taurus, B.mutus, B.indicus x B.taurus, B.grunniens, B.b.bison*

7.3 Artistic and historical sample: Thomas Gainsborough drawing

Lastly, the developed protocol was adopted to investigate milk-based fixatives present in the drawing *"Landscape with Horse and Cart Descending a Hill"* of Thomas Gainsborough. As reported in Chapter II, the samples for the proteomic analysis were collected with minimally-invasive techniques because of the exceptional importance of the artworks. Specifically, the micro-sample collected through fine polishing films of 6-µm grit was analysed with the top-down strategy.

Even if the amount of material investigated was extremely low (critical point in the investigation and fragmentation of intact proteins), the analysis succeeded in characterising different proteoforms attributable to the four caseins (Table III-5).

For instance, two cleaved proteoforms were attributed to α_{S1}-CN sequences (119-145), presenting one phosphorylation on S130 (MW 3182.55 Da, Δm= -1.6 ppm) and (116-151) having a deamidation modification localised on N129 (4199.19 Da, Δm= -6.9 ppm). Furthermore, MS and MS/MS analyses of the 7^+ charged m/z 688.937 allowed identifying the truncated proteoform of β-CN (114-155) with one deamidation localised on Q138 (4815.50 Da, Δm= 2.1 ppm).

The proteolytic proteoform κ-CN variant A of 104 amino acids (86-190), which represent more than 60% of sequence coverage (without signal peptide), was observed with a great extent of modifications, including phosphorylations and a high number (up to eleven) of deaminations on asparagine and glutamine residues (no glycosylation was identified possibly due to the very low sample amount analysed). The deconvolution of the multiply charged patterns from 10^+ to 7^+ charges eluted in the RT 33.30 – 33.75 min range corresponding to the monoisotopic mass of 11358.43 Da Δ= -1.58 ppm ($C_{487}H_{781}N_{123}O_{178}S_3P_2$). The EThcD fragmentation contributed to its characterisation. The weak fragmentation obtained was expected, considering the extremely low concentration of the sample.

Figure III-9 shows the MS pattern at RT 33.71 of the one proteoform of κ-CN variant A. It is interesting to note that the bottom-up proteomics data obtained for the same sample and top-down data are covering complementary sequences of the protein. The lack of the last part of the protein sequence (138-190) in bottom-up data, most likely related to the elevated length of the tryptic peptide.

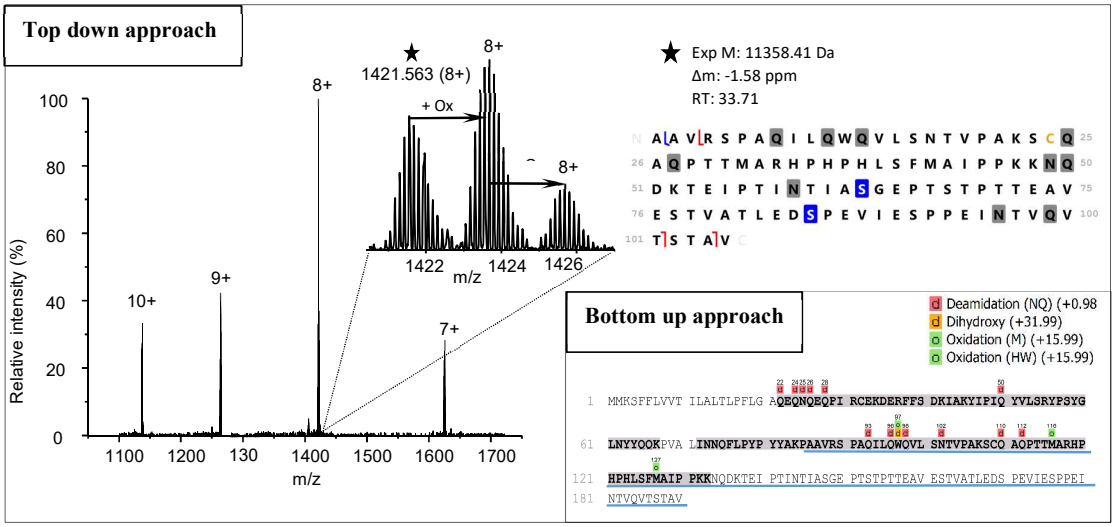


Figure III-9 MS spectrum of hydrolysed κ-CN forms of κ-CN variant A (86-190) and its oxidised forms identified using a top-down approach applied to a sample taken from a historic Gainsborough drawing. The identified sequence by MS and MS/MS experiments is shown on the right corner; the blue dots are phosphorylations, and the grey dots are other modifications (e.g. N/Q deamidations). The box on the right corner illustrates the sequence coverage of kappa casein observed with a bottom-up approach from the same drawing. The identified peptides are highlighted in grey, while the truncated proteoform identified using the top-down method is underlined in blue; the localisation of deamidations, oxidation and di-hydroxylations is also reported.

Two oxidised forms of the truncated κ-CN proteoform, presenting the same modifications, i.e. 2 phosphorylations and 10 deamidations, also eluted at the same RT (MW 11374.44 Da and 1139.41 Da, for one and two oxidations, respectively). The bottom-up data suggested the localisation of the oxidations on the methionine residues (M116 and M127) or tryptophan (W97).

A more in-depth examination of the MS spectra in this elution range revealed other patterns corresponding to various deamidation forms. For example, the proteoform κ-CN variant A (86-190) presenting 10 deamidations (besides two phosphorylations) was eluted and fragmented at RT 33.39 min (peak 100% 8^+ m/z 1423.558, MW 11380.40). Two other proteoforms were detected at RT 33.45 min (peak 100% m/z 1423.439, MW 11379.45) and 33.63 min (peak 100% m/z 1423.318, MW 11378.48 Da) characterised by two phosphorylations plus 9 deamidations and 8 deamidations, respectively.

Figure III-10 B shows the observed patterns of the 9 to 11 deamidated forms of the considered κ-CN variant A (86-190), and Figure III-10 A shows the corresponding theoretical profiles.

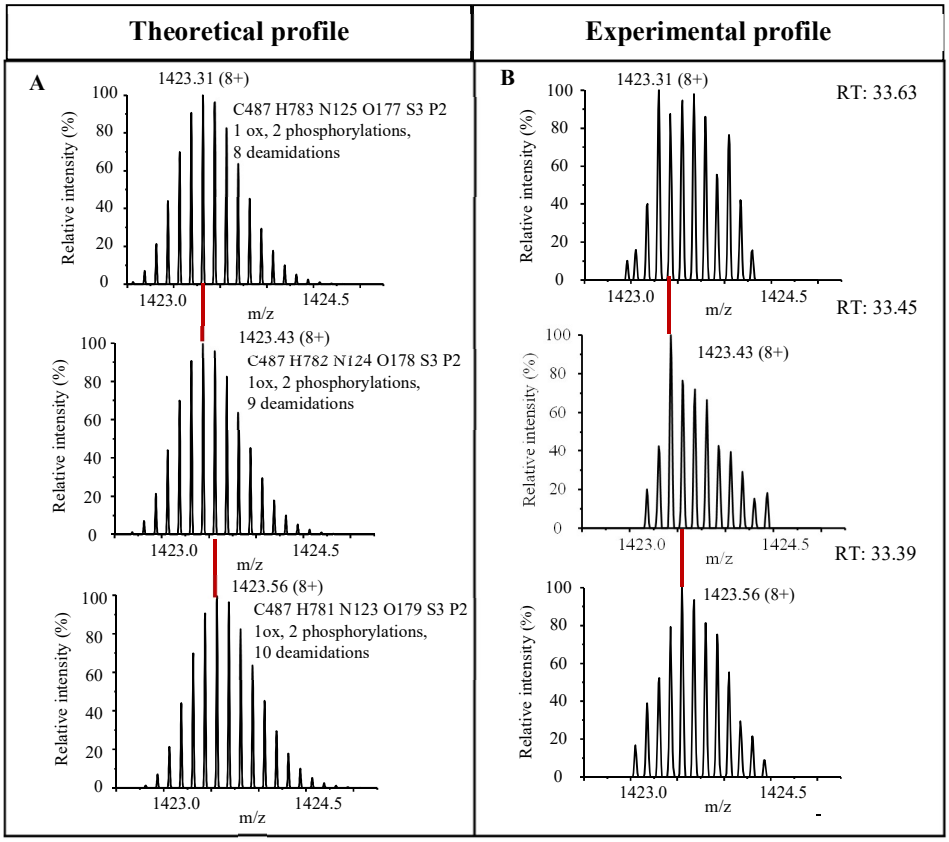
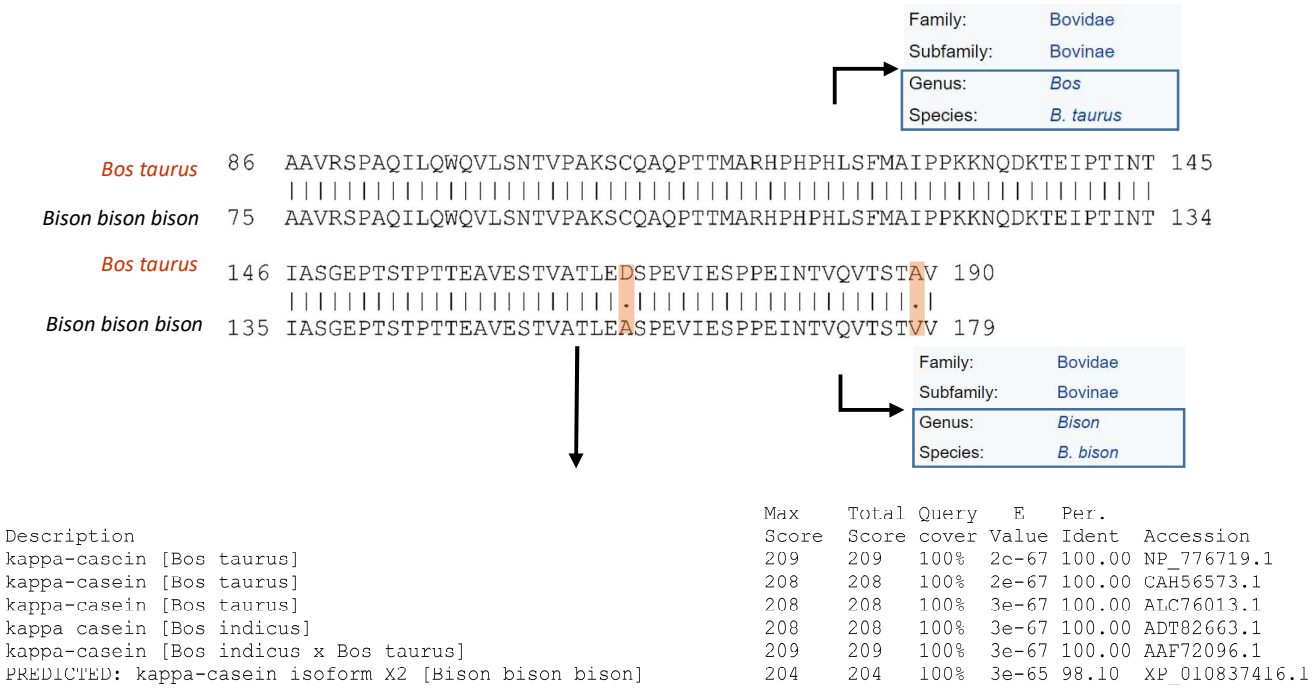


Figure III-10 The isotopic pattern of the 8^+ charged ion of the truncated proteoform κ-CN variant A (86-190) modified by 1 oxidation, 2 phosphorylations and an increasing number of deaminations, eight, nine and ten respectively **(A)** theoretical patterns and **(B)** experimental patterns observed in MS spectra obtained from the analysis of the Gainsborough artwork sample.

A great extent of modifications was also observed for the other cleaved proteoforms such as αS2-CN 56-154, which was detected with two phosphorylations and six deamidations, both in the non-oxidised form (MW 11624.57, Δm= -1.72 ppm) and with one oxidation (MW 11640.56, Δm= -1.72 ppm). For the same protein, two other proteoforms were also identified.

The results achieved from the peptide analysis (Chapter II) suggested that the milk applied by the artist was from *Bovinae* subfamily origin. Nonetheless, the unique peptides detected did not provide a confident distinction between *Bison* and *Bos* Genus. The detection of longer protein forms achieved via top-down approach increased the accuracy of the species identification: the last part of the κ-casein truncated form (138-190), which was not completed in any peptidic analyses, enabled the recognition of two amino acids D169 and A189 distinctive only for *Bos Taurus* (domesticated cattle) specie (Figure III-11).



Description	Max Score	Total Score	Query cover	E Value	Per. Ident	Accession
kappa-cascin [Bos taurus]	209	209	100%	2e-67	100.00	NP_776719.1
kappa-casein [Bos taurus]	208	208	100%	2e-67	100.00	CAH56573.1
kappa-casein [Bos taurus]	208	208	100%	3e-67	100.00	ALC76013.1
kappa casein [Bos indicus]	208	208	100%	3e-67	100.00	ADT82663.1
kappa-casein [Bos indicus x Bos taurus]	209	209	100%	3e-67	100.00	AAF72096.1
PREDICTED: kappa-casein isoform X2 [Bison bison bison]	204	204	100%	3e-65	98.10	XP_010837416.1

Figure III-11 The truncated proteoform 86-190 of the κ-CN variant A provided the discrimination between *B.taurus* and *B.bison* species, which was not achieved with the BUP studies. The alignment of the experimental sequence 86-190 against the whole nrNCBI database via Basic Local Alignment Search Tool (BLAST) showed the unicity of the polypeptide for this species (*Bos taurus indicus* is a subspecies of the *B. taurus* species). Particularly, the two amino acids D169 and A189 in the last segment of the protein enabled two species' distinction.

Table III-5 List of the casein proteoforms identified in Thomas Gainsborough drawing *"Landscape with Horse and Cart Descending a Hill"*. For the proteoforms detected manually (without EthcD fragmentation) is reported the peak m/z and the corresponding observed MW (Da) with a 100% abundance.

Protein	Accession	Position	Theorical MW (Da)	Observed MW (Da)	Error (ppm)	RT	Modifications	Species
α_{S1}-CN	P02662	119-145	3182.56	3182.55	-1.6	21.1	1 phospho (S130)	*B.taurus, B.mutus, B.indicus, B.grunniens, B.b. bison*
		116-151	4199.22	4199.19	-6.9	28.77	1 deam (N129)	*B. taurus, B.b.bison*
β-CN	P02666	114-155	4815.49	4815.50	2.1	33.32	1 deam (Q138)	*B.taurus, B.mutus, B.indicus, B.indicus x B.taurus, B.grunniens, B.b.bison, B.bubalis*
κ-CN	P02668	86-190	11358.45	11358.43	-1.58	33.71	2 phospho (S148-S170), 10 deam (Q93-Q96-Q98-Q110-Q112-N134-Q135-N144-N181-Q184)	*B.taurus, B.indicus, B.indicus x B.taurus*
			11374.44	11374.42	-1.76	33.39	2 phospho, 10 deam, 1ox.	
			11390.44	11390.41	-2.6	33.63	2 phospho, 10 deam, 2ox	
			11373.46 11379.47 (100%)	11379.45 (100%)	1.8	33.45	2 phospho 9 deam, 1ox	
			11372.47 11378.49 (100%)	11378.48 (100%)	-0.44	33.63	2 phospho, 8 deam, 1ox	
		127-190	6799.33	6799.33	0.1	27.7	1 ox (M127)	*B.taurus, B.indicus, B.indicus x B.taurus*
α_{S2}-CN	P02663	28-46	2180.08	2180.07	-0.9	30.69		*B.taurus, B.mutus, B.indicus x B.taurus, B.grunniens, B.b.bison*
		56 - 154	11624.59	11624.57	-1.72	33.59	2 phospho (S144-S146), 6 deam (Q94-N98-N101-N102-N130-N139)	*B.taurus, B.mutus, B.b.bison, B.bubalis*
			11640.58	11640.56	-1.72	33.55	2 phospho, 6 deam, 1 ox	
		120-153	4051.98	4051.97	-2.2	30.85	2 phospho (S146-S150)	*B.taurus, B.mutus, B.indicus x B.taurus, B.grunniens, B.b.bison, B.bubalis*

8. Conclusion

A top-down proteomics methodology capable of separating and characterising proteoforms, starting from very low sample amounts, was successfully developed and applied to historic samples. The presented protocol consists of an adapted filter-based sample preparation combined with protein separation using a monolithic column and high-resolution protein analysis and fragmentation using EThcD. The protocol developed on modern casein standards was successively tested on ancient casein-based glue samples, proving the presence of several truncated casein proteoforms with age-related modifications, such as oxidation and deamidation. The method was finally applied to the study of Gainsborough's drawing *"Landscape with Horse and Cart Descending a Hill",* whose sample was collected with a minimally-invasive sampling, providing for the first time the detection of non-digested proteins from an incredibly minute amount of analyte.

The identification of various casein proteoforms, both hydrolysed and highly modified, suggested that the artist used milk as pigment fixative during the drawing (1780) production. These results confirmed the artist writings and the data obtained through the bottom-up approach.

The top-down study provided new insights into the biological species of milk and the degradation of the sample, complementing the peptidic-analysis results. Bottom-up and top-down proteomics are complementary approaches: currently, in proteomics applied to the cultural heritage, the method with peptidic digestion is more robust and sensitive, leading to more proteins; whey and MGMF milk proteins were scarcely detected in the sample under analysis.

Nonetheless, the data achieved have proven that the top-down approach can disclose missing information; at this stage, top-down can bring added value to data formerly investigated with bottom-up methodology. Several improvements are still required, mostly in the sensitivity and automatisation of data elaboration. Nonetheless, shortly TDP will offer a suitable strategy for a more in-depth investigation of protein and protein modifications in artistic and archaeological artefacts.

Reference

1. Catherman, A.D., O.S. Skinner ,N.L. Kelleher, *Top down proteomics: facts and perspectives.* Biochemical and biophysical research communications, 2014. **445**(4): p. 683-693.
2. Chait, B.T., *Mass spectrometry: bottom-up or top-down?* Science, 2006. **314**(5796): p. 65-66.
3. Chen, B., K.A. Brown, Z. Lin ,Y. Ge, *Top-down proteomics: ready for prime time?* Analytical chemistry, 2017. **90**(1): p. 110-127.
4. Yates, J.R., C.I. Ruse ,A. Nakorchevsky, *Proteomics by mass spectrometry: approaches, advances, and applications.* Annual review of biomedical engineering, 2009. **11**: p. 49-79.
5. Zabrouskov, V., M.W. Senko, Y. Du, R.D. Leduc ,N.L. Kelleher, *New and automated MSn approaches for top-down identification of modified proteins.* Journal of the American Society for Mass Spectrometry, 2005. **16**(12): p. 2027-2038.
6. Fornelli, L., T.K. Toby, L.F. Schachner, P.F. Doubleday, K. Srzentić, C.J. DeHart ,N.L. Kelleher, *Top-down proteomics: where we are, where we are going?* Journal of proteomics, 2018. **175**: p. 3.
7. Smith, L.M., N.L. Kelleher, M. Linial, D. Goodlett, P. Langridge-Smith, Y.A. Goo, G. Safford, L. Bonilla, G. Kruppa ,R. Zubarev, *Proteoform: a single term describing protein complexity.* Nature methods, 2013. **10**(3): p. 186-187.
8. Yang, X., J. Coulombe-Huntington, S. Kang, G.M. Sheynkman, T. Hao, A. Richardson, S. Sun, F. Yang, Y.A. Shen ,R.R. Murray, *Widespread expansion of protein interaction capabilities by alternative splicing.* Cell, 2016. **164**(4): p. 805-817.
9. Capriotti, A.L., C. Cavaliere, P. Foglia, R. Samperi ,A. Laganà, *Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics.* Journal of Chromatography A, 2011. **1218**(49): p. 8760-8776.
10. Toby, T.K., L. Fornelli ,N.L. Kelleher, *Progress in top-down proteomics and the analysis of proteoforms.* Annual review of analytical chemistry, 2016. **9**: p. 499-519.
11. Donnelly, D.P., C.M. Rawlins, C.J. DeHart, L. Fornelli, L.F. Schachner, Z. Lin, J.L. Lippens, K.C. Aluri, R. Sarin ,B. Chen, *Best practices and benchmarks for intact protein analysis for top-down mass spectrometry.* Nature methods, 2019. **16**(7): p. 587-594.
12. Franz, T. ,X. Li, *Step-by-step preparation of proteins for mass spectrometric analysis*, in *Proteomic Profiling*. 2015, Springer. p. 235-247.
13. Cristobal, A., F. Marino, H. Post, H.W. van den Toorn, S. Mohammed ,A.J. Heck, *Toward an optimized workflow for middle-down proteomics.* Analytical chemistry, 2017. **89**(6): p. 3318-3325.
14. Kachuk, C. ,A.A. Doucette, *The benefits (and misfortunes) of SDS in top-down proteomics.* Journal of Proteomics, 2018. **175**: p. 75-86.
15. Henderson, L., S. Oroszlan ,W. Konigsberg, *A micromethod for complete removal of dodecyl sulfate from proteins by ion-pair extraction.* Analytical biochemistry, 1979. **93**: p. 153-157.
16. Doucette, A. ,A. Crowell, *Precipitation of Detergent-Containing Samples for Top-Down and Bottom-Up Proteomics*, in *Proteomics Technologies and Applications*. 2019, IntechOpen.
17. Botelho, D., M.J. Wall, D.B. Vieira, S. Fitzsimmons, F. Liu ,A. Doucette, *Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation.* Journal of proteome research, 2010. **9**(6): p. 2863-2870.

18. Puchades, M., A. Westman, K. Blennow ,P. Davidsson, *Removal of sodium dodecyl sulfate from protein samples prior to matrix-assisted laser desorption/ionization mass spectrometry.* Rapid communications in mass spectrometry, 1999. **13**(5): p. 344-349.
19. Wiśniewski, J.R., A. Zougman, N. Nagaraj ,M. Mann, *Universal sample preparation method for proteome analysis.* Nature methods, 2009. **6**(5): p. 359-362.
20. Wiśniewski, J., *Filter-aided sample preparation: the versatile and efficient method for proteomic analysis*, in *Methods in enzymology*. 2017, Elsevier. p. 15-27.
21. Kachuk, C., M. Faulkner, F. Liu ,A.A. Doucette, *Automated SDS depletion for mass spectrometry of intact membrane proteins though transmembrane electrophoresis.* Journal of proteome research, 2016. **15**(8): p. 2634-2642.
22. Ghezellou, P., V. Garikapati, S.M. Kazemi, K. Strupat, A. Ghassempour ,B. Spengler, *A perspective view of top-down proteomics in snake venom research.* Rapid Communications in Mass Spectrometry, 2019. **33**: p. 20-27.
23. Valeja, S.G., L. Xiu, Z.R. Gregorich, H. Guner, S. Jin ,Y. Ge, *Three dimensional liquid chromatography coupling ion exchange chromatography/hydrophobic interaction chromatography/reverse phase chromatography for effective protein separation in top-down proteomics.* Analytical chemistry, 2015. **87**(10): p. 5363-5371.
24. Cui, W., H.W. Rohrs ,M.L. Gross, *Top-down mass spectrometry: recent developments, applications and perspectives.* Analyst, 2011. **136**(19): p. 3854-3864.
25. Gargano, A.F., L.S. Roca, R.T. Fellers, M. Bocxe, E. Domínguez-Vega ,G.W. Somsen, *Capillary HILIC-MS: a new tool for sensitive top-down proteomics.* Analytical chemistry, 2018. **90**(11): p. 6601-6609.
26. Padula, M.P., I.J. Berry, B. Raymond, J. Santos ,S.P. Djordjevic, *A comprehensive guide for performing sample preparation and top-down protein analysis.* Proteomes, 2017. **5**(2): p. 11.
27. Shen, Y., N. Tolić, P.D. Piehowski, A.K. Shukla, S. Kim, R. Zhao, Y. Qu, E. Robinson, R.D. Smith ,L. Paša-Tolić, *High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics.* Journal of Chromatography A, 2017. **1498**: p. 99-110.
28. Tran, J.C. ,A.A. Doucette, *Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation.* Analytical chemistry, 2008. **80**(5): p. 1568-1573.
29. Wang, L., T. Bo, Z. Zhang, G. Wang, W. Tong ,D. Da Yong Chen, *High resolution capillary isoelectric focusing mass spectrometry analysis of peptides, proteins, and monoclonal antibodies with a flow-through microvial interface.* Analytical chemistry, 2018. **90**(15): p. 9495-9503.
30. Patrie, S.M., *Top-down mass spectrometry: proteomics to Proteoforms*, in *Modern Proteomics–Sample Preparation, Analysis and Practical Applications*. 2016, Springer. p. 171-200.
31. Zheng, S., C. Yoo, N. Delmotte, F.R. Miller, C.G. Huber ,D.M. Lubman, *Monolithic column HPLC separation of intact proteins analyzed by LC-MALDI using on-plate digestion: An approach to integrate protein separation and identification.* Analytical chemistry, 2006. **78**(14): p. 5198-5204.
32. Wu, M., R.a. Wu, Z. Zhang ,H. Zou, *Preparation and application of organic-silica hybrid monolithic capillary columns.* Electrophoresis, 2011. **32**(1): p. 105-115.
33. Eeltink, S., S. Wouters, J.L. Dores-Sousa ,F. Svec, *Advances in organic polymer-based monolithic column technology for high-resolution liquid chromatography-mass spectrometry profiling of antibodies, intact proteins, oligonucleotides, and peptides.* Journal of Chromatography A, 2017. **1498**: p. 8-21.
34. Liang, Y., Y. Jin, Z. Wu, T. Tucholski, K.A. Brown, L. Zhang, Y. Zhang ,Y. Ge, *Bridged Hybrid Monolithic Column Coupled to High-Resolution Mass Spectrometry for Top-Down Proteomics.* Analytical chemistry, 2019. **91**(3): p. 1743-1747.

35. Díaz-Bao, M., R. Barreiro, J.M. Miranda, A. Cepeda ,P. Regal, *Recent advances and uses of monolithic columns for the analysis of residues and contaminants in food.* Chromatography, 2015. **2**(1): p. 79-95.
36. Svec, F., *Less common applications of monoliths: Preconcentration and solid-phase extraction.* Journal of chromatography B, 2006. **841**(1-2): p. 52-64.
37. Kelleher, N.L., *Peer reviewed: Top-down proteomics*. 2004, ACS Publications.
38. Spraggins, J.M., D.G. Rizzo, J.L. Moore, K.L. Rose, N.D. Hammer, E.P. Skaar ,R.M. Caprioli, *MALDI FTICR IMS of intact proteins: using mass accuracy to link protein images with proteomics data.* Journal of the American Society for Mass Spectrometry, 2015. **26**(6): p. 974-985.
39. Schwamborn, K. ,R.M. Caprioli, *MALDI imaging mass spectrometry–painting molecular pictures.* Molecular oncology, 2010. **4**(6): p. 529-538.
40. Wu, S., H. Ma, I. Prytkova, D. Stenoien ,L. Paša-Tolić, *Proteomics, Top-Down.* 2017.
41. Dittwald, P., D. Valkenborg, J. Claesen, A.L. Rockwood ,A. Gambin, *On the fine isotopic distribution and limits to resolution in mass spectrometry.* Journal of The American Society for Mass Spectrometry, 2015. **26**(10): p. 1732-1745.
42. Bogdanov, B. ,R.D. Smith, *Proteomics by FTICR mass spectrometry: top down and bottom up.* Mass spectrometry reviews, 2005. **24**(2): p. 168-200.
43. Michalski, A., E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Müller, R. Viner, J. Schwartz, P. Remes ,M. Belford, *Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes.* Molecular & Cellular Proteomics, 2012. **11**(3).
44. Denisov, E., E. Damoc, O. Lange ,A. Makarov, *Orbitrap mass spectrometry with resolving powers above 1,000,000.* International Journal of Mass Spectrometry, 2012. **325**: p. 80-85.
45. Fornelli, L., K.R. Durbin, R.T. Fellers, B.P. Early, J.B. Greer, R.D. LeDuc, P.D. Compton ,N.L. Kelleher, *Advancing top-down analysis of the human proteome using a benchtop quadrupole-orbitrap mass spectrometer.* Journal of proteome research, 2017. **16**(2): p. 609-618.
46. Scheltema, R.A., J.-P. Hauschild, O. Lange, D. Hornburg, E. Denisov, E. Damoc, A. Kuehn, A. Makarov ,M. Mann, *The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer.* Molecular & Cellular Proteomics, 2014. **13**(12): p. 3698-3708.
47. Erickson, B.K., M.P. Jedrychowski, G.C. McAlister, R.A. Everley, R. Kunz ,S.P. Gygi, *Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer.* Analytical chemistry, 2015. **87**(2): p. 1241-1249.
48. Bourgoin-Voillard, S., N. Leymarie ,C.E. Costello, *Top-down tandem mass spectrometry on RN ase A and B using a Q h/FT-ICR hybrid mass spectrometer.* Proteomics, 2014. **14**(10): p. 1174-1184.
49. Tucholski, T. ,Y. Ge, *Fourier-transform ion cyclotron resonance mass spectrometry for characterizing proteoforms.* Mass spectrometry reviews, 2020.
50. Olsen, J.V., J.C. Schwartz, J. Griep-Raming, M.L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor ,M. Splendore, *A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed.* Molecular & cellular proteomics, 2009. **8**(12): p. 2759-2769.
51. Cleland, T.P., C.J. DeHart, R.T. Fellers, A.J. VanNispen, J.B. Greer, R.D. LeDuc, W.R. Parker, P.M. Thomas, N.L. Kelleher ,J.S. Brodbelt, *High-throughput analysis of intact human proteins using UVPD and HCD on an orbitrap mass spectrometer.* Journal of proteome research, 2017. **16**(5): p. 2072-2079.

52. Olsen, J.V., B. Macek, O. Lange, A. Makarov, S. Horning ,M. Mann, *Higher-energy C-trap dissociation for peptide modification analysis.* Nature methods, 2007. **4**(9): p. 709-712.
53. Gregorich, Z.R. ,Y. Ge, *Top-down proteomics in health and disease: Challenges and opportunities.* Proteomics, 2014. **14**(10): p. 1195-1210.
54. Zubarev, R.A., N.L. Kelleher ,F.W. McLafferty, *Electron capture dissociation of multiply charged protein cations. A nonergodic process.* Journal of the American Chemical Society, 1998. **120**(13): p. 3265-3266.
55. Zubarev, R.A., D.M. Horn, E.K. Fridriksson, N.L. Kelleher, N.A. Kruger, M.A. Lewis, B.K. Carpenter ,F.W. McLafferty, *Electron capture dissociation for structural characterization of multiply charged protein cations.* Analytical chemistry, 2000. **72**(3): p. 563-573.
56. Tian, Z., N. Tolić, R. Zhao, R.J. Moore, S.M. Hengel, E.W. Robinson, D.L. Stenoien, S. Wu, R.D. Smith ,L. Paša-Tolić, *Enhanced top-down characterization of histone post-translational modifications.* Genome biology, 2012. **13**(10): p. R86.
57. Fornelli, L., E. Damoc, P.M. Thomas, N.L. Kelleher, K. Aizikov, E. Denisov, A. Makarov ,Y.O. Tsybin, *Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS.* Molecular & Cellular Proteomics, 2012. **11**(12): p. 1758-1767.
58. Iavarone, A.T. ,E.R. Williams, *Collisionally activated dissociation of supercharged proteins formed by electrospray ionization.* Analytical chemistry, 2003. **75**(17): p. 4525-4533.
59. Floris, F., L. Chiron, A.M. Lynch, M.P. Barrow, M.-A. Delsuc ,P.B. O'Connor, *Application of tandem two-dimensional mass spectrometry for top-down deep sequencing of calmodulin.* Journal of The American Society for Mass Spectrometry, 2018. **29**(8): p. 1700-1705.
60. Cannon, J.R., M.B. Cammarata, S.A. Robotham, V.C. Cotham, J.B. Shaw, R.T. Fellers, B.P. Early, P.M. Thomas, N.L. Kelleher ,J.S. Brodbelt, *Ultraviolet photodissociation for characterization of whole proteins on a chromatographic time scale.* Analytical chemistry, 2014. **86**(4): p. 2185-2192.
61. Riley, N.M., J.W. Sikora, H.S. Seckler, J.B. Greer, R.T. Fellers, R.D. LeDuc, M.S. Westphall, P.M. Thomas, N.L. Kelleher ,J.J. Coon, *The value of activated ion electron transfer dissociation for high-throughput top-down characterization of intact proteins.* Analytical chemistry, 2018. **90**(14): p. 8553-8560.
62. Riley, N.M., M.S. Westphall ,J.J. Coon, *Activated ion-electron transfer dissociation enables comprehensive top-down protein fragmentation.* Journal of proteome research, 2017. **16**(7): p. 2653-2659.
63. Ledvina, A.R., G.C. McAlister, M.W. Gardner, S.I. Smith, J.A. Madsen, J.C. Schwartz, G.C. Stafford Jr, J.E. Syka, J.S. Brodbelt ,J.J. Coon, *Infrared photoactivation reduces peptide folding and hydrogen-atom migration following ETD tandem mass spectrometry.* Angewandte Chemie International Edition, 2009. **48**(45): p. 8526-8528.
64. Park, J., P.D. Piehowski, C. Wilkins, M. Zhou, J. Mendoza, G.M. Fujimoto, B.C. Gibbons, J.B. Shaw, Y. Shen ,A.K. Shukla, *Informed-Proteomics: open-source software package for top-down proteomics.* Nature methods, 2017. **14**(9): p. 909-914.
65. Durbin, K.R., J.C. Tran, L. Zamdborg, S.M. Sweet, A.D. Catherman, J.E. Lee, M. Li, J.F. Kellie ,N.L. Kelleher, *Intact mass detection, interpretation, and visualization to automate Top-Down proteomics on a large scale.* Proteomics, 2010. **10**(20): p. 3589-3597.
66. Horn, D.M., R.A. Zubarev ,F.W. McLafferty, *Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules.* Journal of the American Society for Mass Spectrometry, 2000. **11**(4): p. 320-332.

67. Liu, X., Y. Inbar, P.C. Dorrestein, C. Wynne, N. Edwards, P. Souda, J.P. Whitelegge, V. Bafna ,P.A. Pevzner, *Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach.* Molecular & Cellular Proteomics, 2010. **9**(12): p. 2772-2782.
68. Zamdborg, L., R.D. LeDuc, K.J. Glowacz, Y.-B. Kim, V. Viswanathan, I.T. Spaulding, B.P. Early, E.J. Bluhm, S. Babai ,N.L. Kelleher, *ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry.* Nucleic acids research, 2007. **35**(suppl_2): p. W701-W706.
69. D. LeDuc, R. ,N. L. Kelleher, *Using ProSight PTM and related tools for targeted protein identification and characterization with high mass accuracy tandem MS data.* Current protocols in bioinformatics, 2007. **19**(1): p. 13.6. 1-13.6. 28.
70. Viner, R., S. Sharma, J.D. Canterbury, D.M. Horn ,V. Zabrouskov, *Optimization of LC/MS Intact/Top-Down Protein Analysis on an Orbitrap Fusion Mass Spectrometer.*
71. DeHart, C.J., R.T. Fellers, L. Fornelli, N.L. Kelleher ,P.M. Thomas, *Bioinformatics analysis of top-down mass spectrometry data with ProSight Lite*, in *Protein Bioinformatics*. 2017, Springer. p. 381-394.
72. Fellers, R.T., J.B. Greer, B.P. Early, X. Yu, R.D. LeDuc, N.L. Kelleher ,P.M. Thomas, *ProSight Lite: graphical software to analyze top-down mass spectrometry data.* Proteomics, 2015. **15**(7): p. 1235-1238.
73. Karabacak, N.M., L. Li, A. Tiwari, L.J. Hayward, P. Hong, M.L. Easterling ,J.N. Agar, *Sensitive and specific identification of wild type and variant proteins from 8 to 669 kDa using top-down mass spectrometry.* Molecular & Cellular Proteomics, 2009. **8**(4): p. 846-856.
74. Sirotkin, Y., X. Liu, Y. Shen, G. Anderson, Y.S. Tsai, Y.S. Ting, D.R. Goodlett, R.D. Smith, V. Bafna ,P.A. Pevzner, *Visualization of the MS-Align algorithm results for the protein spectrum matches.* Department of Bioengineering and Bioinformatics of MV Lomonosov Moscow State University, 2011: p. 347.
75. Liu, X., Y. Sirotkin, Y. Shen, G. Anderson, Y.S. Tsai, Y.S. Ting, D.R. Goodlett, R.D. Smith, V. Bafna ,P.A. Pevzner, *Protein identification using top-down spectra.* Molecular & cellular proteomics, 2012. **11**(6).
76. Kou, Q., L. Xun ,X. Liu, *TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization.* Bioinformatics, 2016. **32**(22): p. 3495-3497.
77. Tsai, Y.S., A. Scherl, J.L. Shaw, C.L. MacKay, S.A. Shaffer, P.R. Langridge-Smith ,D.R. Goodlett, *Precursor ion independent algorithm for top-down shotgun proteomics.* Journal of the American Society for Mass Spectrometry, 2009. **20**(11): p. 2154-2166.
78. Cai, W., H. Guner, Z.R. Gregorich, A.J. Chen, S. Ayaz-Guner, Y. Peng, S.G. Valeja, X. Liu ,Y. Ge, *MASH Suite Pro: a comprehensive software tool for top-down proteomics.* Molecular & Cellular Proteomics, 2016. **15**(2): p. 703-714.
79. Kim, S. ,P.A. Pevzner, *MS-GF+ makes progress towards a universal database search tool for proteomics.* Nature communications, 2014. **5**: p. 5277.
80. Dallongeville, S., N. Garnier, C. Rolando ,C. Tokarski, *Proteins in art, archaeology, and paleontology: from detection to identification.* Chemical reviews, 2016. **116**(1): p. 2-79.
81. Tholey, A. ,A. Becker, *Top-down proteomics for the analysis of proteolytic events-Methods, applications and perspectives.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2017. **1864**(11): p. 2191-2199.
82. Sinitcyn, P., J.D. Rudolph ,J. Cox, *Computational methods for understanding mass spectrometry–based shotgun proteomics data.* Annual Review of Biomedical Data Science, 2018. **1**: p. 207-234.

83. Cai, W., T. Tucholski, B. Chen, A.J. Alpert, S. McIlwain, T. Kohmoto, S. Jin ,Y. Ge, *Top-down proteomics of large proteins up to 223 kDa enabled by serial size exclusion chromatography strategy.* Analytical chemistry, 2017. **89**(10): p. 5467-5475.
84. Grupe, G., *Preservation of collagen in bone from dry, sandy soil.* Journal of Archaeological Science, 1995. **22**(2): p. 193-199.
85. Grupe, G. ,S. Turban-Just, *Amino acid composition of degraded matrix collagen from archaeological human bone.* Anthropologischer Anzeiger, 1998: p. 213-226.
86. Tuross, N., *Albumin preservation in the Taima-taima mastodon skeleton.* Applied Geochemistry, 1989. **4**(3): p. 255-259.
87. Embery, G., A. Milner, R.J. Waddington, R.C. Hall, M.S. Langley ,A.M. Milan, *The isolation and detection of non-collagenous proteins from the compact bone of the dinosaur Iguanodon.* Connective tissue research, 2000. **41**(3): p. 249-259.
88. Ostrom, P.H., M. Schall, H. Gandhi, T.-L. Shen, P.V. Hauschka, J.R. Strahler ,D.A. Gage, *New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry.* Geochimica et Cosmochimica Acta, 2000. **64**(6): p. 1043-1050.
89. Nielsen-Marsh, C.M., P.H. Ostrom, H. Gandhi, B. Shapiro, A. Cooper, P.V. Hauschka ,M.J. Collins, *Sequence preservation of osteocalcin protein and mitochondrial DNA in bison bones older than 55 ka.* Geology, 2002. **30**(12): p. 1099-1102.
90. Ostrom, P.H., H. Gandhi, J.R. Strahler, A.K. Walker, P.C. Andrews, J. Leykam, T.W. Stafford, R.L. Kelly, D.N. Walker ,M. Buckley, *Unraveling the sequence and structure of the protein osteocalcin from a 42 ka fossil horse.* Geochimica et Cosmochimica Acta, 2006. **70**(8): p. 2034-2044.
91. Humpula, J.F., P.H. Ostrom, H. Gandhi, J.R. Strahler, A.K. Walker, T.W. Stafford Jr, J.J. Smith, M.R. Voorhies, R.G. Corner ,P.C. Andrews, *Investigation of the protein osteocalcin of Camelops hesternus: Sequence, structure and phylogenetic implications.* Geochimica et Cosmochimica Acta, 2007. **71**(24): p. 5956-5967.
92. Welker, F., J. Ramos-Madrigal, M. Kuhlwilm, W. Liao, P. Gutenbrunner, M. de Manuel, D. Samodova, M. Mackie, M.E. Allentoft ,A.-M. Bacon, *Enamel proteome shows that Gigantopithecus was an early diverging pongine.* Nature, 2019. **576**(7786): p. 262-265.
93. Cappellini, E., F. Welker, L. Pandolfi, J. Ramos-Madrigal, D. Samodova, P.L. Rüther, A.K. Fotakis, D. Lyon, J.V. Moreno-Mayar ,M. Bukhsianidze, *Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny.* Nature, 2019. **574**(7776): p. 103-107.
94. Tokarski, C., C. Cren-Olive, E. Martin ,C. Rolando, *Proteomics approach for binding media studies in art painting.* Non-Destructive Testing and Microanalysis for the Diagnostics and Conservation of the Cultural and Environmental Heritage, Antwerp, Belgium, 2002.
95. Tokarski, C., C. Cren-Olive, C. Rolando ,E. Martin, *Protein studies in cultural heritage.* Molecular biology and cultural heritage: Swets & Zeitlinger, 2003: p. 119-130.
96. CASCO Gluing Guide, C.C.o.A., Division of the Borden Company, ed. *CASCO gluing guide*. 3rd edition ed. 1938.
97. Kirby, D.P., A. Manick ,R. Newman, *Minimally Invasive Sampling of Surface Coatings for Protein Identification by Peptide Mass Fingerprinting: A Case Study with Photographs.* Journal of the American Institute for Conservation, 2019: p. 1-11.
98. Erde, J., R.R.O. Loo ,J.A. Loo, *Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments.* Journal of proteome research, 2014. **13**(4): p. 1885-1895.

99. Pierri, G., D. Kotoni, P. Simone, C. Villani, G. Pepe, P. Campiglia, P. Dugo ,F. Gasparrini, *Analysis of bovine milk caseins on organic monolithic columns: An integrated capillary liquid chromatography–high resolution mass spectrometry approach for the study of time-dependent casein degradation.* Journal of Chromatography A, 2013. **1313**: p. 259-269.

Chapter IV Protein Structural Analysis in artworks by Mass Spectrometry: Hydrogen/Deuterium Exchange and Cross-linking investigation

Abstract

This chapter outlines the studies performed to improve the comprehension of proteins' structural and conformational alterations in proteinaceous binder applied in tempera paintings. Two innovative techniques not yet used in cultural heritage studies were conducted by combining mass spectrometry with hydrogen/deuterium exchange HDX-MS (implementing an intact protein analysis mode) and cross-linking investigations (bottom-up approach). The studies were performed on the lysozyme model system to simplify the complexity of egg protein composition. HDX-MS studies aimed to inquire about the role of some of the most common inorganic pigments (lead white/zinc white/cinnabar/ red lead) on proteins' molecular and structural changes during formulation before the protein ageing. The dynamics of the isotopic exchange in unpigmented lysozyme was compared with the results of lysozyme-based painting models to get an insight into the protein structural differences.

Complementary information on proteins' ageing and oxidation degradations were achieved by performing cross-linking investigations with MS analysis. A new strategy based on the elaboration of data from the bottom-up studies was established to investigate the protein network formations through cross-linked peptidic pairs' localisation and characterisation. The methodology was initially tested on mock-ups of paintings (formulated with lysozyme mixed with lead white pigment) treated with oxidising agents or naturally aged. The detection of different reticulated formations led to defining various molecular patterns characteristic of oxidative-based cross-links in lysozyme, subsequently detected in more complex samples from historical tempera paintings. This new molecular information provides new insights into the study of proteins and their networks in historic artworks.

1. Introduction

The current challenge in cultural heritage studies is being driven towards a greater understanding of the alterations of organic compounds within the sample matrix with considerations on its environment, ageing or restorative treatments. So far, the research mainly focused on oily media [1-10] or lipid in egg tempera [11-14], while few studies inquired about the protein structural modifications within paint layers. Specifically, little has been understood on the formation of aggregation or networks in proteins consequent to interactions between proteins/proteins, amongst various substances present in the artwork (e.g. different organic molecules or pigments) or later, during ageing.

However, meaningful information on the alterations of the proteins' secondary structure in painting models has been already achieved by performing the thermogravimetric analysis (TG), Differential Scanning Calorimetry (DSC) and Fourier Transform Infrared Spectroscopy (FTIR) [15-18]. Investigations on the most common proteinaceous compounds (collagen of animal glue, casein and ovalbumin) have proven that pigments react with proteins, modifying the molecular conformation directly (e.g. with the establishment of metal complexes) or indirectly, with the promotion of oxidative stress. The reduction of thermal stability observed in the fresh painting models was explained with possible pigment intercalations in β-sheets or α-helices, as well as to the partial breakage of interprotein interactions (with a decrease of α-helices and decrease of β-sheets) [17].

During the ageing, proteinaceous binders undergo further alterations, including modifications of amino acid residues (like oxidations, deamidations) and partial hydrolysis of the proteins [19-21]. Besides the stable complexes with pigments, during ageing, proteins can form other aggregations (through hydrophobic, electrostatic, van der Waals forces) and establish strong covalent bonds (cross-linking formation), which cause an increase in thermal stability [18, 22-25]. These processes can reduce the solubility of protein fractions, decreasing the proteolytic susceptibility and hindering the analysis. An in-depth comprehension of the mechanism and products of aggregation and cross-linking is critical to address those limitations, providing a proper and representative investigation of the proteinaceous compounds in ancient artworks.

2. Aim of the research

The research's principal aims were the investigation of the protein's structure and its modifications at the conformational level in artworks, with a particular focus on tempera egg-based paintings at different stages (from the paint formulation to its ageing). The structural study was divided into two parts where two innovative techniques, not yet performed in the field of cultural heritage, were applied to gain information on the protein modifications subsequent to the paint formulation with different pigments and induced by oxidative reactions and natural ageing.

1. The interactions of different pigments with lysozyme protein were inquired through Hydrogen/Deuterium exchange (HDX) studies by mass spectrometry. The isotopic exchanges of the intact protein structure of unpigmented lysozyme were compared with the dynamics of lysozyme mixed with four of the most commonly implemented pigments (lead white, zinc white, cinnabar and red lead). The objective was to investigate their potential structural differences in freshly made paint formulations.

2. The second objective was to develop a strategy to detect the cross-links established in ancient and degraded proteinaceous compounds. The study aimed to optimise the data elaboration of bottom-up ESI-MS/MS analyses to provide the identification and localisation of distinctive cross-linked peptidic pairs. The achievement of molecular fingerprints characteristic of lysozyme networks in aged and oxidised materials was envisaged.

The studies were performed on lysozyme, which is a short and well structurally-characterised model protein. The purpose was to simplify the egg tempera system by studying only one egg white protein. The sequence has 147 amino acids, of which the first eighteen correspond to the pre-peptide; its molecular formula is $C_{613}H_{959}N_{193}O_{185}S_{10}$ with a monoisotopic molecular weight of 14303.8 Da. The native protein presents four intramolecular disulfide bonds: two weaker (C24-C145, C48-C133) and two stronger (C82-C98 and C94-C112) [21, 26].

3. Hydrogen deuterium exchange combined with mass spectrometry (HDX-MS)

Hydrogen deuterium exchange combined with mass spectrometry (HDX-MS) has its bases on Linderstrøm-Lang and colleagues' studies at the laboratories of Carlsberg (Copenhagen, Danemark). The researchers demonstrated a measurable correlation between the differences in the kinetics of HDX in protein's backbone amide group and protein dynamics in conformations [27-29] as well as the solvent accessibility [30]. HDX-MS is principally considered a comparative analysis that investigates the differences between two similar systems or the variations in the same system. Its main applications are in the validation of biosimilars and drugs and the investigation of protein denaturation, folding, ligands, and protein-protein interactions [31]. The deuterium labelling is commonly performed incubating the protein in a deuterated solvent (D2O) at variable labelling time (so-called on-exchange strategy) followed by the examination and calculation of the deuterium incorporation at each time [32]. The incubation provokes an isotope exchange in any site of the amide backbone and amino acid side chain having labile hydrogen (-OH, -SH, -NH).

However, only the amide hydrogen exchanges have kinetic rates compatible with the duration of the HX-MS study. The other kinetics of the side chain H, being too fast, are lost during the analysis (Figure IV-1).

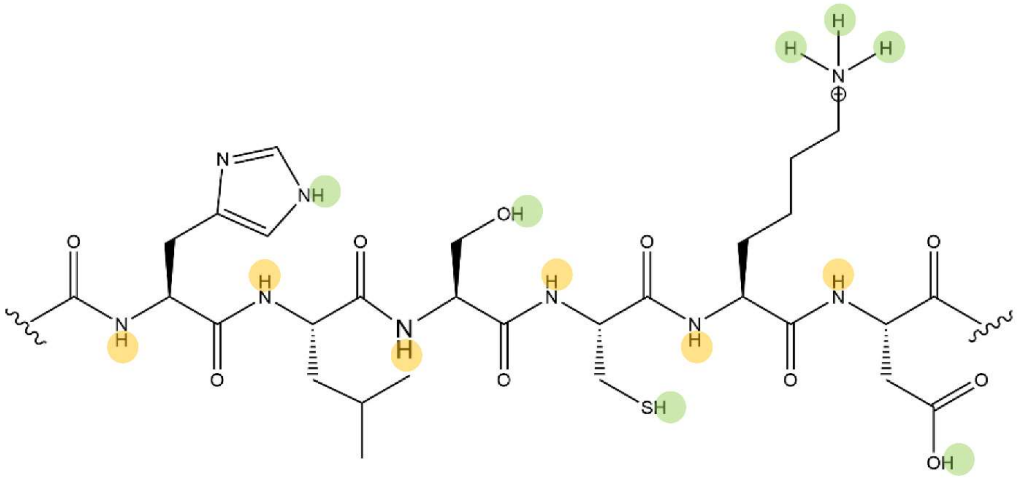


Figure IV-1 Amino acid sequence constituted by H-L-S-C-K-D. The backbone amide hydrogens are highlighted in yellow; their kinetics rates of H/D exchanges are compatible with the HX-MS analysis duration. The less stable side-chain protons are highlighted in light green. Hydrogen-carbon atoms are not showed since not exchangeable. Figure elaborated from Oganesyan et al., 2018 using ChemDraw Prime.

The deuterium uptake in amide hydrogens is highly representative of even minor structure changes [33]. The reasons behind this are the even distribution of hydrogens along the protein chain and the formation of primary H bonds within and between the protein's secondary structural elements [34, 35].

A less frequent alternative is the off-exchange strategy consisting of an initial protein preparation in deuterated water to exchange all amide hydrogens to deuterium, followed by dilution in a non-deuterated buffer the measurement of the loss of deuterium at different time intervals [36]. This strategy is limited by the high time some protons require to initially exchange, which can happen after years [37].

The labelling can be performed following two methods, depending on the research objectives and the proteinaceous compound investigated. The first is the continuous labelling, which leads to an investigation of protein conformational fluctuations and a comparison between the different dynamics of the same protein in various conditions (for example, in the presence of pigments). The method consists of dividing the sample into aliquots with identical conditions and their exposure to D2O buffer for different durations (e.g. 1 min, 5 min, 10min) [38]. Alternatively, the pulse labelling leads to depict the protein variations from one state to another due to a triggering event, as the denaturation. In this case, protein samples are collected at different time intervals (min/hours/days) from the stock protein, which is subject to the trigger (for example, a continuous unfolding process). Each sample is then incubated to D2O for a specific time duration and quenched. The results provide a set of data for each conformational and dynamic moment of the protein in that condition [36].

Regardless of the approach selected, samples should be prepared and labelled, controlling with cautions the pH and temperature to standardise the HDX rate and thus the analysis. Following the exposure, the protein solution is quenched to minimise the kinetics of exchanges (acid and base catalysed processes) by decreasing the temperature to 0-3°C and the pH to 2.5 – 2.8. Commonly during the quenching, a denaturant (e.g. guanidinium chloride) is added. The proteins then can be digested through a protease active in acid conditions, like pepsin (bottom-up approach), or they can be analysed intact, with or without fragmentation in the so-called global or native strategy. The digestion step is commonly performed online in an LC pepsin column, while the intact protein can be either directly introduced to MS or desalted through an LC-separation step, on condition that the system preserves the quenching conditions to prevent the back exchange [32, 38].

Labelling, digestion and separation steps in an HDX workflow have been recently fully automated and standardised by developing robotic liquid autosampler systems.

The automation provides a more reliable reproduction of the preparation steps, improving the precision of the analysis duration, the amount of transferred liquid, and the temperature and pH control. Eluted analytes are then analysed using the mass spectrometer providing the isotopic profiles, which will be gathered from the other measurements and compared [39].

In general, the intact analysis (also called the global approach) identifies the global structural variations over the entire protein while the bottom-up approach provides local information of the deuterium exchange dynamics [40]. The global approach is particularly useful in providing a first insight into protein aggregation and to compare biosimilar. Nonetheless, a disadvantage results in the risk of no detect any change in the deuterium absorption due to an equal decrease and increase of exchanges in two different areas of the proteins. Contrarily, the peptide analysis enables the localisation of the area, presenting the difference in exchange, providing more detailed knowledge of protein structure [41].

4. Hydrogen/Deuterium exchange (HDX) studies via mass spectrometry of painting models formulated with lysozyme and different pigments

4.1 Methodology

4.1.1 Samples

The studies were performed on painting mock-ups prepared by mixing 100 µL of a solution of lysozyme 10% (w/v) (protein purchased from Sigma-Aldrich, L6876) with a different pigment (100 mg): lead-white pigment ($2PbCO_3$ $Pb(OH)_2$); zinc white (ZnO), cinnabar (HgS), and red lead/minium (Pb_3O_4), purchased from Kremer (Figure IV-2). The paint mixture was drawn up on inert glass slides and dried in the open air.

Lysozyme + lead white $(PbCO_3)_2 \cdot Pb(OH)_2$ | **Lysozyme + zinc white (ZnO)** | **Lysozyme + cinnabar (HgS)** | **Lysozyme + red lead/minium (Pb_3O_4)**

Figure IV-2 Painting models formulated by mixing lysozyme with four different pigments often implemented in artistic painting formulations.

4.1.2 Sample preparation

The proteins were extracted from the different mock-up (2mg) by adding 200 µl of 0.1% formic acid (AF) to the powder, crushing vehemently with a pestle designed for the Eppendorf 1.5 mL, and leaving stirring overnight. The next day, the solutions were centrifuged to remove the solid particles, then evaporated and re-diluted with 50 µl of water to obtain the sample's desired concentration; the final dilution was in deionised water.

All samples, including the lysozyme standard, were then subjected to two-cycles of filter cleaning using Amicon 10kDa with 10-fold water dilution.

4.1.3 HDX-MS analysis

In this study, the isotopic exchange was performed following the on-exchange strategy with continuous labelling. The whole procedure of H/D exchange was conducted on an HDx-3 PAL™ automated platform (Trajan, former Leap) coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo). During the HDX-MS experiment, the stock sample was kept at 0°C, while D_2O and H_2O stocks were stored at 20°C. The labelling times were set at 0s (no-exchange), 120s, 900s and 3600s; three and four replicas were conducted for each time-point for the lysozyme sample without and with pigment, respectively. The HDX exchange process was initiated by incubating the sample (10 µM) at 20°C in deuterated water D_2O with a ratio protein/buffer of 1:9 (2µL of the sample in 18µL of D_2O). The un-exchanged sample of t0 sec was incubated in a water buffer. Quenching of H/D isotopic exchange reaction was achieved by adding 30 µL of pre-chilled (at 0°C) quenching solution containing 1.5 M guanidinium hydrochloride, 100 mM TCEP, 0.8 % FA pH 3 and mixing.

The sample was then immediately injected into the automated LC-system maintained at 0 °C in the cold box of the automate. The quenched protein solution was loaded into a Pepmap 300 C18 trap column (ThermoFisher Scientific) and subsequently separated by a C18 column (1.9 µm Hypersil Gold, 1 mm × 50 mm, ThermoFisher Scientific) with a linear gradient of 13% to 35% buffer B over 10 min (buffer A, 0.1% FA in water; buffer B, 95% acetonitrile, 0.1% FA in water; flow rate 40 µL/min) in elution configuration. The trap and analytical column were washed with 100% buffer B for 15 min. The column was then re-equilibrated to 13% buffer B for 10 min. The gradient was intentionally left short to avoid the risk of back-exchange of the proteins. MS analyses allowing mass protein determination were carried out with an LTQ Orbitrap XL mass spectrometer (ThermoFisher Scientific). The mass spectrometer operated in positive ion mode at a 4.10-kV needle voltage. The settings used for mass spectrometry (MS) analysis were: resolution 60,000; AGC target 5e5; scan range m/z 375–2000. Data were analysed using Qual browser and Xtract Xcalibur software (ThermoFisher Scientific). The result for each time point was represented as an average of three (lysozyme standard) or four (painting mock-ups) replicate measurements.

Proteoforms eluted at each labelling time (for all samples) were singularly investigated by a manual calculation of the average mass of the majority peaks (at least two) and then averaging between the replica results.

All calculated intact proteoform masses were further confirmed through the automated deconvolution. This approach was not chosen as the only strategy (despite being very advantageous in terms of speed) because the presence of poorly resolved peaks would have led to an incorrect estimation of the precursor mass.

The data obtained at each labelling moment for lysozyme with and without pigments were then compared to investigate whether the samples' exchange rate was equivalent.

4.2 Analysis and discussion of results

The protein structures of lysozyme in the presence and absence of four pigments were compared through the HDX-MS strategy to investigate the potential structural differences induced in the proteins by the presence of various metal ions.

The first positive results regarded the remarkable resemblance and superimposition of the different replica spectra obtained for each exchange times (0, 120 s, 900 s, 3600 s). These findings assured the analysis robustness and reproducibility, and the stability of lysozyme structure (with or without the ligand) for all experiment duration, proving that the continuous labelling was the proper strategy to apply. Both unpigmented lysozyme and lead white/lysozyme system were analysed multiple times, repeating the experiments on different days: the similarity observed among the results for the runs of the same sample strengthened the strategy's robustness. Some replica runs did not succeed in providing any signal, probably due to a local technical issue. Nonetheless, the decision of analysing three replicas for lysozyme and four replicas for each lysozyme/pigment ensured for the majority of points at least two replicates of the sample.

4.2.1 Mass spectrometry results at t0 (no incubation in deuterated water)

Initial observations on the intact proteins were achieved from the classical mass spectrometry analysis without deuterated water. In our experimental conditions, all lysozyme samples showed two prominent peaks of elution at two different times (about 16.15 and 16.80 min) (Figure IV-3).

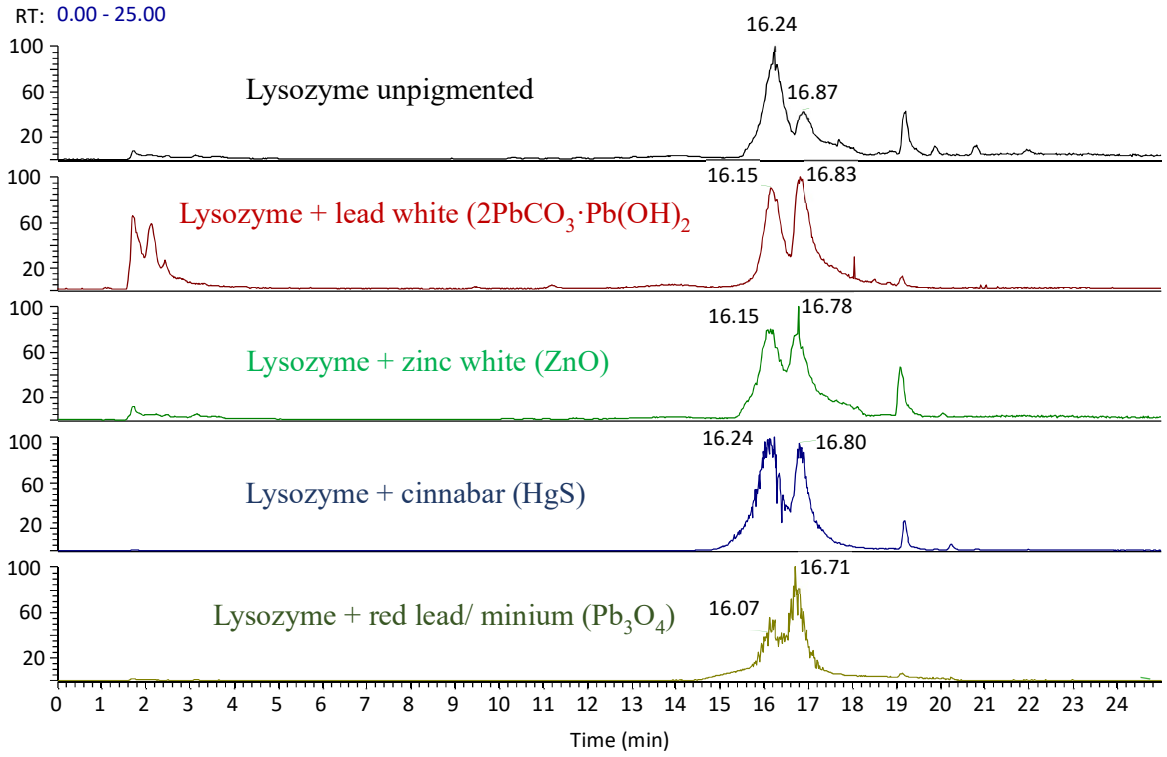


Figure IV-3 Total ion current (TIC) chromatograms of lysozyme unpigmented and painting models formulated with four pigments widely implemented in artistic productions: lead white, zinc white, cinnabar and red lead. All samples showed two main peaks of elutions.

In all samples, the predominant proteoform firstly eluted presented patterns corresponding to 15^+ to 8^+ charges with an average mass of 14304.8 Da. The value reported was calculated from the 11^+ charged ion m/z 1301.44 (peak at 100% abundance measured from the unpigmented lysozyme) (Figure IV-4 on the top). This value was attributed to the lysozyme proteoform presenting four disulphide bonds [21]

Interestingly, considering the first main elution peak at RT 16.00-16.25 min in the unpigmented lysozyme, only one form was detected while different oxidation forms (with an average of three forms) were identified in the four painting models, with the greatest extent observed in red lead/lysozyme (Figure IV-4 on the bottom).

ELUTION MAIN PEAK 1

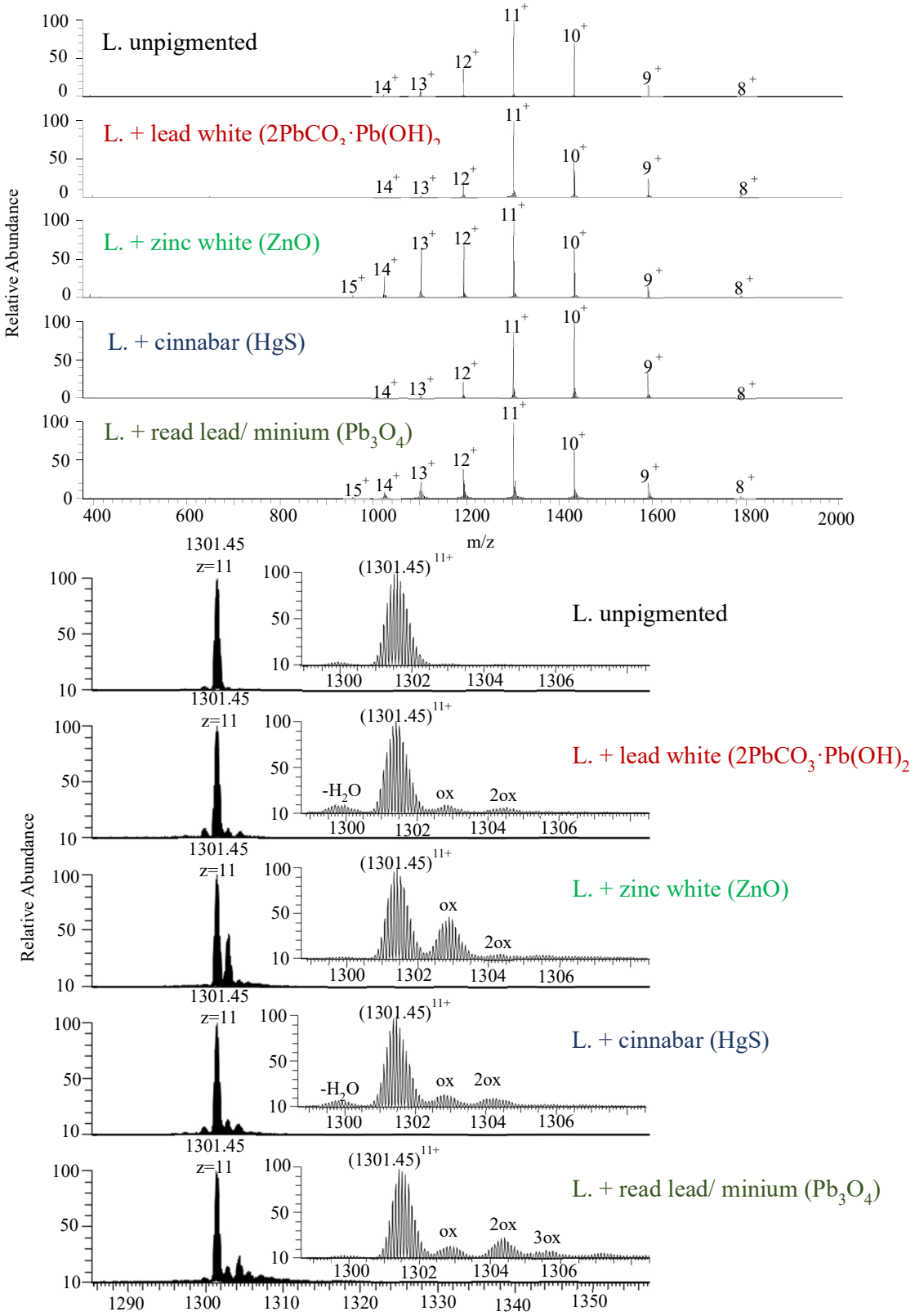


Figure IV-4 MS spectra from the first main peak of elution in all samples under investigation at time zero (no deuterium exchange). The complete massive charge of the predominant proteoform is shown on the top, while on the bottom are represented two different magnifications of the 100% abundance peak. Various oxidation forms were detected in the painting models.

Successively in the gradient (2nd main eluted peak at RT 16.60-16.90 min), in all samples, a second lysozyme proteoform eluted with a distinct ionisation pattern (Figure IV-5 on the top). In most of the samples, the pattern's main peak was evidence at charge 13^+ (m/z 1101.53); only in lysozyme/ lead white, the principal peak was observed at charge 12^+ (m/z 1193.24). The experimental masses calculated for all samples were equivalent to an average mass of 14306.8 Da corresponding to the lysozyme with one disulphide bond cleaved (-2.033Da).

A significant extent of oxidations lysozyme proteoforms was highlighted in all painting models, with the attribution of at least four oxidation forms in the red lead sample, three in cinnabar and zinc white samples, and two in lysozyme with lead white pigment. Moreover, in the model with zinc white pigment, the most abundant peak was attributed to the di-oxidised lysozyme form (Figure IV-5 on the bottom).

These findings significantly suggested that in the presence of pigments, proteins were subjected to elevated oxidative stress. The observations agreed with previous studies reporting oxidations of side chains in proteinaceous based paints in which pigments like cinnabar were identified as a sensitiser of oxidising stress [18, 42, 43]. Other metal ions from pigments such as Pb or Fe are known to catalyse oxidation reactions [16] Furthermore, the numerous sulfur-containing amino acids (cysteine, cystine and methionine) and residues with aromatic side chains (histidine, tryptophan and tyrosine) in the lysozyme sequence make the protein particularly vulnerable to oxidation stress.

ELUTION MAIN PEAK 2

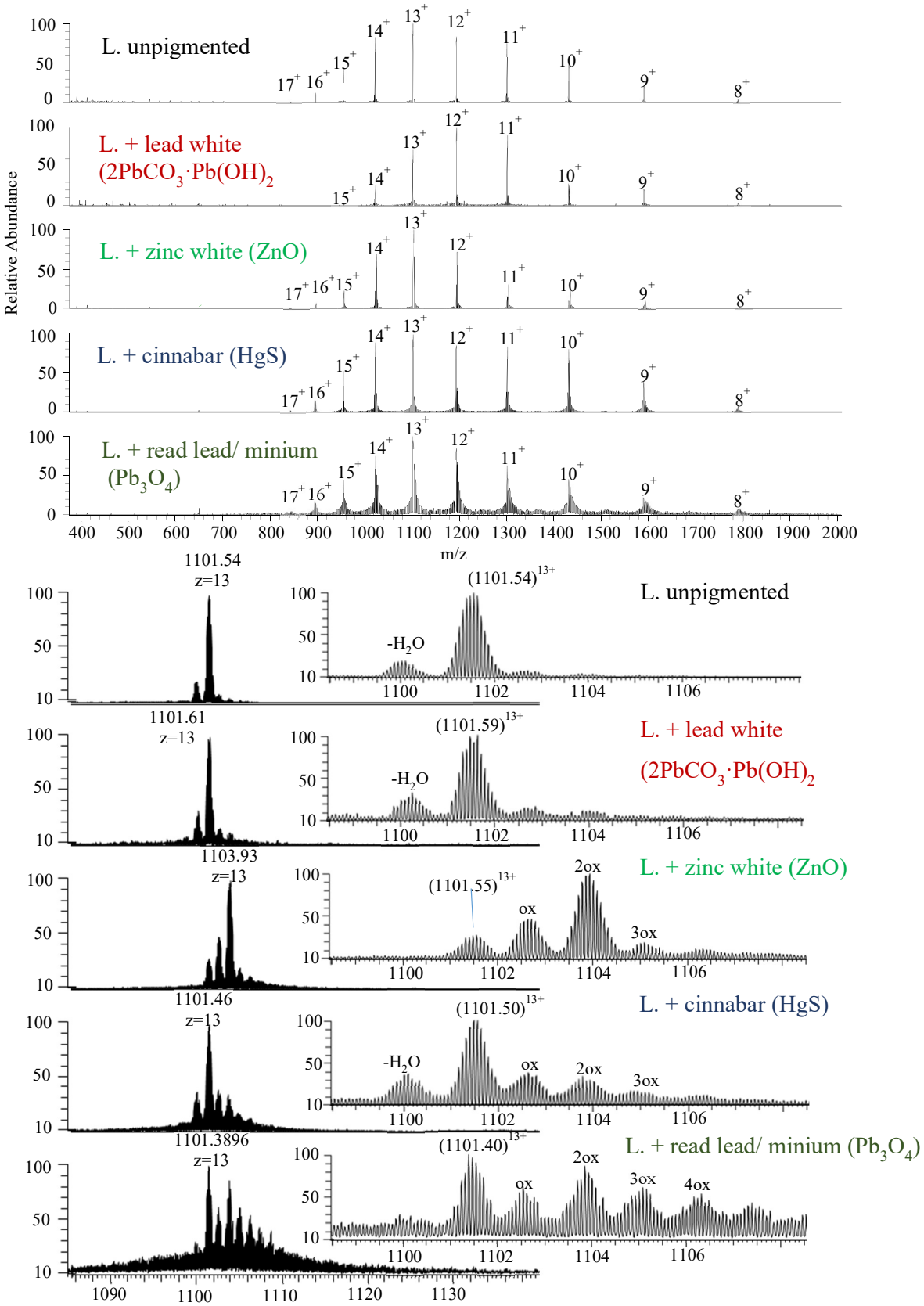


Figure IV-5 MS spectra second main peak of elution in all samples under investigation at time zero (no deuterium exchange). The predominant proteoform's multiply charged patterns are shown on the top, while on the bottom are represented two different magnifications of the 100%abundance peak. Various oxidation forms were detected in all painting models.

4.2.2 Mass spectrometry results following deuterium incubation at an increasing rate

In all investigated samples, a proper isotopic exchange was recorded for all eluted proteoforms from the first analysis (at 120 s) with a gradual increase of exchange rate correlated to the rise of incubation time (900 s and 3600 s). The study, although, did not concern the increment of isotopic exchange in the single protein over time but the comparison of the outcomes observed between the unpigmented lysozyme and each pigmented model.

For example, in the painting model lysozyme/lead white, a variance in the exchange rate was detected already at short incubation times (120 s and 900 s), although the most remarkable effect was highlighted after 3600 seconds (Figure IV-6). After one hour of incubation, the mass ranges eluted in each of the two peaks of elutions differed between unpigmented lysozyme and lysozyme/lead white. Specifically, the difference of average mass for the proteform detected in the first elution peak between lysozyme without (14323.3 Da) and with lead white pigment (14320.9 Da) showed the exchange of 3 hydrogens. (Figure IV-8).

Similarly, the average masses of the proteoforms detected in the second peak of elution differed in the two samples for 3 Da: 14324.6 Da (with a most abundant peak at charge 13^+) was calculated in lysozyme standard and 14321.9 Da (with the main peak of charge 12^+) in lysozyme pigmented (Figure IV-8). These observations suggest that, independently of the cleavage of a disulfide bond, in the unpigmented lysozyme, the overall exchange was of three amide hydrogens more than in lysozyme with lead white pigment.

The reliability of these results and the strategy's validity were proven with the superimpositions of spectra achieved in different replicas (Figure IV-7). The error bars represent the standard deviation calculated on the average of the 3 values (lysozyme unpigmented) or 4 values (painting models). The similarity obtained in two analyses of the sample lysozyme/lead white carried out on different days allowed the strengthening of the accuracy of the findings and the observation of the stability of the new protein structure redefined by the presence of the pigment.

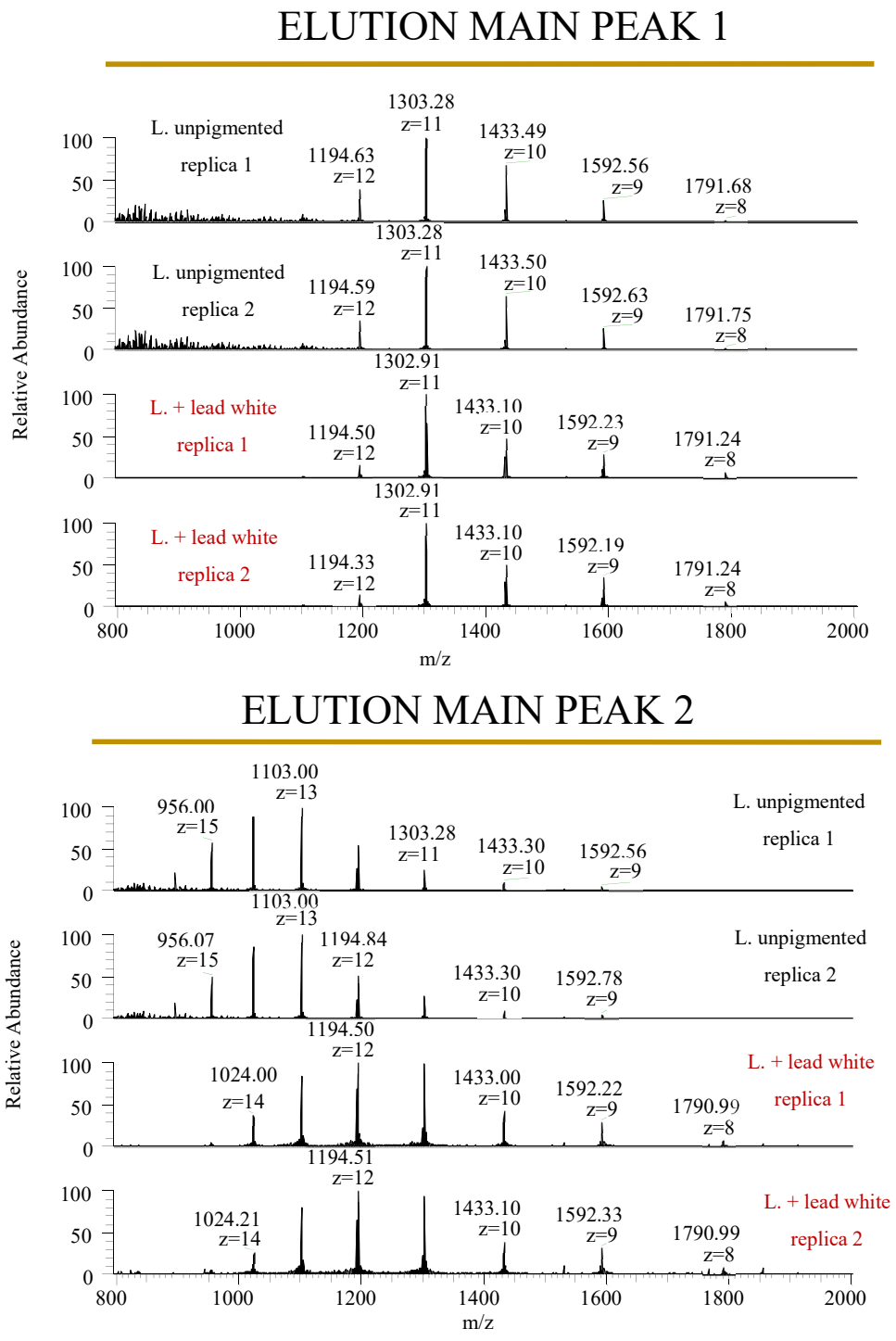


Figure IV-6 MS spectra of the predominants proteoforms of the first and second main elution peaks in lysozyme with and without white lead pigment after 3600 seconds of incubation with the deuterated solution. For each sample, two experimental replicates are shown to prove the repeatability of the experiment.

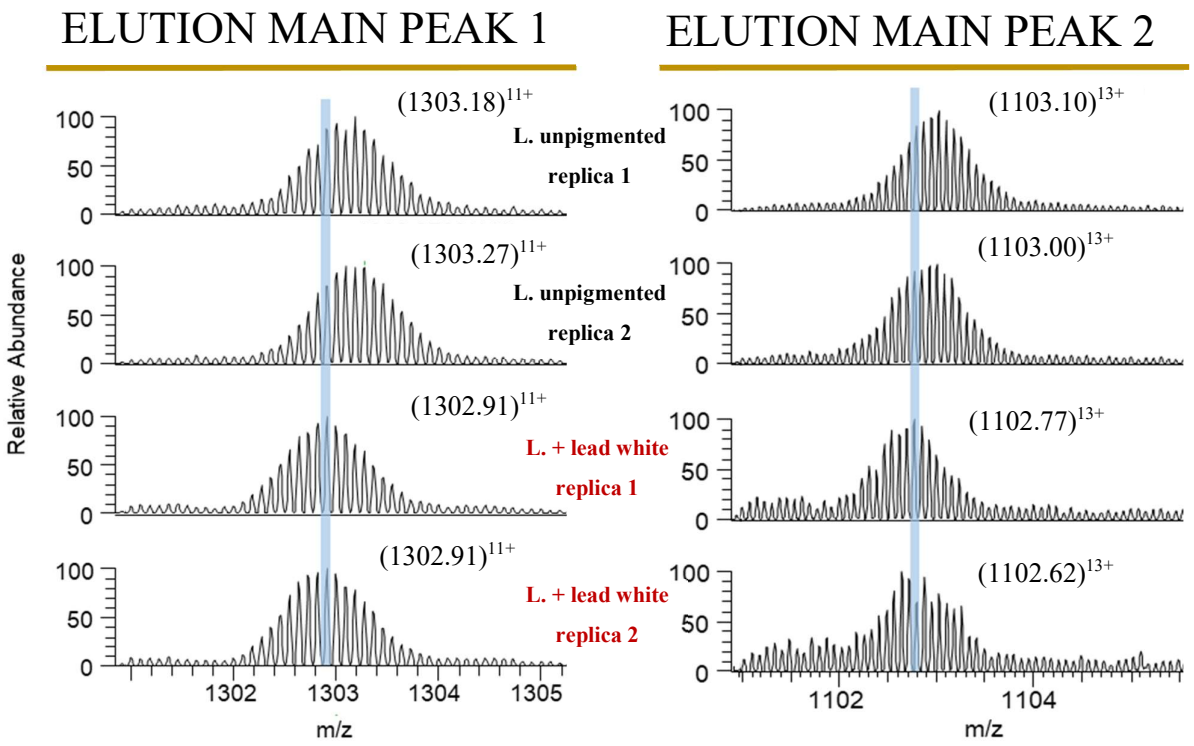


Figure IV-8 Magnification of the peak of 100% detected in lysozyme with and without lead white pigment. The distinction between the average masses of proteoform from both the first and second peak elutions was 3 Da which corresponds to three hydrogens exchange. For each sample, two experimental replicates are shown to prove the repeatability of the experiment.

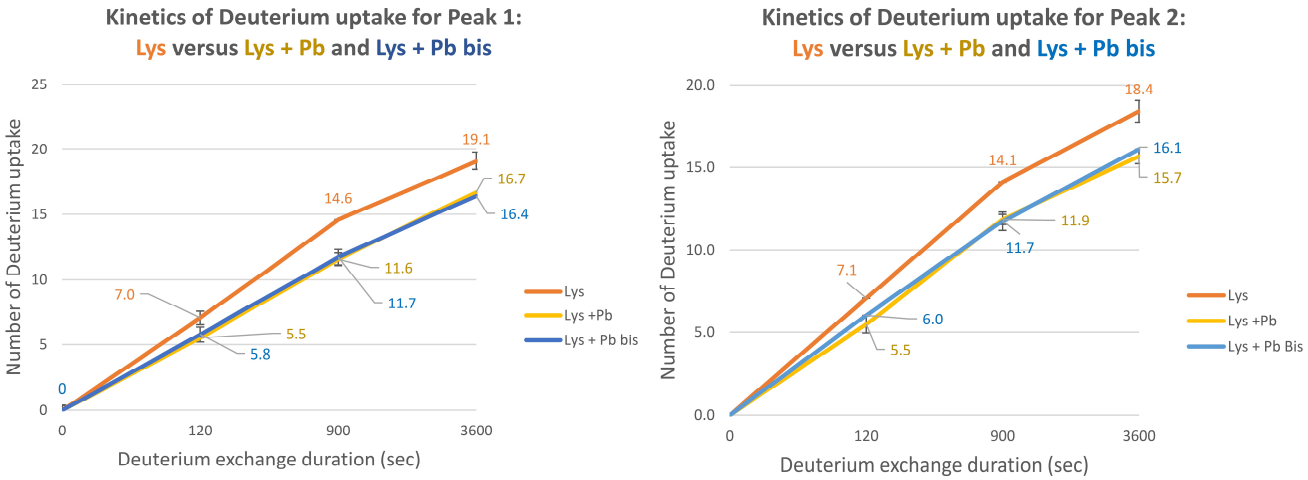


Figure IV-7 Graphs of comparison between the deuterium uptake of unpigmented lysozyme and lead white painting model. The kinetics of deuterium absorption for the proteoforms identified in the first and second elution peaks are shown. Lysozyme is reported in orange, while in grey and yellow are reported two curves corresponding to two experiments run on different days on samples of lysozyme/lead white pigment. The error bars represent the standard deviation calculated on the average of the 3 values (lysozyme unpigmented) or 4 values (lysozyme/lead white pigment).

Different kinetics of deuterium uptake were observed for each one of the other three painting models. In lysozyme mixed with zinc white pigment, no divergences were detected in the overall deuterium exchange regarding the unpigmented lysozyme (Figure IV-9-A).

Contrarily, lysozyme in the presence of cinnabar pigment demonstrated an elevated exchange rate, exhibiting after 3600 seconds a variation of 7 Da in the average mass compared to the unpigmented lysozyme (i.e. both in the first and second peak of elution) (Figure IV-9-B). The complexity and multitude of eluted forms (caused principally by oxidation processes) observed from the red lead pigment model painting led to a problematic interpretation of the results achieved for the deuterium kinetics uptakes at the protein level (further investigation are currently on going on the digested sample).

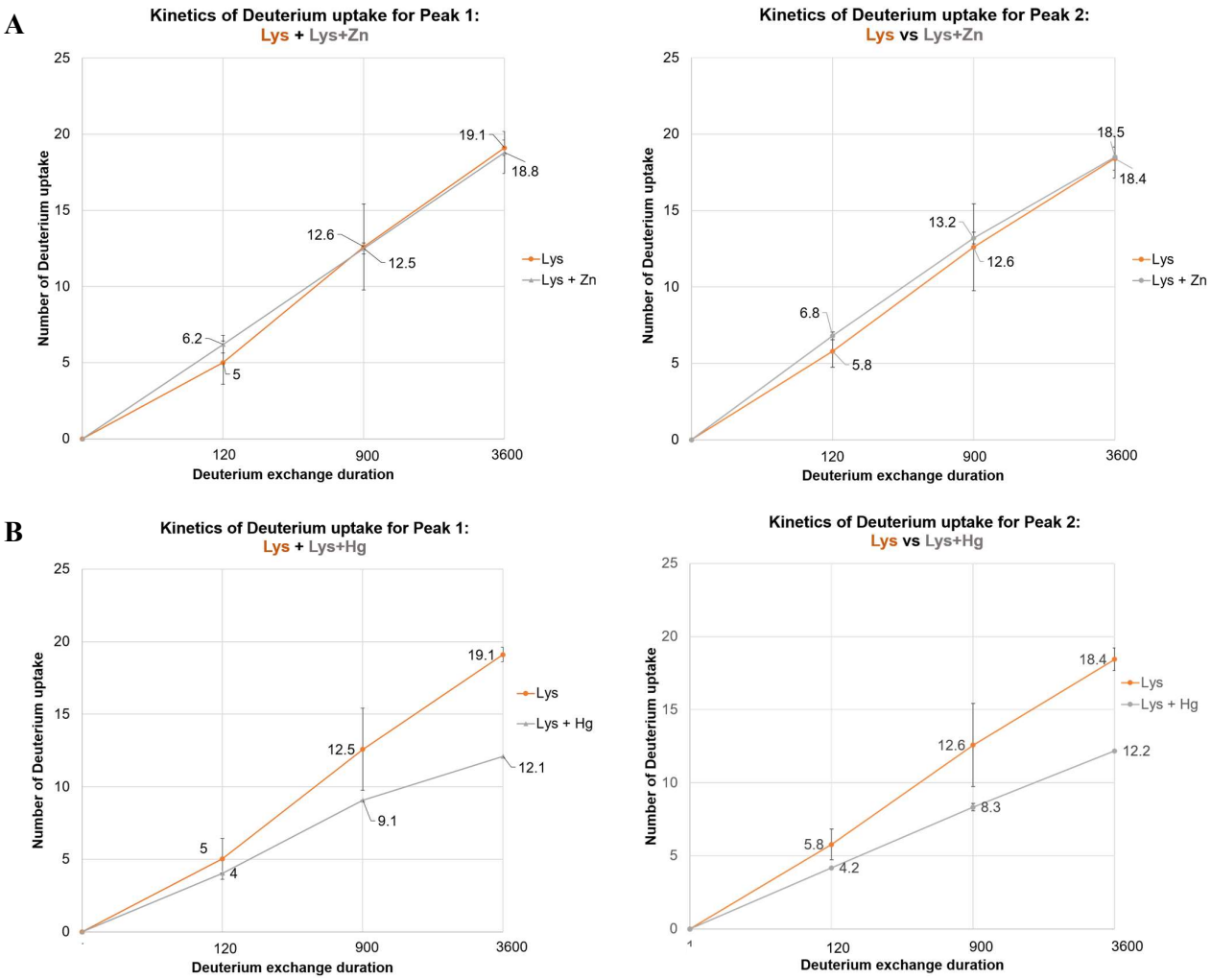


Figure IV-9 Graphs of comparison between the deuterium uptake of unpigmented lysozyme and **(A)** zinc white model painting, **(B)** cinnabar model painting. The kinetics of deuterium absorption for the proteoforms identified in the first and second elution peaks are shown. Lysozyme unpigmented is orange while the painting-mock-ups in grey. The error bars represent the standard deviation calculated on the average of the 3 values (lysozyme unpigmented) or 4 values (painting mock-ups).

The reduction in the exchange rate observed in the painting models, predominantly in lysozyme with lead white and cinnabar, compared to the unpigmented lysozyme might result from various causes: (i) the pigment interposed itself between the amide hydrogen and the solvent, hampering its accessibility, (ii) the pigment locally modified the protein conformation lowering the accessibility of these regions to the solvent, (iii) the simultaneity of both previous effects.

Considering that the lysozyme sample (in different conditions) was confirmed stable throughout the analysis, the implementation of longer labelling incubations (for several hours) will be explored; in fact, it would theoretically increase the observed effect, leading to a more straightforward interpretation of the results, particularly for zinc white.

The global HDX analysis provides an overview of the alteration of the structure. However, if the structure had varied in one region by increasing its number of available amide hydrogens but had an equal number of hydrogen ions been hindered in another area of the same structure, the double effect would have been zeroed. That is the reason why, from this first exploratory study, further bottom-up analyses on the same samples are currently investigated to provide more accurate insight on the events that induced the change of deuterium incorporations leading to a local investigation and localisation of the conformation changes.

5. Investigation of protein cross-linking

5.1 Protein cross-linking

Proteins can undergo cross-linking. The term refers to the establishment of one or more covalent bonds of amino acid residues within the same subunit of a protein complex (intra-protein cross-link) or between different proteins (inter-protein cross-link) with the consequent formation of a protein network [44].

Proteins may carry several modifications besides the cross-links that, following the proteolysis for BUP analysis, result in a complex mixture of (i) non-reticulated/linear peptides, modified or not; (ii) intra peptide cross-links constituted by one peptide having two amino acid residues linked, modified or not; inter peptide cross-links, which can be (iii) cross-links between peptides of the same protein, or (iv) between peptides of different proteins. (Figure IV-10).

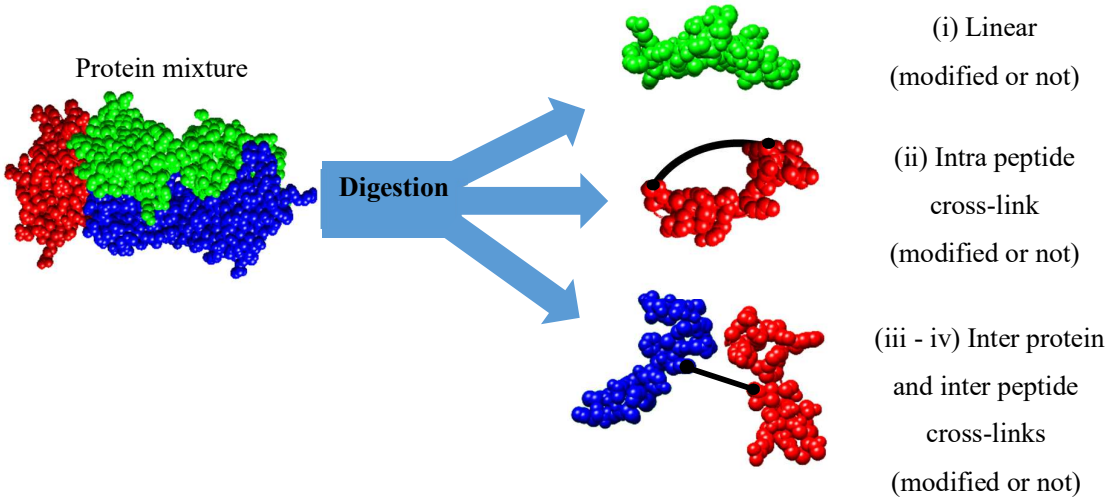


Figure IV-10 Representation of the potential peptidic constituents obtained from the digestion of a cross-linked protein mixture. The illustration depicts the chains A-B and C of hen egg-white lysozyme and has been realised using UCSF Chimera software.

5.2 Principal causes of cross-linking formations

Proteins might undergo cross-linking formation due to different reasons that can be either natural or anthropogenic.

- Physiological processes. Various forms of molecular liaison, including weak interactions (hydrophobic interactions or hydrogen bonds) and strong covalent cross-links, spontaneously form to stabilise the 3D protein structures.

For instance, disulfide bonds are endogenous cross-links that naturally occur in various proteins between two sulfhydryls (SH) side chains of two cysteine residues, stabilising the secondary and tertiary protein structure [25]. Due to their ubiquity and stability, these covalent bonds are probably the most studied and well-known cross-links [45-47].

- Enzymatic activity. Numerous enzymes catalyse cross-linking formations, for example, transglutaminases, peroxidases, transferase or polyphenol oxidases [48].

- Non-enzymatic reactions such as Maillard reactions induced by the interaction with sugars [49, 50].

- Insult-induced cross-linking. Cross-links might be induced by alkaline pH [51, 52], physical/chemical oxidative treatments (e.g., chemical agent or UV radiation [53-56]), or heat exposure[57, 58].

- Intentional introduction of chemical cross-linkers. The implementation of chemical cross-linking is often combined with mass spectrometry analysis to conduct structural proteomics studies (XL-MS) on protein-protein interactions [59-61]. Alternatively, cross-linking is applied in the biomedical field like in fixative and preservative treatments (i.e., Formalin-FixedParaffin-Embedded-FFPE tissues) [62, 63].

- Interactions of metal ions with proteins (e.g. pigments in artistic production). Metals can provoke either protein cleavages or folding, depending on their chemical composition [15, 17, 18, 64-66].

- In painting systems, besides the action of pigments (metal salts), other factors can promote protein networking. For example, various organic or inorganic compounds in the paint system, storage conditions (environmental factors), and products applied during restoration treatments (such as the unintentional introduction of chemical cross-linkers).

Independently from the cause of origin, cross-link formations are principally influenced by the structural closeness of the two amino acid residues and by the steric hindrance that affects the accessibility of amino acid residues to the linker [67].

5.2.1 Protein cross-links induced by oxidative reactions

Different cross-links can be induced by oxidation/photo-oxidation processes [68-70]. Generally, these reticulations are provoked by endogenous or exogenous oxidative factors. For instance, enzymatic actions both direct (such as oxidase and peroxidase) and indirect (formation of intermediate species as dehydroalanine, DHA), or radiations (photo-oxidation) and oxidative reactions with the subsequent production of radicals and ROS (reactive oxidant species) [71-73]. Sulfur-containing residues (such as Cys and Met) and aromatic residues (His, Phe, Tyr, and Trp), being electron-rich sites, are the primary targets of the oxidants, which are electrophilic [74].

Currently, the mechanisms of oxidation-induced cross-links are still not wholly explained. Their identification and localisation are extremely challenging since the cross-linking formation is unpredictable and often in competition with other PTMs, such as oxidations of side chains. Furthermore, this latter's signal often hampers the detection of the reticulation in the analysis [75].

Besides the disulfide bonds, the principal oxidation induced cross-links that have been described to involve tyrosine, tryptophan (Tyr-Tyr; Trp-Trp and Tyr-Trp covalent bonds) and histidine (His-Lys; His-His and His-Cys) amino acid residues.

Tyrosine-Tyrosine cross-links

The di-tyrosine formation is composed of two Tyr-phenoxyl-radicals bonded trough an ortho-ortho di-Tyr cross-link following an oxidative or nitrative stress [72, 73, 76]. During the dimerisation process, the extensive electron delocalisation from the phenoxyl oxygen around the aromatic ring could form two alternative species characterised by carbon-carbon (C3-C3, o, o') bonds or carbon-phenoxyl oxygen bonds (C3-O) [77] (Figure IV-11).

Recent studies suggested that tyrosine and tryptophan secondary products (as quinone arising from 3,4 dihydroxyphenylalanine (DOPA) in Tyr and N-formylkynurenine or kynurenine in Trp) may be involved in cross-linking generation via Michael addition of a nucleophile [70]. The dityrosine formation process is characterised by the loss of two hydrogens which causes a mass variation of - 2.0159 Da in the dimer [78].

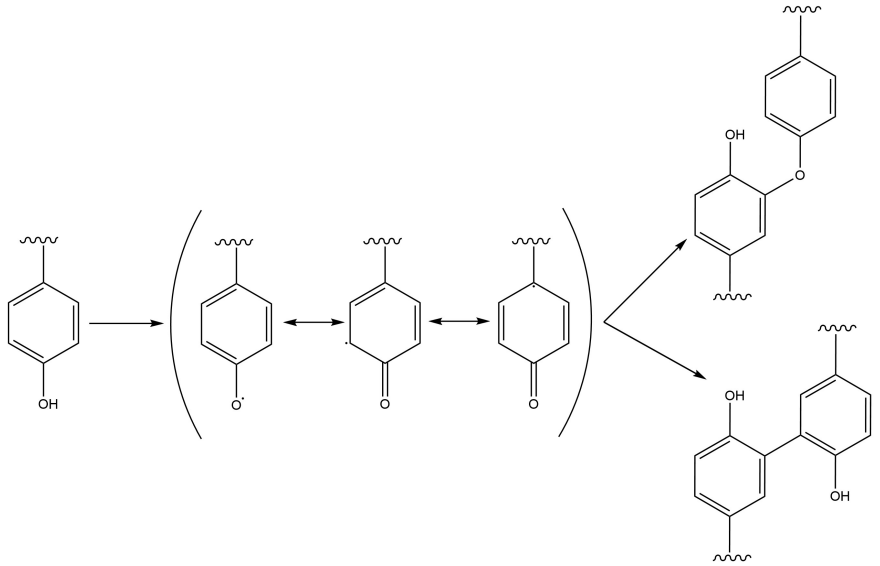


Figure IV-11 Proposed mechanism of Tyr-Tyr cross-links formation (di-tyrosine compound). The bond can involve carbon and oxygen (product at the top) or carbon-carbon (product on the bottom). Figure elaborated from Hägglund et al., 2018 with ChemPrime.

Tryptophan-trypthophan cross-link

Tryptophan is considered the principal oxidation target. In the formation of covalent cross-links involving these residues, the relevance of Trp indolyl radicals (Trp•) formation has clearly ascertained, but their role is still not entirely explained [79]. The mechanism could create either carbon-carbon C3-C3 or carbon-nitrogen C3-N bonds. Multiple stereoisomers can arise from both products due to the extensive spin delocalisation over the indole ring in Trp• followed by covalent bond formation at various sites having high spin density (Figure IV-12) [80-83]. This significant variability of different reticulated products often results in a highly heterogeneous mixture of challenging characterisation.

Likewise, the dityrosine formation, the Trp-Trp bond, causes two hydrogens' loss with a consequent reduction of mass variation of - 2.0159 Da [78]. Analogous mechanisms might prompt covalent bonds between tryptophan and tyrosine [84].

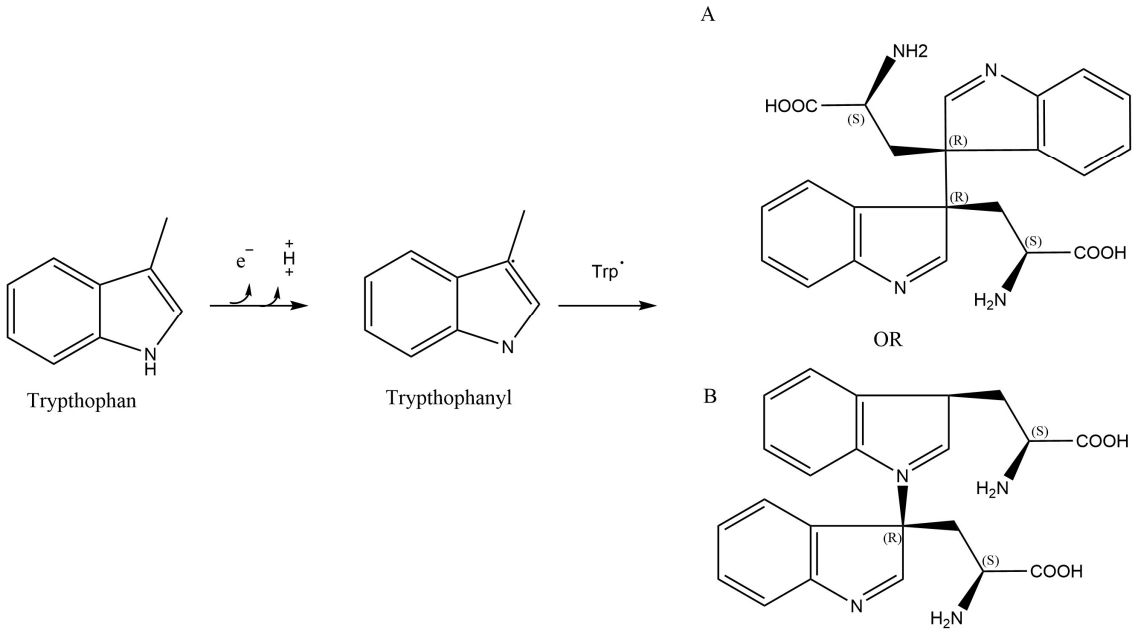


Figure IV-12 Proposed mechanism of Trp-Trp formation (di-tryptophan compound). The bond can involve carbon-carbon (A) or carbon-nitrogen (B). Furthermore, multiple stereoisomers have been observed for both compounds. Figure elaborated from Hägglund et al., 2018 with ChemPrime.

Histidine-lysine cross-links

Histidine-oxidation products with carbonyl functions may form cross-links with another histidine, lysine or cysteine through nucleophile reactions [85] (Figure IV-13). The mechanism is still not entirely understood, but it may involve a reaction of singlet oxygen with histidine to form an endoperoxide intermediate. Being unstable and reactive, this compound turns rapidly into two potential intermediates: 2-oxo-histidine (2-oxo-His with +14 Da) or His+ 32 species (+32 Da). This latter could be involved in the formation of a cross-link via nucleophile attack with different amino acid residues, such as lysine, cysteine or another histidine. The entire cross-linking process is characterised by a mass addition of 13.9840 Da, equivalent to the mass of an oxygen minus two hydrogen masses [70, 85-87].

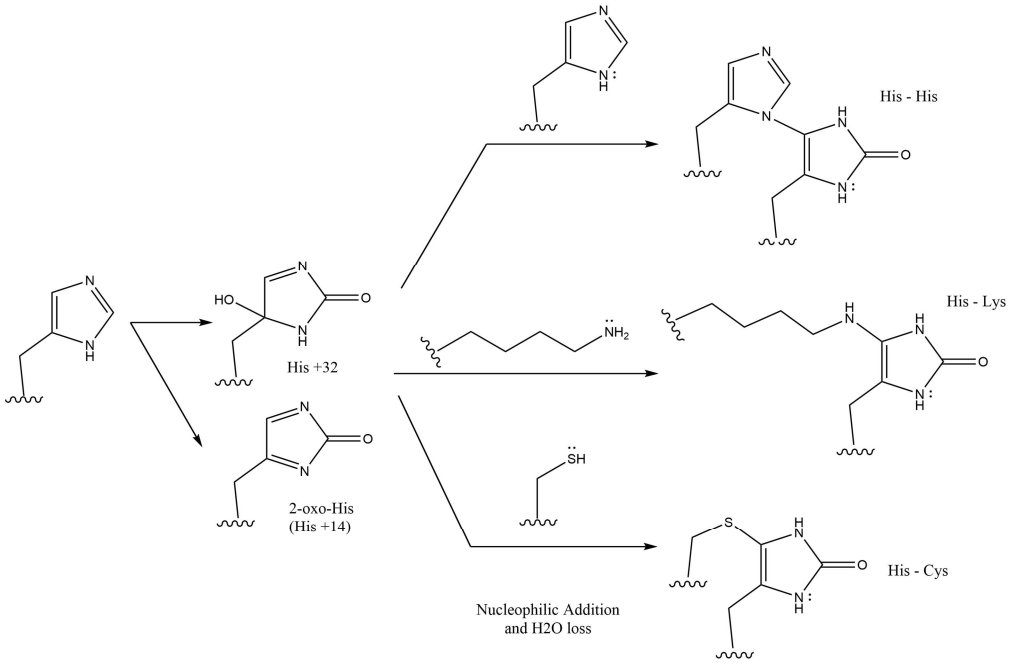


Figure IV-13 Proposed mechanisms of formation of His-His, His-Lys and His-Cys cross-links. Figure elaborated from Hägglund et al., 2018 with ChemPrime.

5.3 Mass spectrometry-based proteomics for cross-linking investigation

MS-based proteomics has proven to be one of the most effective methods to identify and map protein cross-links, providing a better understanding of its working mechanisms. Protein chemical cross-linking coupled with mass spectrometry (XL-MS) is the central branch of this growing field focused on structural proteomics studies, such as investigating protein-protein interactions conformation [59, 88]. The analyses are based on introducing a chemical cross-linker to the protein solution, followed by the digestion and mass spectrometry analysis of the resulting reticulated peptides to get an insight into the mechanisms of interaction with protein regions in proximity. Other studies, instead, have been conducted to break the chemical cross-links formerly provoked; for example, different researches focused on breaking the formaldehyde-based cross-links to improve the investigation of Formalin-Fixed Paraffin-Embedded (FFPE) tissues years after the preparation [62, 89-92].

A more recent ramification of the research field consists of detecting and localising native and insult-induced cross-link formations. This research pursues the comprehension of the products and mechanisms of cross-linking induced by protein ageing or provoked during the degradation.

The understanding of how cross-linking affects the original structure of proteins and, in turn, their physical properties and functionalities is critical in various fields, including food science and technology [53, 93-95], tissue engineering [96-99], and biomedicine [100-102].

In cultural heritage research, the investigation of proteins' structural alterations induced by the painting matrix and during ageing is critical in understanding the degradation processes of the compounds and the history of the object (for example, its storage conditions or previous restorations). These studies are at the very beginning and, to date, no investigations have been performed on reticulated proteins in artistic objects applying a proteomic strategy.

Different studies have focused on increasing the extractable and identifiable protein extent for the mass spectrometry analysis that, due to cross-links formation, shows a reduced solubility [19, 23, 103]. For example, new strategies of mass spectrometry analysis have been applied in the study of degraded samples (characterised by low solubility), such as Evolved gas analysis mass spectrometry (EGA-MS) that involves pyrolysis before the analysis leading to the additional identification of aggregated and cross-linked components [24, 25, 103, 104].

5.3.1 Challenges in the cross-link analysis to overcome

Cross-linking studies via MS-proteomics are predominantly conducted through a bottom-up approach. Compared to the classical characterisation of non-cross-linked peptides (also called linear), reticulated peptides' detection is more complicated. So far, it has been hindered by numerous challenges encountered in each step of the workflow, from sample preparation to data acquisition and interpretation (Table IV-1).

Table IV-1 Challenges encountered in a cross-linking investigation with mass spectrometry proteomics analysis and possible solutions to overcome them.

Steps of proteomic protocol	Challenges	Possible solutions
Extraction	Protein cross-links reduce protein solubility and extraction.	Partial acid hydrolysis of the sample would lead to the cleavage of some (not all) cross-links, improving the extraction [19].
Digestion	In the presence of cross-links involving Lys and Arg, the trypsin action is hampered. Difficulties in reaching an appropriate peptide length.	Implementation of combined or sequential proteases [100] other than trypsin.
Mass spectrometry analysis	Reticulated peptides are less abundant than linear peptides.	- Increase of the yield and reduction of the sample complexity through enrichment strategies [105]: - Size exclusion chromatography (SEC) since reticulated cross-links have a greater mass than linear [106, 107]. - Sodium Dodecyl Sulphate – Poly Acrylamide Gel Electrophoresis (SDS PAGE). -Strong cation exchange (SCX) chromatography and charged-base fractional diagonal chromatography (ChaFRADIC). Reticulated peptides have four terminal residues that increase their net charge compared to linear ones [108]. - ^{18}O isotope labelling. Cross-linked peptides have two C-termini than one in linear peptides; therefore, four ^{16}O will exchange with ^{18}O (8Da shift concerning usual 4Da) [109].

Mass spectrometry analysis	The intensity of cross-linked peptides is lower than linear ones. Their signals are often hidden, hardly selected and fragmented.	- Data independent analysis - Exclusion of low charge state precursors (one and two) for fragmentation and high-resolution measurements. - Inclusion list, whether the cross-linked peptide values are already known.
Data interpretation	- n-square problem: combination explosion of search space caused by the consideration of each possible pairwise combination of peptides in the database. With n peptides, there are $(n^2+n)/2$ possible pairwise combinations [95, 100], consequent increment of time search and false positives.	- Customised computational tools for cross-linking detection. - Meticulous manual control of the results.
Fragmentation interpretation	The fragmentation behaviour is different and not fully understood [110, 111]. Multiple fragmentation patterns cause complex MS/MS spectra.	- Automated annotation of MS2. The majority of algorithms are elaborated for known cross-linked masses (XL-MS) and not for insult-induced cross-links. - There is still a high necessity for manual control to complete the automated interpretation of the MS2 spectra.

5.3.2 Fragmentation of cross-linked peptides

The fragmentation of reticulated peptides differs from the classical one, and it is not yet fully understood. The MS/MS of a reticulated formation is a complex and heterogeneous mixture of different fragment ion possibilities. This vast extent of potential fragmentations ions results in complex MS2 spectra of difficult interpretation [112].

The simplest case is characterised by the collisional fragmentation (CID or HCD) of an intra-peptide cross-link since only one peptide is involved. Nonetheless, the fragment ions formed could be low in number since only the peptide regions not bonded to the link produce fragments ions (Figure IV-14-A).

The fragmentation pattern of a pair of two reticulated peptides results in much broader and heterogeneous MS/MS spectra (inter-peptide cross-link). The set of values generated is constituted both by fragments of every single peptide and by segments of different lengths of one peptide linked by the bond to parts of the other peptide, plus their counterparts (Figure IV-14–B). The detection of these ions can lead to an accurate structural comprehension of the reticulation and the localisation of the covalent bridge.

Depending on the covalent bond's strength, on the dissociation technique implemented in the analysis and on the fragmentation energy, the fragmentation of the covalent bonds might also occur [113]. In this case, two pairs of ions composed of the component peptides could be originated, one with the link while the other without (Figure IV-14–C). Occasionally, two fragmentation events could occur within the backbones and at the covalent bridge, forming b- and y- ions in which the cross-link is attached to only one of the peptides while the other fragments as if it were linear.

In this doctoral research, the widely accepted Schilling nomenclature [114] is implemented to annotate the peptides and the fragmentation ions. The longer peptide is defined as alpha (α), while the shorter is beta (β). Specific annotations were used for values composed of fragments of both peptides with the intact cross-linker (e.g. y2αy2β, y2αb3β) and values presenting fragments of α peptide linked with the entire peptide β, preserving intact the cross-linker (y2α_β).

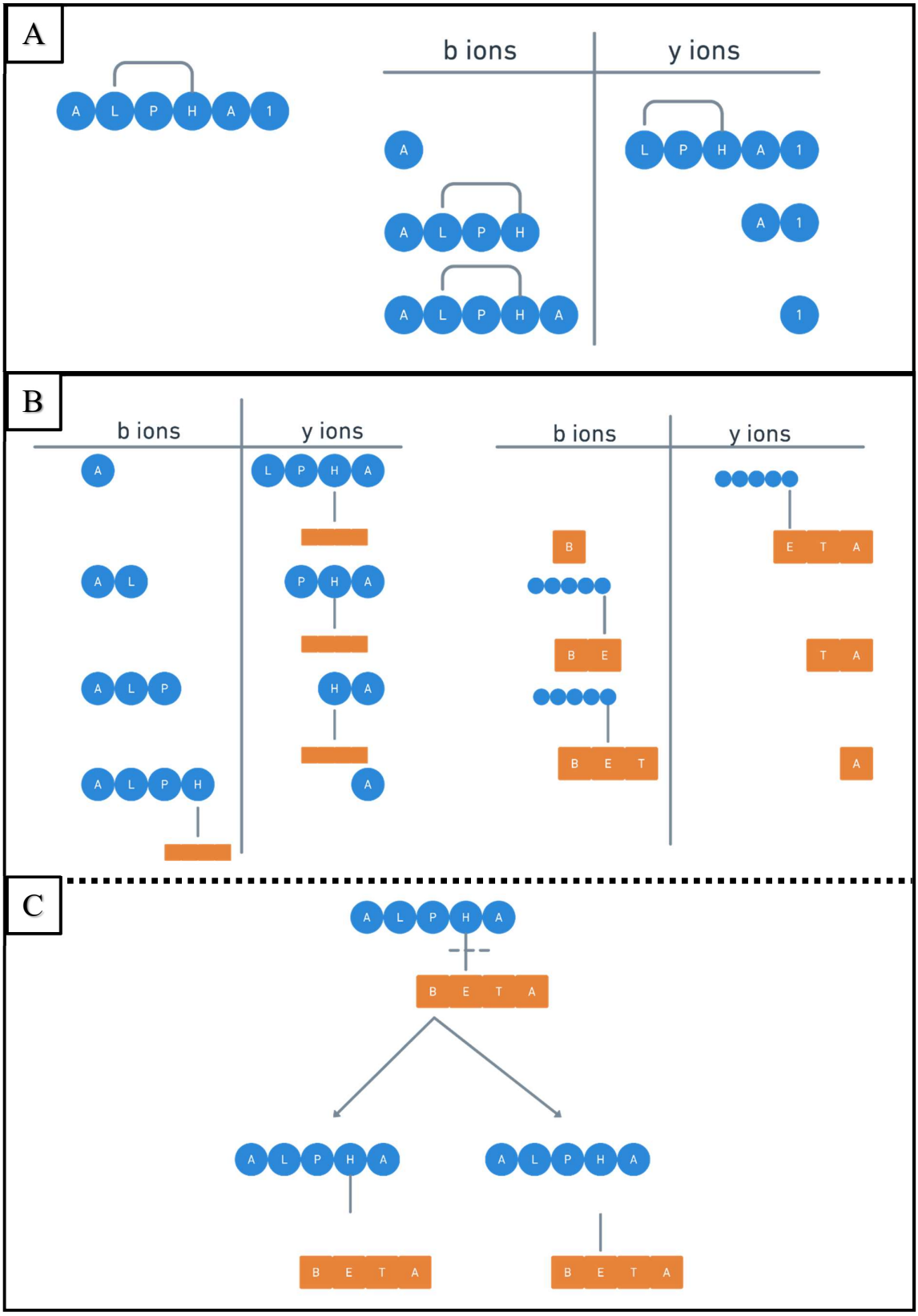


Figure IV-14 Potential fragmentation patterns in the presence of an intrapeptidic cross-link **(A)** and a pair of reticulated peptides **(B-C)**. The fragmentation might occur at the covalent bond **(B)** or within the peptide backbone **(C)**. Figure elaborated from Maes et al., 2017.

5.3.3 Data analysis of cross-linked peptides

The majority of computational software has been developed for identifying cross-links provoked by predefined chemical cross-linker implemented in structural studies (XL-MS) [115-117]. In these circumstances, the characteristic fragment signatures of the cross-linkers are known *a priori* [116]. The algorithms can be adapted to investigate native and insult-induced bonds, but these studies result more complicated due to the incertitude of the cross-links formed.

The great majority of the software package available require the definition of numerous parameters, including cross-linker unit, reactive amino acids residues, protein sequence, potential modifications, and protease implemented. Specific algorithms have been elaborated to assign fragments to the two peptides simultaneously: a library of each possible cross-linked peptide pairs is created according to the defined parameters. Then in-silico and experimental values are compared, and the matches are scored. The software tools most widely applied that implement this strategy are pLink [118], Kojack [119], and MS studio [120]. An alternative approach consists of the individual analysis of the two peptides: one is searched as if it were a linear peptide while the other is searched, adding the cross-link as modification (one example is Popitam software [121]).

The automated MS2 assignment provides critical help in the elucidation of the cross-linked peptide structures and on the assignment of the covalent bond site; nevertheless, there is still a need for several improvements, and manual verification of the results is inevitably required.

6. Study of protein cross-linking in art paintings using mass spectrometry-based bottom-up proteomics

6.1 Materials

6.1.1 Formulation of mock-up paints

The analyses were conducted on two painting mock-ups prepared on inert glass slides in 2010 with lead white pigment ($2PbCO_3*Pb(OH)_2$) mixed to lysozyme 10% (w/v). At the time of formulation, one model was treated with Hydroxy Peroxide H_2O_2 pure while the other was not subjected to any treatment. The model paints were left ageing at room temperature without control of the atmosphere environment.

6.1.2 Artistic and historical tempera paintings

The study was investigated in partnership with Dr Julie Arslanoglu (*Department of Scientific Research, The Metropolitan Museum of Art, New York*), who also collected a few µg of four egg-based historical artworks.

The first specimen was sampled from a tempera on wood, dated the early 1490s and representing "Madonna Adoring the Child with the Infant Saint John the Baptist and an Angel". The author was the Italian artist Lorenzo di Credi (Lorenzo d'Andrea d'Oderigo) (Accession number 03.197). The second sample was collected from the tempera on canvas painted by Spinello Aretino (b. 1345–52, d. 1410) and depicting "Saint Mary Magdalen Holding a Crucifix" on one side and "The Flagellation" on the reverse (Accession number 13.175). The third sample was collected from the adhesive on "Gilt-leather wall hanging" made around 1650-1670 in Amsterdam (Dutch) by De Gecroonde Son or by De Vergulde Roemer (Accession number 2012.332.2). The fourth sample analysed was provided by Dr Dobrochna Zielinska of the University of Warsaw. It was collected from the Royal Church's wall painting in Dongola (Nubian area) and dated 9th century (Accession number D_BV_6 for more detailed information of this sample, please refer to Chapter II5.1 on page 40). The table with the pictures of the paintings is reported in Annexe iii.

All the samples were already analysed through LC-ESI-MS/MS analysis starting from very low sample amounts estimated to few micrograms (bottom-up approach with liquid digestion using eFasp protocol described in Chapter II4.1 on page 35). The research consisted of the elaboration of the raw data for the identification of oxidation-based cross-links.

6.2 Methodology

6.2.1 Extraction of proteins from standard painting models

Samples collected from the painting models (2 mg) were subjected to protein extraction by adding 200 µl of 0.1% formic acid (AF), crushing with a pestle designed for the Eppendorf 1.5 mL, and was left stirring overnight. The next day, the solution was centrifuged to remove the maximum of solid particles, then evaporated and re-diluted in 0.1% of formic acid.

6.2.2 Separation with SDS-PAGE (standard painting models)

Separation by SDS-PAGE electrophoresis gel was done before the digestion: the analyte solution of 5 µL was mixed with 10 µL of Laemmli buffer and then placed in the wells of an SDS-PAGE (separation gel at T= 15% and concentration gel at T = 4.7%). A solution of 10 µl of molecular weight standard with 5 µL of Laemmli buffer was included. Electrophoresis was carried out at a constant current of 50 mA and 200 V for 90 min. Coomassie-Blue was used for protein detection (0.2g of Brilliant blue G / 7% Acetic acid / 20% Ethanol). The gel's discolouration was done after one night through an aqueous solution containing 7% Acetic acid and 20% Ethanol.

6.2.3 Digestion in gel (standard painting models)

Each singular band identified in the gel (multimers of lysozyme protein) was cut into cubes and digested singularly. The decolourisation was performed using a 50 mM solution of 50 mM Ammonium bicarbonate pH 8.8 to 50/50 with pure ACN by covering the gel for 20 minutes (3 times). Pure ACN was added for 5 minutes. Finally, the samples were dried for 10 min at 37 ° C.

The reduction was conducted via 15 mM DTT in 50 mM Ammonium Bicarbonate pH 8.8 (60 min at 46 °C), followed by alkylation with a solution of 55 mM iodoacetamide in 50mM Ammonium Bicarbonate pH 8.8 (45 min at 37 °C). The same procedure of discolouration was repeated twice.

All the samples were digested with a 20 ng/µL solution of trypsin in 50 mM Ammonium bicarbonate pH 8.8 (50 µL, at 4 ° C. for 15 minutes). Then, the solution was substituted with one of 50 mM Ammonium Bicarbonate pH 8.8 left overnight at 37 °C.

Once the digestion solution was displaced in another Eppendorf, the gel cubes were treated with 100 µl ACN 0.1% in formic acid (20 minutes, twice) to collect all the extracted peptides from the gel. The analyte solutions were evaporated and resuspended in 10µL of acidified water (0.1% FA).

6.2.4 LC-MS/MS analysis

The details regarding the mass spectrometry analysis conducted have been described in Chapter II2.4 on page 32.

6.2.5 Data analysis

PEAKS Studio (Bioinformatics Solutions Inc.) was used to identify non-cross-linked peptides with the following parameters: oxidation of methionine residues, deamidation of asparagine and glutamic acid, carbamidomethylation and propionamide of cysteine residues were searched as variable modifications. The propionamide modification was added only to analyse standard samples digested in-gel since cysteine side chains can react with non-polymerised free acrylamide resulting in acrylamide adduction [122, 123].

A maximum of 3 missed tryptic cleavages, parent mass error tolerance of 10 ppm and fragment mass error tolerance of 0.02 Da were set.

Two software tools explicitly designed for cross-linked analysis were implemented in the identification and localisation of reticulated peptides, confronting and combining the obtained results: Mass Spec Studio framework (Structural Mass Spectrometry group at the University of Calgary, Alberta, Canada) [117] and MaxQuant software (v1.6.14) supplied with the extension of cross-linking search (collaboration with MaxQuant group, specifically with J.Cox and Ş.Yılmaz).

For both bioinformatic investigations, the following variable modifications were included: carbamidomethylation and acrylamide adduct of cysteine residues (for samples with gel digestion), oxidation of methionine and deamidation of asparagine and glutamine.

A maximum of two missed tryptic cleavages, parent mass tolerance of 10 ppm, and an MS/MS peak tolerance of 0.02 m/z were set. Lysozyme C *Gallus gallus* (Chicken) P00698 (FASTA format) or 1DPX (PDB format) were included as references.

Under advanced settings, HCD fragmentation was selected. Dityrosine, tryptophan−tyrosine, di-tryptophan, lysine−histidine, histidine-histidine and histidine-cysteine were sequentially selected potential cross-links with the appropriate mass differences. The minimum charge for candidate cross-linked peptides was defined as 3. Cross-linked peptides were ordered and considered based on their score, the MS/MS spectra quality and the fragment ion coverage. Furthermore, all automatically identified cross-linked peptides were validated by a careful manual inspection of both MS1 and MS2 spectra to validate the candidate and avoid false positives.

6.3 Analysis and discussion of the results

6.3.1 Model paintings

The gel electrophoresis SDS-PAGE was performed to obtain a visual confirmation of the potential cross-linking formations and facilitate their investigation by enriching the yield and reducing the sample complexity. The two samples from the naturally aged painting models (treated with H_2O_2 pure and untreated) and the lysozyme without pigment (standard) were investigated in the same gel (Figure IV-15). Samples were reduced and alkylated (no disulfide bonds). The same amount of material was loaded to provide an accurate comparison. The separation of the protein populations based on their charge and molecular weight allowed the visible discrimination of monomers, dimers and other higher forms for the aged compounds. Except for the standard, which displayed only the band around 14300 Da characteristic of the monomeric form of lysozyme, the paint mock-up showed two bands corresponding to the dimer (28607 Da) and trimer (42911 Da) forms. A tetramer (57215 Da) form was also detected on lysozyme mock-up treated with H_2O_2 pure.

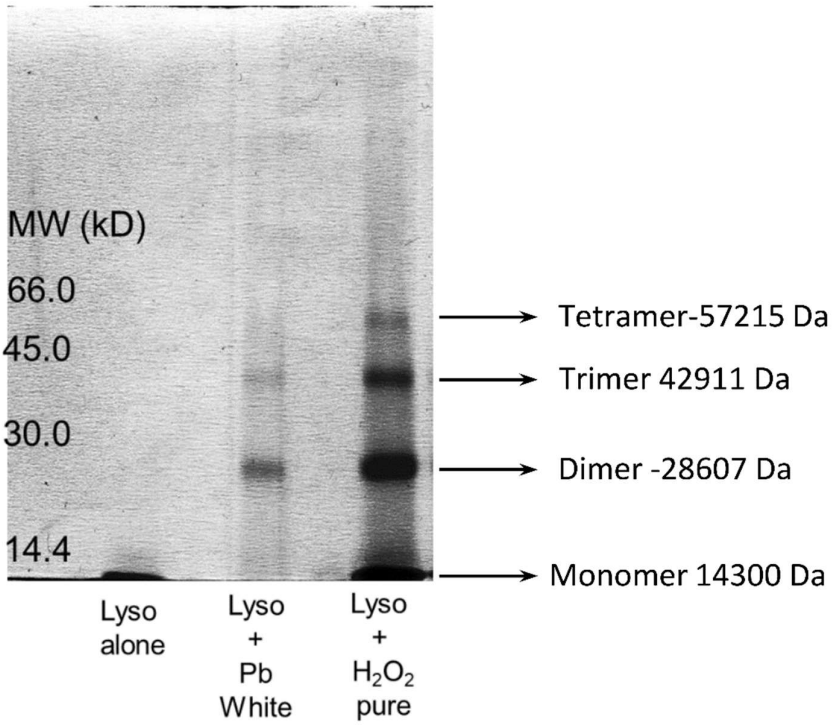


Figure IV-15 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of lysozyme protein standard and two naturally aged painting models (lysozyme/ lead white pigment). Two multiforms were observed for both aged samples (dimers and trimers). Furthermore, a tetramer form was highlighted in the mock-up, which was treated with pure H2O2.

Considering the sample treatment and the high number of oxidations identified from the linear investigation of the MS data, the study focused on the research of oxidation-induced cross-links (Tyr-Tyr; Trp-Tyr; Trp-Trp; His-His; His-Lys; His-Cys). The confident identification of numerous cross-linked pairs allowed the demonstration of almost all the researched cross-linking types in the old model paints (only His-His and His-Cys were not observed). The greatest extent of reticulated peptides was identified in the dimeric lysozyme form of the aged painting model treated with pure H_2O_2. Several reticulated peptides were also highlighted in the untreated aged painting model. Both intra- and inter- peptidic cross-links were pointed out; some inter-protein bonds were also identified. No cross-links have been highlighted in lysozyme standard without pigment, in agreement with the observation made from SDS-PAGE experiments. The complete list of the cross-links detected in the two model paintings is reported in Table IV-2.

Di-tyrosine cross-linking

In the dimeric form of lysozyme treated with H_2O_2 pure, the MS and MS2 spectra of the triply charged ion m/z 1031.4832 (mass monoisotopic mass 3091.4193 Da Δm=2.19) suggested the detection of a di-tyrosine formation constituted by two distant peptides ^{46}nTDGSTD**Y53**GILQInSR^{61} - ^{22}G**Y23**SLGNWVcAAK^{33}. (Figure IV-16-A). The mass loss of 2.014 Da concerning the sum of the parent peptides is characteristic of a Tyr-Tyr cross-link (Y53-Y23 ~24 Å distance) [109, 124]. In the fragmentation spectrum, the detection of the almost complete sequence of y- and b- fragment ions characteristic of both peptides suggested that two fragmentations occurred, one in the peptide backbone and the other at the cross-link bridge. These fragment ions led to a straightforward localisation of two deaminations (+0.984) at the asparagine residues (N46-N59) of the peptide α; one acrylamide adduction (+71.037) was instead identified on the cysteine residue of the peptide β (C30). The observation of the precursors' ions $(\alpha)^{2+}$ and $(\beta)^{2+}$ suggested a single event of fragmentation at the cross-link bridge between two aromatic rings [125].

Fragment ions composed of the two peptides bonded segments were not detected, preventing the confirmation of the Y53-Y23 bond. The implementation of collision energy too strong could be a potential explanation of the observed fragmentation. In support of the findings, the same di-tyrosine compound was already observed through ^{18}O isotopic labelling [109].

Oxidising forms of the triply charged ion m/z 1031.4832 were also detected in the close elution range (Figure IV-16-B). The MS2 spectrum of the triply charged ion m/z 1036.4819 (MW 3106.4222 Da with Δm=-2.63 ppm) determined the localisation of one oxidation of the tryptophan (W28) in the peptide β ^{22}G**Y23**SLGNWVcAAK^{33}. Besides this, a propionamide modification (C30) was detected on the same peptide, and one asparagine deamination (N46) was localised in the peptide α ^{46}nTDGSTD**Y53**GILQINSR^{61}. The confident assignment of the oxidation on the tryptophan residue strengthens the covalent bond hypothesis between the tyrosine residues.

The triply charged ion m/z 1041.8188 (monoisotopic mass 3122.4329 Da) corresponds to a further oxidation form (+ 31.990 Da respect to 3091.4193 Da). However, the absence of comprehensive data in its fragmentation spectrum did not allow a confident definition and localisation of the modifications.

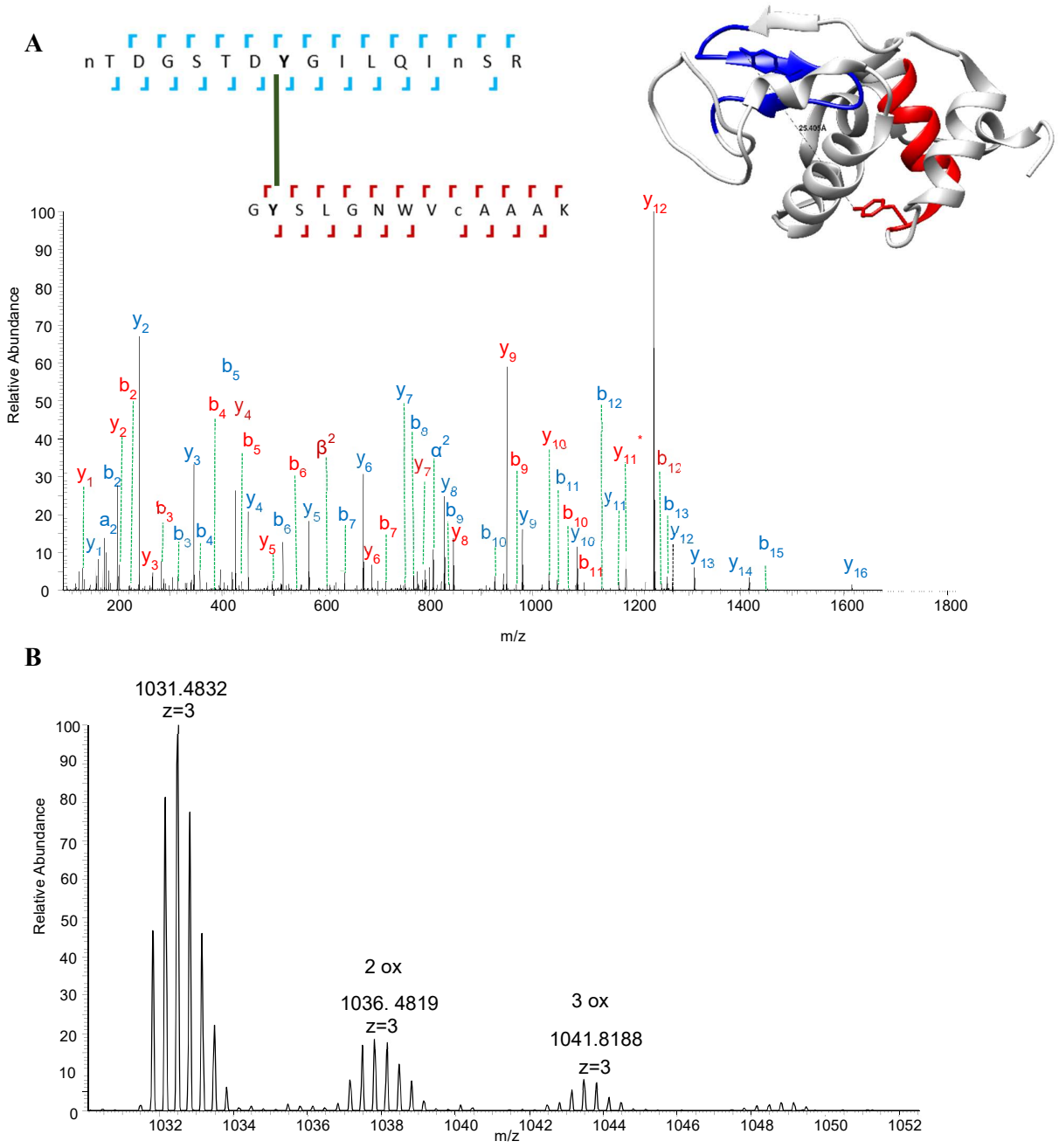


Figure IV-16 (A) MS2 of the triply charged ion m/z 1031.4832 (Δm=2.19 ppm) attributed to the Tyr-Tyr (-2H) reticulated peptide pair ^{46}nTDGSTDY53GILQInSR^{61} - ^{22}GY23SLGNWVcAAK^{33}. In the peptide α, two deaminations (N46-N59) were detected on asparagine residues, while one acrylamide adduction was identified on the cysteine residue of the peptide β (C30). On the top right of the figure is presented the structure of the lysozyme C of the hen egg. The peptide α is coloured in blue, and the peptide β in red; the two side chains of tyrosine involved in the covalent bond are also highlighted. **(B)** MS spectrum representing the oxidation forms of the triply charged ion m/z 1031.4832 (potential di-tyrosine). The oxidation localised on the tryptophan in the MS2 spectrum of the triply charged ion 1036.4819 strengthened the hypothesis of a covalent link between the tyrosine residues.

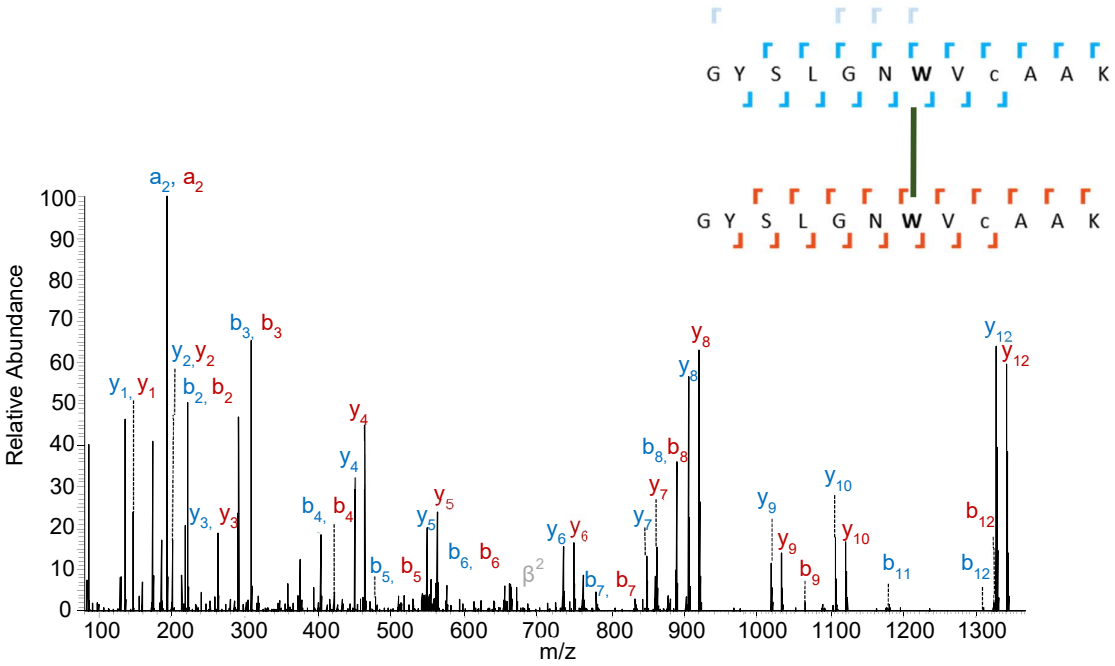
Furthermore, two different interprotein di-tyrosine compounds were outlined in the lysozyme model painting treated with H2O2 pure: Y53-Y53 (^{46}nTDGSTD**Y53**GILQInSR^{61})$_2$ and (Y20-Y20) (^{14}RHGLDN**Y20**R^{21})$_2$. This latter cross-link was also identified in the naturally aged mock-up (untreated) without the miscleavages (^{15}HGLDN**Y20**R^{21})$_2$. Noteworthy, in the MS/MS spectrum of the charged 4^+ ion m/z 515.2574 (Δm=-3.25ppm) assigned to the homodimeric form (^{14}RHGLDN**Y20**R^{21})$_2$ two fragment ions attributable to segments of both peptides jointed via diTyr (-2H) were highlighted: the doubly charged ion m/z 507.745 $(y2\alpha y5\beta)^{2+}$ and m/z 674.351 $(y2\alpha y2\beta)^{1+}$. These values led to the precise localisation of the cross-link between the two tyrosine residues.

Di-tryptophan cross-linking

The second type of cross-link formation observed in the samples involved Trp-Trp residues. Di-tryptophan formations were already identified in various proteins, including lysozyme [80, 82, 83] due to the oxidation process [79, 81].

Similarly to di-tyrosine compounds, both intra- and inter-cross links were detected in the samples under analysis. The formation of the homodimeric form (^{22}GYSLGN**W28**VcAAK^{33})$_2$ was assumed from the MS and MS2 analyses of the triply charged ion m/z 888.0911 (MW 2661.2501 Da Δm=1.20 ppm) (Figure IV-17- A). In the peptide α, one carbamidomethylation was localised on the cysteine residue, while in the peptide β, a propionamide modification was detected on the cysteine C30. The investigation of the fragmentation spectrum resulted critically in the confirmation of the di-tryptophan covalent bond (W28-W28): the HCD fragmentation detected some y- fragments of the peptide α (y_6, y_7, y_8 and y_{12}, all containing the Trp residue) also presenting -2.0156 Da (Figure IV-17- B). This behaviour was already described for the same homodimeric peptide detected in lysozyme [80] and a Trp-Trp dimer in hSOD1 protein [79, 126]. In agreement with the previous studies, these data were explained with a cross-linking bond between C3-N1 or C3-C3 of two tryptophans.

A



B

y Fragment	Sequence	m/z		Sequence	m/z
y_6	WVCAAK	734.3668		W(-2H)VCAAK	732.3512
y_7	NWVCAAK	848.4094		NW(-2H)VCAAK	846.3938
y_8	GNWVCAAK	905.4306		GNW(-2H)VCAAK	903.4150
y_{12}	GYSLGNWVCAAK	1325.6326		GYSLGNW(-2H)VCAAK	1323.6170

Figure IV-17 (A) MS/MS spectrum of the triply charged ion m/z 888.0911 (MW 2661.2501 Da Δm=1.20 ppm) attributed to the homodimeric form (^{22}GYSLGNW28VcAAK^{33})$_2$. In the peptide α, one carbamidomethylation was localised on the cysteine residue, while in the peptide β a propionamide modification was detected on the cysteine C48. **(B)** Table of the two series of y-ions fragments (detected with and without loss of -2.0156 Da).

The reticulations involving Trp residues of the same peptide were observed in all painting models under investigation. For instance, the doubly charged ion at m/z 845.3932 (MW 1688.7707Da, Δm=-1.30 ppm) detected in the aged mock-up (not treated) was attributed to the peptide ^{98}IVSDGNGMNA**W108**VA**W111**R^{112} with an intra-peptide cross-link between two tryptophan residues W108 and W111. The MS2 spectrum provided the peptic sequence's characterisation and the localisation of methionine oxidation (M105). Furthermore, the localisation of cross-link at Trp-Trp was presumed by detecting ion fragments (y_5, y_6, y_7 and y_9) with -2.014 Da (-2H).

In all samples, an analogue intra-cross-link was identified between Trp62-Trp63 of the peptide ^{62}**W62W63**cNDGR^{68}. This cross-linked peptide was observed in both samples under analysis, presenting the cysteine modified with either acrylamide adduction or a carbamidomethylation; for instance, in the untreated aged model, its presence was ascribed to the detection of the doubly charged ions at m/z 503.1945 (Δm=-1.39 ppm) (di-Trp with +57.021) and m/z 496.1863 (Δm=-2.22 ppm) (di-Trp with +71.037).

Tyrosine-tryptophan cross-linking

In the untreated painting model, a tyrosine-tryptophan cross-linking formation was assumed between the peptides ^{46}NTDGSTD**Y53**GILQInSR^{61} and ^{97}KIVSDGNGMnA**W108**VAwR^{112}. The dimeric form's detection was achieved from the fragmentation of the triply charged ion m/z 1191.5618 (MW 3571.6619 Da) (Figure IV-18). The MS/MS spectrum also provided the localisation of asparagine deamidation (N59) in peptide α through the characterisation of fragment ions from y_1 to y_{10} and b_5. The ions y_1, y_2 and b_2 to b_5,b_7, b_9 assigned to the peptide β provided the localisation of the asparagine deamidation (N106) and the tryptophan oxidation (W111). Moreover, ions from the fragmentation of the covalent bond α (1754.8191) and β $(910.9411)^{+2}$ were detected.

The lack of values comprising fragments with the linked peptides hindered the confirmation of the cross-linking site. Nevertheless, the localisation of the oxidation on the W111 residue in the peptide β strengthened the assumption of a Y53-W108 bond.

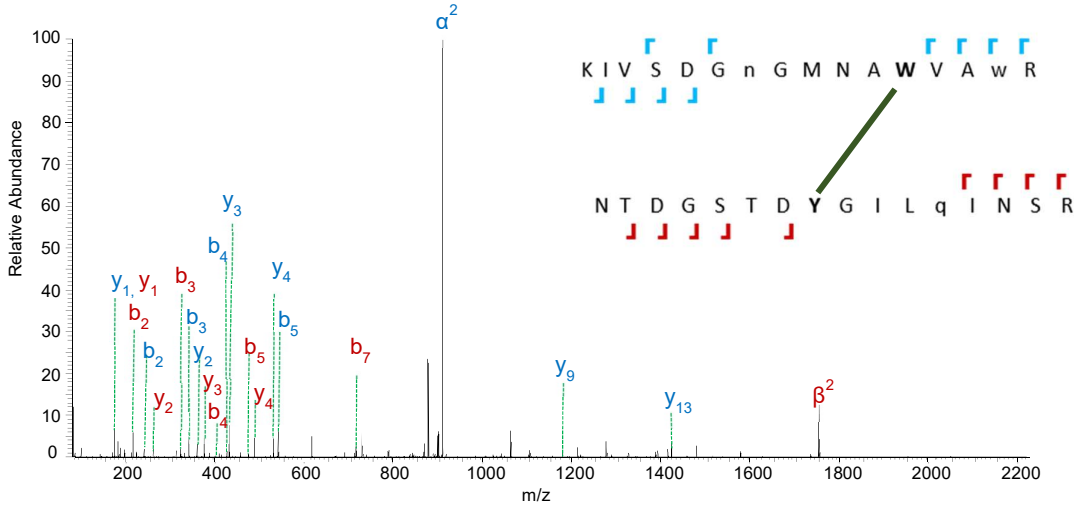


Figure IV-18 MS/MS spectrum of the triply charged ion m/z 1191.5618 (MW 3571.6619 Da) from the untreated painting model. The value was attributed to the inter-peptide cross-link (Tyr-Trp) between the peptides ^{46}NTDGSTDY53GILQInSR^{61} and ^{97}KIVSDGNGMnAW108VAwR^{112}. A deamidation of asparagine was localised on N59 of the peptide α while an asparagine deamidation (N106) and tryptophan oxidation (W111) were identified on peptide β.

Histidine- lysine cross-linking

The last type of cross-link detected in the painting mock-ups involved histidine and lysine residues. The 4^+charged ion at m/z 374.1953 was attributed to the reticulated form of ^{15}**H15**GLDNYR^{21} – ^{1}**K1**VFGR^{5} having a covalent bond between the His15 and Lys1 residues. The link formation was characterised by the mass increment of 13.9840 Da, equivalent to the loss of two hydrogen atoms and the acquisition of oxygen.

The MS2 spectrum of the charged ion provided the dimeric compound's characterisation and the localisation of the cross-link (Figure IV-19). Peptide α was identified through the ion fragments from y_1 to y_4 and b_4, the detection of y_1 to y_3 and b3 instead provided the peptide β determination. Furthermore, the observation of four other values provided strong evidence of the proposed cross-link: m/z 393.723 $(b2\alpha\beta)^{2+}$, m/z 407.719 $(b2\alpha\beta)^{2+}$, m/z 464.262 $(b3\alpha\beta)^{2+}$ and m/z 521.776 $(b4\alpha\beta)^{2+}$. These signals reflected different HGLDNYR chain fragments summed with the entire β peptide (KVFGR), preserving intact the H-K cross-link.

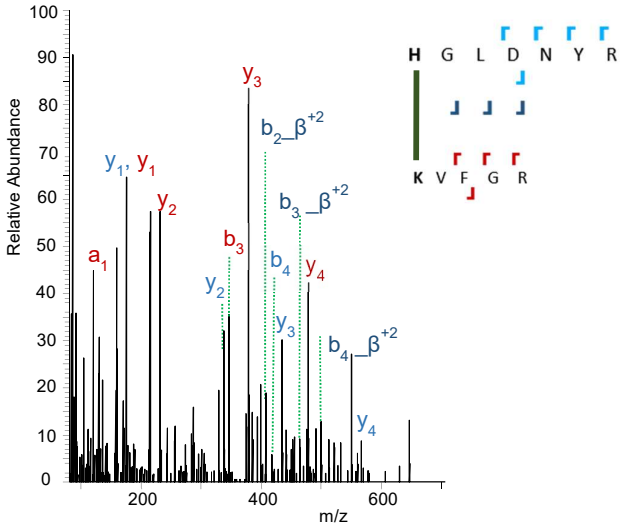


Figure IV-19 MS/MS spectrum of the 4^+charged ion at m/z 374.1953 from the treated model painting. The values were attributed to the reticulated form of (^{15}H15GLDNYR^{21}) – (^{1}K1VFGR^{5}), having a covalent bond between His15 and Lys1 residues. His-Lys cross-link hypothesis was strengthened by observing four fragment ions constituted by different fragments of the HGLDNYR chain summed with the entire β peptide (KVFGR), preserving intact the H-K cross-link.

Table IV-2 Cross-links identified in natural aged painting mock-ups (lysozyme/ lead white)

Cross-link type	Cross-linked peptides	m/z	Experimental mass (Da)	Theoric mass (Da)	Mass error ppm	RT	Charge (z)	modifications	Sample
Y-Y	αNTDGSTD**Y53**GILQINSR βG**Y23**SLGNWVCAAK	1026.4843	3076.4296	3076.4197	3.21	84.56	3	α: deam (N46) β: carb (C30)	H2O2 pure
		1031.4832	3091.4261	3091.4193	2.19	85.86		α: deam (N46-N59) β: prop (C30)	H2O2 pure
		1036.4819	3106.4222	3106.4302	-2.63	81.53		α: deam (N46) β: prop (C30), ox (W28)	H2O2 pure
		1041.8188	3122.4329	3122.4251	2.49	85.35		α: deam β: prop (C30), 2ox	H2O2 pure
	α NTDGSTD**Y53**GILQINSR β NTDGSTD**Y53**GILQINSR	1169.2178	3504.6299	3504.6241	1.65	81.45	3	α: deam(N46)	H2O2 pure
	α RHGLDN**Y20**R β RHGLDN**Y20**R	515.2574	2056.9985	2057.0052	-3.25	67.28	4	-	H2O2 pure
	α HGLDN**Y20**R β HGLDN**Y20**R	437.2067	1744.7955	1744.8030	-4.29	91.41	4	-	Untreated (ageing)
W-W	IVSDGNGMNA**W108**VA**W111**R	837.3962	1672.7767	1672.7780	-0.77	84.99	2	-	H2O2 pure
		837.3939	1672.7721	1672.7780	-3.52	90.13			Untreated (ageing)
		845.3939	1688.7721	1688.7729	-0.47	81.37		ox (M105)	H2O2 pure
		845.3932	1688.7707	1688.7729	-1.30	81.32			Untreated (ageing)
		837.8879	1673.7601	1673.7620	-1.13	87.48		deam (N106)	H2O2 pure
		837.8877	1673.7597	1673.7620	-1.37	85.00			Untreated (ageing)
	W62W63CNDGR	496.1884	990.36.12	990.3593	1.92	88.78	2	α: carb (C64)	H2O2 pure
		496.1863	990.3571	990.3593	-2.22	72.52			Untreated (ageing)
		503.1960	1004.3764	1004.3749	1.49	87.42		α: prop (C64)	H2O2 pure
		503.1945	1004.3735	1004.3749	-1.39	73.78			Untreated (ageing)
	α GYSLGN**W28**VCAAK β GYSLGN**W28**VCAAK	888.0911	2661.2501	2661.2469	1.20	85.85	3	α: carb (C30) β: prop (C30)	H2O2 pure
		898.4210	2692.2395	2692.2414	-0.70	85.79		α: prop (C30);ox (Y23) β: prop (C30);deam (N27)	H2O2 pure
Y-W	α nTDGSTD**Y53**GILQINSR βKIVSDGNGMNA**W108**VAwR	1191.5618	3571.6619	3571.6638	-0.53	81.56	3	α: 1dea (N46) β: 1 dea (N106); ox (W111)	Untreated (ageing)
H-K	α **H15**GLDNYR β **K1**VFGR	374.1953	1492.7499	1492.7583	-5.62	69.99	4	-	H2O2 pure

6.3.2 Artistic and historic tempera paintings

After the promising results achieved in the study of the simplified painting mock-ups, the developed protocol for cross-linking investigation was applied to more complex proteinaceous binders of historical tempera paintings. Four samples were chosen from different artworks (canvas, wood, wall supports and an adhesive on "Gilt-leather wall hanging") in which a former MS-analysis (BUP) demonstrated the implementation of egg and egg-white proteins. The research proved to be more challenging than the previous two main reasons: the analyte was extremely low in amount and with a protein composition more complex and heterogeneous. Various egg proteins were detected in all proteinaceous binder. Moreover, in some samples, other proteins were also observed, for example, collagens as adhesive. Another difficulty was the lack of the gel separation step, which enabled the enrichment of the cross-linked proteins and facilitated their peptide fragmentation and detection.

The study's initial objective was to research oxidation-induced cross-links in the historical samples focusing on the lysozyme protein. Despite the outlined challenges, various cross-links were successfully highlighted in all historical artworks. In Table IV-3 the observed reticulation forms are listed. Both interpeptide and intrapeptide cross-links involving Trp-Trp a Tyr-Tyr were reported for all samples. Furthermore, one cross-link His15-Lys1 bond was identified in the painting on wood made by Dicredi.

Figure IV-20 reports the MS/MS spectrum of the triply charged ion at m/z 1169.5380 (MW 3505.5908 Da) as an example, detected in the tempera on wood (Lorenzo di Credi). The MS and fragmentation spectra analysis provided the attribution of the charged ion to the homodimer (^{46}NTDGSTD**Y53**GILQInSR^{61}) involving a covalent bond between two Tyr53 of different lysozyme proteins. Almost all fragment ions were detected, most likely due to a double fragmentation at the covalent bond and peptidic backbone. The peptides involved in the cross-link resulted analogous, both presenting a deamidation on N46. Unfortunately, the lack of fragment ions comprising segments from the two bonded peptides prevented the accurate confirmation of the Y53-Y53 bond. However, it can be remarked that the same homodimer was detected and localised in the painting model treated with H_2O_2 pure. The findings highlighted the potentialities of the developed strategy capable of identifying cross-links in complex and aged proteinaceous binders, notwithstanding their low abundance and intensity.

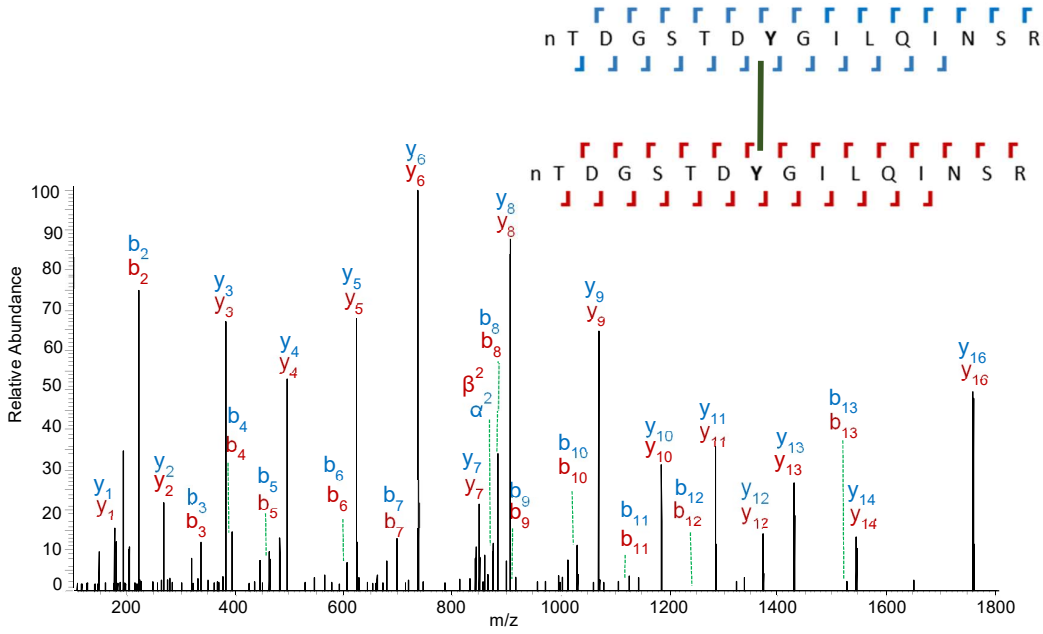


Figure IV-20 The MS/MS spectrum of the triply charged ion at m/z 1169.5380 (MW 3505.5908 Da), detected in the tempera on wood (Dicredi). The charged ion was attributed to the homodimer (^{46}NTDGSTDY53GILQInSR^{61}) involving a covalent bond between two Tyr53 of different lysozyme proteins. The detection of almost all fragment ions suggested fragmentation at the covalent bond and the peptide backbone. The peptides involved in the cross-link resulted analogous, both presenting a deamidation on N46.

However, the number of detected oxidation-induced protein cross-links is potentially lower than the real extent present in the paintings. The investigation of oxidation-induced cross-links in artworks remained subordinated to the list of the potential covalent bonds previously identified in the standards. The software tools failed in identifying numerous cross-linked forms due to their low fragmentation and resulted in necessary manual research of the values determined in the painting standards.

Despite the need for other improvements, the achievements of this study potentially opened to new perspectives in the investigation of covalent bonds in aged proteinaceous binders.

The determination and collection of molecular and fragmentation fingerprints of different reticulated peptides from various origin and proteins would critically contribute to the improvement of cross-linking examination in complex systems. Our current investigations are also focused on the cross-linked forms' enrichment strategies using various separation supports (i.e. chromatography, Native GELFrEE).

Furthermore, other dissociation techniques will be tested, such as EthCD and UVPD, which have been proven to provide higher fragment ion coverage of the cross-linked peptides.

Table IV-3 Cross-links identified in four historical tempera based artworks

Cross-link	Cross-linked peptides	m/z	Experimental mass (Da)	Theoric mass (Da)	Mass error (ppm)	RT	Charge (z)	modifications
Madonna Adoring the Child (Lorenzo di Credi) Tempera on canvas								
Y-Y	α NTDGSTD**Y53**GILQINSR β G**Y23**SLGNWVCAAK	1026.4830	3076.4255	3076.4197	1.885	83.65	3	α: deam (N46) β: carb (C30)
		1031.4820	3091.4225	3091.4305	-2.587	85.90	3	α: ox β: carb (C30)
	α NTDGSTD**Y53**GILQINSR β NTDGSTD**Y53**GILQINSR	1169.5380	3505.5908	3505.5908	-2.767	122.07	3	α: deam(N46) β: deam(N46)
W-W	IVSDGNGMNA**W108**VA**W111**R	837.3955	1672.7753	1672.7780	-1.614	122.99	2	-
		845.3926	1688.7695	1688.7729	-2.013	33.10		ox (M105)
H-K	α **H15**GLDNYR β **K1**VFGR	498.5956	1492.7633	1492.7583	3.349	70.05	3	-
Saint Mary Magdalen Holding a Crucifix (Spinello Aretino) Tempera on wood								
Y-Y	α RHGLDN**Y20**R β RHGLDN**Y20**R	692.0164	2072.9984	2073.0052	-3.280	73.04	3	Ox
W-W	IVSDGNGMNA**W108**VA**W111**RNR	988.4637	1974.9117	1974.9119	-0.101	92.99	2	diox (M105)
Royal Church, Dongola, Nubian region wall painting								
Y-Y	α NTDGSTD**Y53**GILQINSR β G**Y23**SLGNWVCAAK	773.8626	3091.4191	3091.4305	-3.687	117.18	4	β: carb (C30) ox (W28)
		1026.4808	3076.4189	3076.4197	-0.260	124.07	3	α: deam β: carb (C30)
	α RHGLDN**Y20**R β RHGLDN**Y20**R	692.0081	2073.0008	2073.0052	-2.122	87.55	3	α: ox
Dutch Adhesive Egg-based adhesive								
Y-Y	α NTDGSTD**Y53**GILQINSR β G**Y23**SLGNWVCAAK	1026.1525	3075.4340	3075.4356	-0.520	46.67	3	β: carb (C30)

7. Conclusions

Mass spectrometry analyses were combined with hydrogen/deuterium exchange and cross-linking investigations to better understand the structure and modifications at the conformational level in proteinaceous binders (tempera paintings). Both techniques were not yet performed in cultural heritage studies. The studies were performed on a lysozyme model system to simplify the complexity of egg protein composition. HDX-MS highlighted the variations that lysozyme structures undergo induced by solely introducing inorganic pigments (before the ageing). The results suggested that the inorganic salts influenced the elevated extent of oxidations PTMs on amino acids residues (intact analysis before the exchange) and reduced the availability of hydrogens for an isotopic exchange with deuterium.

The hindering action observed for cinnabar, lead white, and zinc white was interpreted with the folding of certain protein regions or due to the pigment interposition between hydrogens and solvent. Cinnabar showed a more significant rate of alteration of deuterium, while zinc white pigment proved no modifications of the overall extent with respect to unpigmented lysozyme. Nonetheless, further HDX-MS studies have to be performed following a bottom-up approach to provide the localisation and a better elucidation of the interactions established.

Complementary information on proteins' ageing and oxidation degradations were achieved by performing cross-linking investigations with MS analysis. A new strategy based on the elaboration of data from the bottom-up analysis was optimised to investigate the protein network formations through the localisation and characterisation of cross-linked peptidic pairs. The detection of different reticulated formations (Trp-Trp; Tyr-Tyr; His-Lys) provided the confident definition of various molecular patterns characteristic of lysozyme oxidative-based cross-links. The identification of some of these cross-link peptide pairs also in historical tempera paintings (a few tens of μg of a sample) highlighted the great potentialities of the developed methodology. . It can be underlined that the study was performed without the use of chemical cross-linkers introduced in the sample as it is usually done for biological samples. The establishment of a list of reticulated peptides for different proteins could be implemented as a complement of the linear analyses to reveal the protein fraction potentially undetected because of their reticulations.

Both techniques were highly promising in understanding the proteins' structural and conformational alterations within an artwork. This new molecular information provides new insights into the study of proteins and their networks in historic artworks.

Reference

1. Bonaduce, I., L. Carlyle, M.P. Colombini, C. Duce, C. Ferrari, E. Ribechini, P. Selleri ,M.R. Tine, *New insights into the ageing of linseed oil paint binder: a qualitative and quantitative analytical study.* PLoS One, 2012. **7**(11): p. e49333.
2. Cotte, M., E. Checroun, W. De Nolf, Y. Taniguchi, L. De Viguerie, M. Burghammer, P. Walter, C. Rivard, M. Salomé ,K. Janssens, *Lead soaps in paintings: Friends or foes?* Studies in Conservation, 2017. **62**(1): p. 2-23.
3. Modugno, F., F. Di Gianvincenzo, I. Degano, I.D. van der Werf, I. Bonaduce ,K.J. van den Berg, *On the influence of relative humidity on the oxidation and hydrolysis of fresh and aged oil paints.* Scientific reports, 2019. **9**(1): p. 1-16.
4. Bonaduce, I., C. Duce, A. Lluveras-Tenorio, J. Lee, B. Ormsby, A. Burnstock ,K.J. van den Berg, *Conservation issues of modern oil paintings: a molecular model on paint curing.* Accounts of Chemical Research, 2019. **52**(12): p. 3397-3406.
5. Garrappa, S., E. Kočí, S. Švarcová, P. Bezdička ,D. Hradil, *Initial Stages of Metal SoapsFormation in Model Paints: The Role of Humidity.* Microchemical Journal, 2020: p. 104842.
6. Casadio, F., K. Keune, P. Noble, A. Van Loon, E. Hendriks, S.A. Centeno ,G. Osmond, *Metal soaps in art*. 2019: Springer.
7. van den Berg, K.J., A. Burnstock ,M. Schilling, *Notes on metal soap extenders in modern oil paints: history, use, degradation, and analysis*, in *Metal Soaps in Art*. 2019, Springer. p. 329-342.
8. Baij, L., J.J. Hermans, K. Keune ,P.D. Iedema, *Time-dependent ATR-FTIR spectroscopic studies on solvent diffusion and film swelling in oil paint model systems.* Macromolecules, 2018. **51**(18): p. 7134-7144.
9. Lee, J., I. Bonaduce, F. Modugno, J. La Nasa, B. Ormsby ,K.J. van den Berg, *Scientific investigation into the water sensitivity of twentieth century oil paints.* Microchemical Journal, 2018. **138**: p. 282-295.
10. van Dam, E.P., K.J. van den Berg, A.N.P. Gaibor ,M. van Bommel, *Analysis of triglyceride degradation products in drying oils and oil paints using LC–ESI-MS.* International Journal of Mass Spectrometry, 2017. **413**: p. 33-42.
11. Odlyha, M., N.S. Cohen ,G.M. Foster, *Dosimetry of paintings: determination of the degree of chemical change in museum exposed test paintings (smalt tempera) by thermal analysis.* Thermochimica Acta, 2000. **365**(1-2): p. 35-44.
12. Odlyha, M., N.S. Cohen, G.M. Foster ,R.H. West, *Dosimetry of paintings: determination of the degree of chemical change in museum exposed test paintings (azurite tempera) by thermal and spectroscopic analysis.* Thermochimica acta, 2000. **365**(1-2): p. 53-63.
13. van den Brink, O.F., E.S. Ferreira, J. van der Horst ,J.J. Boon, *A direct temperature-resolved tandem mass spectrometry study of cholesterol oxidation products in light-aged egg tempera paints with examples from works of art.* International Journal of Mass Spectrometry, 2009. **284**(1-3): p. 12-21.
14. van den Brink, O.F., J.J. Boon, P.B. O'Connor, M.C. Duursma ,R.M. Heeren, *Matrix-assisted laser desorption/ionization Fourier transform mass spectrometric analysis of oxygenated triglycerides and phosphatidylcholines in egg tempera paint dosimeters used for environmental monitoring of museum display conditions.* Journal of mass spectrometry, 2001. **36**(5): p. 479-492.
15. Pellegrini, D., C. Duce, I. Bonaduce, S. Biagi, L. Ghezzi, M.P. Colombini, M.R. Tinè ,E. Bramanti, *Fourier transform infrared spectroscopic study of rabbit glue/inorganic pigments mixtures in fresh and aged reference paint reconstructions.* Microchemical Journal, 2016. **124**: p. 31-35.

16. Duce, C., E. Bramanti, L. Ghezzi, L. Bernazzani, I. Bonaduce, M.P. Colombini, A. Spepi, S. Biagi ,M.R. Tine, *Interactions between inorganic pigments and proteinaceous binders in reference paint reconstructions.* Dalton Transactions, 2013. **42**(17): p. 5975-5984.
17. Ghezzi, L., C. Duce, L. Bernazzani, E. Bramanti, M.P. Colombini, M.R. Tiné ,I. Bonaduce, *Interactions between inorganic pigments and rabbit skin glue in reference paint reconstructions.* Journal of Thermal Analysis and Calorimetry, 2015. **122**(1): p. 315-322.
18. Duce, C., L. Ghezzi, M. Onor, I. Bonaduce, M.P. Colombini, M.R. Tine ,E. Bramanti, *Physico-chemical characterization of protein–pigment interactions in tempera paint reconstructions: casein/cinnabar and albumin/cinnabar.* Analytical and bioanalytical chemistry, 2012. **402**(6): p. 2183-2193.
19. Orsini, S., A. Yadav, M. Dilillo, L.A. McDonnell ,I. Bonaduce, *Characterization of degraded proteins in paintings using bottom-up proteomic approaches: new strategies for protein digestion and analysis of data.* Analytical chemistry, 2018. **90**(11): p. 6403-6408.
20. Leo, G., I. Bonaduce, A. Andreotti, G. Marino, P. Pucci, M.P. Colombini ,L. Birolo, *Deamidation at asparagine and glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings.* Analytical chemistry, 2011. **83**(6): p. 2056-2064.
21. Dallongeville, S., N. Garnier, C. Rolando ,C. Tokarski, *Proteins in art, archaeology, and paleontology: from detection to identification.* Chemical reviews, 2016. **116**(1): p. 2-79.
22. Elert, K. ,C. Cardell, *Weathering behavior of cinnabar-based tempera paints upon natural and accelerated aging.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019. **216**: p. 236-248.
23. Orsini, S., F. Parlanti ,I. Bonaduce, *Analytical pyrolysis of proteins in samples from artistic and archaeological objects.* Journal of Analytical and Applied Pyrolysis, 2017. **124**: p. 643-657.
24. Degano, I., F. Modugno, I. Bonaduce, E. Ribechini ,M.P. Colombini, *Recent advances in analytical pyrolysis to investigate organic materials in heritage science.* Angewandte Chemie International Edition, 2018. **57**(25): p. 7313-7323.
25. Orsini, S., E. Bramanti ,I. Bonaduce, *Analytical pyrolysis to gain insights into the protein structure. The case of ovalbumin.* Journal of Analytical and Applied Pyrolysis, 2018. **133**: p. 59-67.
26. Canfield, R.E. ,A.K. Liu, *The disulfide bonds of egg white lysozyme (muramidase).* Journal of biological chemistry, 1965. **240**(5): p. 1997-2002.
27. Hvidt, A. ,S.O. Nielsen, *Hydrogen exchange in proteins*, in *Advances in protein chemistry*. 1966, Elsevier. p. 287-386.
28. Englander, S.W. ,N.R. Kallenbach, *Hydrogen exchange and structural dynamics of proteins and nucleic acids.* Quarterly reviews of biophysics, 1983. **16**(4): p. 521-655.
29. Englander, S., L. Mayne, Y. Bai ,T. Sosnick, *Hydrogen exchange: the modern legacy of Linderstrøm-Lang.* Protein science, 1997. **6**(5): p. 1101-1109.
30. Englander, S., N. Downer ,H. Teitelbaum, *Hydrogen exchange.* Annual review of biochemistry, 1972. **41**(1): p. 903-924.
31. Oganesyan, I., C. Lento ,D.J. Wilson, *Contemporary hydrogen deuterium exchange mass spectrometry.* Methods, 2018. **144**: p. 27-42.
32. Pascal, B.D., S. Willis, J.L. Lauer, R.R. Landgraf, G.M. West, D. Marciano, S. Novick, D. Goswami, M.J. Chalmers ,P.R. Griffin, *HDX workbench: software for the analysis of H/D exchange MS data.* Journal of the American Society for Mass Spectrometry, 2012. **23**(9): p. 1512-1521.

33. Englander, S.W., *Hydrogen exchange and mass spectrometry: A historical perspective.* Journal of the American Society for Mass Spectrometry, 2006. **17**(11): p. 1481-1489.
34. Marcsisin, S.R. ,J.R. Engen, *Hydrogen exchange mass spectrometry: what is it and what can it tell us?* Analytical and bioanalytical chemistry, 2010. **397**(3): p. 967-972.
35. Engen, J.R., *Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS*. 2009, ACS Publications.
36. Masson, G.R., M.L. Jenkins ,J.E. Burke, *An overview of hydrogen deuterium exchange mass spectrometry (HDX-MS) in drug discovery.* Expert opinion on drug discovery, 2017. **12**(10): p. 981-994.
37. Englander, S.W. ,L. Mayne, *The nature of protein folding pathways.* Proceedings of the National Academy of Sciences, 2014. **111**(45): p. 15873-15880.
38. Masson, G.R., J.E. Burke, N.G. Ahn, G.S. Anand, C. Borchers, S. Brier, G.M. Bou-Assaf, J.R. Engen, S.W. Englander ,J. Faber, *Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments.* Nature methods, 2019. **16**(7): p. 595-602.
39. Narang, D., C. Lento ,D.J. Wilson, *HDX-MS: An Analytical Tool to Capture Protein Motion in Action.* Biomedicines, 2020. **8**(7): p. 224.
40. Claesen, J. ,T. Burzykowski, *Computational methods and challenges in hydrogen/deuterium exchange mass spectrometry.* Mass spectrometry reviews, 2017. **36**(5): p. 649-667.
41. Bonnington, L., I. Lindner, U. Gilles, T. Kailich, D. Reusch ,P. Bulau, *Application of hydrogen/deuterium exchange-mass spectrometry to biopharmaceutical development requirements: improved sensitivity to detection of conformational changes.* Analytical Chemistry, 2017. **89**(16): p. 8233-8237.
42. Nevin, A., D. Anglos, S. Cather ,A. Burnstock, *The influence of visible light and inorganic pigments on fluorescence excitation emission spectra of egg-, casein-and collagen-based painting media.* Applied Physics A, 2008. **92**(1): p. 69-76.
43. Nevin, A., I. Osticioli, D. Anglos, A. Burnstock, S. Cather ,E. Castellucci, *The analysis of naturally and artificially aged protein-based paint media using Raman spectroscopy combined with Principal Component Analysis.* Journal of Raman Spectroscopy: An International Journal for Original Work in all Aspects of Raman Spectroscopy, Including Higher Order Processes, and also Brillouin and Rayleigh Scattering, 2008. **39**(8): p. 993-1000.
44. Feeney, R.E. ,J.R. Whitaker, *Importance of cross-linking reactions in proteins.* Advances in cereal science and technology (USA), 1987.
45. Wedemeyer, W.J., E. Welker, M. Narayan ,H.A. Scheraga, *Disulfide bonds and protein folding.* Biochemistry, 2000. **39**(15): p. 4207-4216.
46. Hogg, P.J., *Disulfide bonds as switches for protein function.* Trends in biochemical sciences, 2003. **28**(4): p. 210-214.
47. Betz, S.F., *Disulfide bonds and the stability of globular proteins.* Protein Science, 1993. **2**(10): p. 1551-1558.
48. Heck, T., G. Faccio, M. Richter ,L. Thöny-Meyer, *Enzyme-catalyzed protein crosslinking.* Applied microbiology and biotechnology, 2013. **97**(2): p. 461-475.
49. A Ajith, T. ,P. Vinodkumar, *Advanced glycation end products: association with the pathogenesis of diseases and the current therapeutic advances.* Current clinical pharmacology, 2016. **11**(2): p. 118-127.
50. Monnier, V.M., G.T. Mustata, K.L. Biemel, O. Reihl, M.O. Lederer, D. Zhenyu ,D.R. Sell, *Cross-linking of the extracellular matrix by the maillard reaction in aging and diabetes: an update on "a puzzle nearing resolution".* Annals of the New York Academy of Sciences, 2005. **1043**(1): p. 533-544.
51. Guo, X.-N., X.-M. Wei ,K.-X. Zhu, *The impact of protein cross-linking induced by alkali on the quality of buckwheat noodles.* Food chemistry, 2017. **221**: p. 1178-1185.

52. Guo, X.-N., S. Yang ,K.-X. Zhu, *Influences of alkali on the quality and protein polymerization of buckwheat Chinese steamed bread.* Food chemistry, 2019. **283**: p. 52-58.
53. Gerrard, J., *Protein cross-linking in food.* Food biochemistry and food processing, 2006: p. 223.
54. Leo, G., C. Altucci, S. Bourgoin-Voillard, A.M. Gravagnuolo, R. Esposito, G. Marino, C.E. Costello, R. Velotta ,L. Birolo, *Ultraviolet laser-induced cross-linking in peptides.* Rapid Communications in Mass Spectrometry, 2013. **27**(14): p. 1660-1668.
55. Itri, F., D.M. Monti, B. Della Ventura, R. Vinciguerra, M. Chino, F. Gesuele, A. Lombardi, R. Velotta, C. Altucci ,L. Birolo, *Femtosecond UV-laser pulses to unveil protein–protein interactions in living cells.* Cellular and molecular life sciences, 2016. **73**(3): p. 637-648.
56. Itri, F., D.M. Monti, M. Chino, R. Vinciguerra, C. Altucci, A. Lombardi, R. Piccoli, L. Birolo ,A. Arciello, *Identification of novel direct protein-protein interactions by irradiating living cells with femtosecond UV laser pulses.* Biochemical and biophysical research communications, 2017. **492**(1): p. 67-73.
57. Jiang, Z., C. Wang, T. Li, D. Sun, H. Gao, Z. Gao ,Z. Mu, *Effect of ultrasound on the structure and functional properties of transglutaminase-crosslinked whey protein isolate exposed to prior heat treatment.* International Dairy Journal, 2019. **88**: p. 79-88.
58. Mitra, B., R. Lametsch, I. Greco ,J. Ruiz-Carrascal, *Advanced glycation end products, protein crosslinks and post translational modifications in pork subjected to different heat treatments.* Meat science, 2018. **145**: p. 415-424.
59. Holding, A.N., *XL-MS: Protein cross-linking coupled with mass spectrometry.* Methods, 2015. **89**: p. 54-63.
60. Yu, C. ,L. Huang, *Cross-linking mass spectrometry: an emerging technology for interactomics and structural biology.* Analytical chemistry, 2018. **90**(1): p. 144-165.
61. Iacobucci, C., C. Piotrowski, R. Aebersold, B.C. Amaral, P. Andrews, K. Bernfur, C. Borchers, N.I. Brodie, J.E. Bruce ,Y. Cao, *First community-wide, comparative cross-linking mass spectrometry study.* Analytical chemistry, 2019. **91**(11): p. 6953-6961.
62. Magdeldin, S. ,T. Yamamoto, *Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues.* Proteomics, 2012. **12**(7): p. 1045-1058.
63. Coscia, F., S. Doll, J.M. Bech, L. Schweizer, A. Mund, E. Lengyel, J. Lindebjerg, G.I. Madsen, J.M. Moreira ,M. Mann, *A streamlined mass spectrometry–based proteomics workflow for large-scale FFPE tissue analysis.* The Journal of pathology, 2020. **251**(1): p. 100-112.
64. Cardell, C., A. Herrera, I. Guerra, N. Navas, L.R. Simón ,K. Elert, *Pigment-size effect on the physico-chemical behavior of azurite-tempera dosimeters upon natural and accelerated photo aging.* Dyes and Pigments, 2017. **141**: p. 53-65.
65. Romero-Pastor, J., N. Navas, S. Kuckova, A. Rodríguez-Navarro ,C. Cardell, *Collagen-based proteinaceous binder-pigment interaction study under UV ageing conditions by MALDI-TOF-MS and principal component analysis.* Journal of mass spectrometry, 2012. **47**(3): p. 322-330.
66. Leal, S.S., H.M. Botelho ,C.M. Gomes, *Metal ions as modulators of protein conformation and misfolding in neurodegeneration.* Coordination Chemistry Reviews, 2012. **256**(19-20): p. 2253-2270.
67. Buchert, J., D. Ercili Cura, H. Ma, C. Gasparetti, E. Monogioudi, G. Faccio, M. Mattinen, H. Boer, R. Partanen ,E. Selinheimo, *Crosslinking food proteins for improved functionality.* Annual review of food science and technology, 2010. **1**: p. 113-138.
68. Pattison, D.I., A.S. Rahmanto ,M.J. Davies, *Photo-oxidation of proteins.* Photochemical & Photobiological Sciences, 2012. **11**(1): p. 38-53.
69. Davies, M.J., *Protein oxidation and peroxidation.* Biochem J, 2016. **473**(7): p. 805-25.

70. Hägglund, P., M. Mariotti ,M.J. Davies, *Identification and characterization of protein cross-links induced by oxidative reactions.* Expert Review of Proteomics, 2018. **15**(8): p. 665-681.
71. Hawkins, C.L. ,M.J. Davies, *Generation and propagation of radical reactions on proteins.* Biochimica et Biophysica Acta (BBA)-Bioenergetics, 2001. **1504**(2-3): p. 196-219.
72. Karam, L.R., M. Dizdaroglu ,M.G. Simic, *OH radical-induced products of tyrosine peptides.* International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine, 1984. **46**(6): p. 715-724.
73. Arenas, A., C. López-Alarcón, M. Kogan, E. Lissi, M.J. Davies ,E. Silva, *Chemical modification of lysozyme, glucose 6-phosphate dehydrogenase, and bovine eye lens proteins induced by peroxyl radicals: role of oxidizable amino acid residues.* Chemical Research in Toxicology, 2013. **26**(1): p. 67-77.
74. Davies, M.J., *Reactive species formed on proteins exposed to singlet oxygen.* Photochemical & Photobiological Sciences, 2004. **3**(1): p. 17-25.
75. Saito, M. ,K. Marumo, *Collagen cross-links as a determinant of bone quality: a possible explanation for bone fragility in aging, osteoporosis, and diabetes mellitus.* Osteoporosis international, 2010. **21**(2): p. 195-214.
76. Leinisch, F., M. Mariotti, M. Rykaer, C. Lopez-Alarcon, P. Hägglund ,M.J. Davies, *Peroxyl radical-and photo-oxidation of glucose 6-phosphate dehydrogenase generates cross-links and functional changes via oxidation of tyrosine and tryptophan residues.* Free Radical Biology and Medicine, 2017. **112**: p. 240-252.
77. Giulivi, C., N. Traaseth ,K. Davies, *Tyrosine oxidation products: analysis and biological relevance.* Amino acids, 2003. **25**(3-4): p. 227-232.
78. Wright, A., W.A. Bubb, C.L. Hawkins ,M.J. Davies, *Singlet Oxygen–mediated Protein Oxidation: Evidence for the Formation of Reactive Side Chain Peroxides on Tyrosine Residues*¶. Photochemistry and photobiology, 2002. **76**(1): p. 35-46.
79. Medinas, D.B., F.C. Gozzo, L.F. Santos, A.H. Iglesias ,O. Augusto, *A ditryptophan cross-link is responsible for the covalent dimerization of human superoxide dismutase 1 during its bicarbonate-dependent peroxidase activity.* Free Radical Biology and Medicine, 2010. **49**(6): p. 1046-1053.
80. Paviani, V., R.F. Queiroz, E.F. Marques, P. Di Mascio ,O. Augusto, *Production of lysozyme and lysozyme-superoxide dismutase dimers bound by a ditryptophan cross-link in carbonate radical-treated lysozyme.* Free Radical Biology and Medicine, 2015. **89**: p. 72-82.
81. Schöneich, C., *Novel chemical degradation pathways of proteins mediated by tryptophan oxidation: tryptophan side chain fragmentation.* Journal of Pharmacy and Pharmacology, 2018. **70**(5): p. 655-665.
82. Carroll, L., D.I. Pattison, J.B. Davies, R.F. Anderson, C. Lopez-Alarcon ,M.J. Davies, *Formation and detection of oxidant-generated tryptophan dimers in peptides and proteins.* Free Radical Biology and Medicine, 2017. **113**: p. 132-142.
83. Paviani, V., G.T. Galdino, J.N.d. Prazeres, R.F. Queiroz ,O. Augusto, *Ditryptophan cross-links as novel products of protein oxidation.* Journal of the Brazilian Chemical Society, 2018. **29**(5): p. 925-933.
84. Bhaskar, B., C.E. Immoos, H. Shimizu, F. Sulc, P.J. Farmer ,T.L. Poulos, *A novel heme and peroxide-dependent tryptophan–tyrosine cross-link in a mutant of cytochrome c peroxidase.* Journal of molecular biology, 2003. **328**(1): p. 157-166.
85. Agon, V.V., W.A. Bubb, A. Wright, C.L. Hawkins ,M.J. Davies, *Sensitizer-mediated photooxidation of histidine residues: evidence for the formation of reactive side-chain peroxides.* Free Radical Biology and Medicine, 2006. **40**(4): p. 698-710.

86. Leinisch, F., M. Mariotti, P. Hägglund ,M.J. Davies, *Structural and functional changes in RNAse A originating from tyrosine and histidine cross-linking and oxidation induced by singlet oxygen and peroxyl radicals.* Free Radical Biology and Medicine, 2018. **126**: p. 73-86.
87. Xu, C.-F., Y. Chen, L. Yi, T. Brantley, B. Stanley, Z. Sosic ,L. Zang, *Discovery and characterization of histidine oxidation initiated cross-links in an IgG1 monoclonal antibody.* Analytical chemistry, 2017. **89**(15): p. 7915-7923.
88. Sinz, A., *Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here?* Angew Chem Int Ed Engl, 2018. **57**(22): p. 6390-6396.
89. Maes, E., V. Broeckx, I. Mertens, X. Sagaert, H. Prenen, B. Landuyt ,L. Schoofs, *Analysis of the formalin-fixed paraffin-embedded tissue proteome: pitfalls, challenges, and future prospectives.* Amino acids, 2013. **45**(2): p. 205-218.
90. Klockenbusch, C., J.E. O'Hara ,J. Kast, *Advancing formaldehyde cross-linking towards quantitative proteomic applications.* Anal Bioanal Chem, 2012. **404**(4): p. 1057-67.
91. Quesada-Calvo, F., V. Bertrand, R. Longuespée, A. Delga, G. Mazzucchelli, N. Smargiasso, D. Baiwir, P. Delvenne, M. Malaise ,M.-C. De Pauw-Gillet, *Comparison of two FFPE preparation methods using label-free shotgun proteomics: Application to tissues of diverticulitis patients.* Journal of proteomics, 2015. **112**: p. 250-261.
92. Coscia, F., S. Doll, J.M. Bech, A. Mund, E. Lengyel, J. Lindebjerg, G.I. Madsen, J.M. Moreira ,M. Mann, *A streamlined mass spectrometry-based proteomics workflow for large scale FFPE tissue analysis.* bioRxiv, 2019: p. 779009.
93. Singh, H., *Modification of food proteins by covalent crosslinking.* Trends in Food Science & Technology, 1991. **2**: p. 196-200.
94. Gerrard, J.A., *Protein–protein crosslinking in food: methods, consequences, applications.* Trends in food science & technology, 2002. **13**(12): p. 391-399.
95. McKerchar, H.J., S. Clerens, R.C. Dobson, J.M. Dyer, E. Maes ,J.A. Gerrard, *Protein-protein crosslinking in food: Proteomic characterisation methods, consequences and applications.* Trends in Food Science & Technology, 2019. **86**: p. 217-229.
96. Oryan, A., A. Kamali, A. Moshiri, H. Baharvand ,H. Daemi, *Chemical crosslinking of biopolymeric scaffolds: Current knowledge and future directions of crosslinked engineered bone scaffolds.* International journal of biological macromolecules, 2018. **107**: p. 678-688.
97. Parenteau-Bareil, R., R. Gauvin ,F. Berthod, *Collagen-based biomaterials for tissue engineering applications.* Materials, 2010. **3**(3): p. 1863-1887.
98. Rose, J.B., S. Pacelli, A.J.E. Haj, H.S. Dua, A. Hopkinson, L.J. White ,F.R. Rose, *Gelatin-based materials in ocular tissue engineering.* Materials, 2014. **7**(4): p. 3106-3135.
99. Yung, C., L. Wu, J. Tullman, G. Payne, W. Bentley ,T. Barbari, *Transglutaminase crosslinked gelatin as a tissue engineering scaffold.* Journal of Biomedical Materials Research Part A: An Official Journal of The Society for Biomaterials, The Japanese Society for Biomaterials, and The Australian Society for Biomaterials and the Korean Society for Biomaterials, 2007. **83**(4): p. 1039-1046.
100. Maes, E., J.M. Dyer, H.J. McKerchar, S. Deb-Choudhury ,S. Clerens, *Protein-protein cross-linking and human health: the challenge of elucidating with mass spectrometry.* Expert Review of Proteomics, 2017. **14**(10): p. 917-929.
101. Swomley, A.M., S. Förster, J.T. Keeney, J. Triplett, Z. Zhang, R. Sultana ,D.A. Butterfield, *Abeta, oxidative stress in Alzheimer disease: evidence based on proteomics studies.* Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2014. **1842**(8): p. 1248-1257.
102. Linetsky, M., J. Hill, R.D. LeGrand ,F. Hu, *Dehydroalanine crosslinks in human lens.* Experimental eye research, 2004. **79**(4): p. 499-512.

103. Sotiropoulou, S., G. Sciutto, A.L. Tenorio, J. Mazurek, I. Bonaduce, S. Prati, R. Mazzeo, M. Schilling ,M.P. Colombini, *Advanced analytical investigation on degradation markers in wall paintings.* Microchemical Journal, 2018. **139**: p. 278-294.

104. Linn, R., I. Bonaduce, G. Ntasi, L. Birolo, A. Yasur-Landau, E.H. Cline, A. Nevin ,A. Lluveras-Tenorio, *Evolved Gas Analysis-Mass Spectrometry to Identify the Earliest Organic Binder in Aegean Style Wall Paintings.* Angewandte Chemie, 2018. **130**(40): p. 13441-13444.
105. Buncherd, H., W. Roseboom, B. Ghavim, W. Du, L.J. de Koning, C.G. de Koster ,L. de Jong, *Isolation of cross-linked peptides by diagonal strong cation exchange chromatography for protein complex topology studies by peptide fragment fingerprinting from large sequence databases.* Journal of Chromatography A, 2014. **1348**: p. 34-46.
106. Maiolica, A., D. Cittaro, D. Borsotti, L. Sennels, C. Ciferri, C. Tarricone, A. Musacchio ,J. Rappsilber, *Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching.* Molecular & Cellular Proteomics, 2007. **6**(12): p. 2200-2211.
107. Chen, Z.A., A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J.C. Bukowski-Wills ,M. Nilges, *Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry.* The EMBO journal, 2010. **29**(4): p. 717-726.
108. Tinnefeld, V., A.S. Venne, A. Sickmann ,R.P. Zahedi, *Enrichment of cross-linked peptides using charge-based fractional diagonal chromatography (ChaFRADIC).* Journal of proteome research, 2017. **16**(2): p. 459-469.
109. Mariotti, M., F. Leinisch, D.J. Leeming, B. Svensson, M.J. Davies ,P. Hägglund, *Mass-spectrometry-based identification of cross-links in proteins exposed to photo-oxidation and peroxyl radicals using 18O labeling and optimized tandem mass spectrometry fragmentation.* Journal of proteome research, 2018. **17**(6): p. 2017-2027.
110. Santos, L.F., A.H. Iglesias ,F.C. Gozzo, *Fragmentation features of intermolecular cross-linked peptides using N-hydroxy-succinimide esters by MALDI-and ESI-MS/MS for use in structural proteomics.* Journal of mass spectrometry, 2011. **46**(8): p. 742-750.
111. Paizs, B. ,S. Suhai, *Fragmentation pathways of protonated peptides.* Mass spectrometry reviews, 2005. **24**(4): p. 508-548.
112. Lebendiker, M. ,T. Danieli, *Production of prone-to-aggregate proteins.* FEBS letters, 2014. **588**(2): p. 236-246.
113. Srinivasa, S., X. Ding ,J. Kast, *Formaldehyde cross-linking and structural proteomics: Bridging the gap.* Methods, 2015. **89**: p. 91-8.
114. Schilling, B., R.H. Row, B.W. Gibson, X. Guo ,M.M. Young, *MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides.* Journal of the American Society for Mass Spectrometry, 2003. **14**(8): p. 834-850.
115. Götze, M., J. Pettelkau, S. Schaks, K. Bosse, C.H. Ihling, F. Krauth, R. Fritzsche, U. Kühn ,A. Sinz, *StavroX—a software for analyzing crosslinked products in protein interaction studies.* Journal of the American Society for Mass Spectrometry, 2011. **23**(1): p. 76-87.
116. Iacobucci, C., M. Götze, C.H. Ihling, C. Piotrowski, C. Arlt, M. Schäfer, C. Hage, R. Schmidt ,A. Sinz, *A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions.* Nature Protocols, 2018. **13**(12): p. 2864-2889.
117. Rey, M., V. Sarpe, K.M. Burns, J. Buse, C.A. Baker, M. van Dijk, L. Wordeman, A.M. Bonvin ,D.C. Schriemer, *Mass spec studio for integrative structural biology.* Structure, 2014. **22**(10): p. 1538-1548.

118. Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. De Bakker ,M.J. Daly, *PLINK: a tool set for whole-genome association and population-based linkage analyses.* The American journal of human genetics, 2007. **81**(3): p. 559-575.
119. Hoopmann, M.R., A. Zelter, R.S. Johnson, M. Riffle, M.J. MacCoss, T.N. Davis ,R.L. Moritz, *Kojak: efficient analysis of chemically cross-linked protein complexes.* Journal of proteome research, 2015. **14**(5): p. 2190-2198.
120. Sarpe, V., A. Rafiei, M. Hepburn, N. Ostan, A.B. Schryvers ,D.C. Schriemer, *High sensitivity crosslink detection coupled with integrative structure modeling in the Mass Spec Studio.* Molecular & Cellular Proteomics, 2016. **15**(9): p. 3071-3080.
121. Hernandez, P., R. Gras, J. Frey ,R.D. Appel, *Popitam: towards new heuristic strategies to improve protein identification from tandem mass spectrometry data.* Proteomics, 2003. **3**(6): p. 870-878.
122. Bonaventura, C., J. Bonaventura, R. Stevens ,D. Millington, *Acrylamide in polyacrylamide gels can modify proteins during electrophoresis.* Analytical biochemistry, 1994. **222**(1): p. 44-48.
123. Yan, J.X., W.C. Kett, B.R. Herbert, A.A. Gooley, N.H. Packer ,K.L. Williams, *Identification and quantitation of cysteine in proteins separated by gel electrophoresis.* Journal of Chromatography A, 1998. **813**(1): p. 187-200.
124. Mukherjee, S., E.A. Kapp, A. Lothian, A.M. Roberts, Y.V. Vasil'ev, B.A. Boughton, K.J. Barnham, W.M. Kok, C.A. Hutton ,C.L. Masters, *Characterization and identification of dityrosine cross-linked peptides using tandem mass spectrometry.* Analytical chemistry, 2017. **89**(11): p. 6136-6145.
125. Annibal, A., G. Colombo, A. Milzani, I. Dalle-Donne, M. Fedorova ,R. Hoffmann, *Identification of dityrosine cross-linked sites in oxidized human serum albumin.* Journal of Chromatography B, 2016. **1019**: p. 147-155.
126. Coelho, F.R., A. Iqbal, E. Linares, D.F. Silva, F.S. Lima, I.M. Cuccovia ,O. Augusto, *Oxidation of the tryptophan 32 residue of human superoxide dismutase 1 caused by its bicarbonate-dependent peroxidase activity triggers the non-amyloid aggregation of the enzyme.* Journal of Biological Chemistry, 2014. **289**(44): p. 30690-30701.

Chapter V Study of protein cross-linking using MS: molecular evidence of restoration treatments applied to historical manuscripts

Abstract

This chapter addresses the chemical and structural modifications that can occur in proteins from historical and artistic artworks following anthropogenic causes, such as introducing reactive products during conservation treatments. The investigation of these degradation products helps in providing a more comprehensive vision of the artwork and its conservation history. In this chapter, the proteomic analyses were extended beyond the classical bottom-up protein identification, investigating the chemical and structural modifications of the proteinaceous compounds induced in a manuscript previously subjected to invasive restoration treatments.

The application of an unbiased modification search followed by a structural study of reticulated peptide pairs provided the first analytical proofs of the use of formaldehyde-based treatment during parchment conservation treatments in the Vatican library. The research has also demonstrated that the more comprehensive approach can refine the results' interpretation by reaching data that remain undetected in the standard method because they are chemically modified or structurally unreachable.

1. Introduction

Besides natural causes (such as environmental agents), proteinaceous compounds in artworks might undergo chemical and structural modifications following anthropogenic factors, such as applying reactive products in restoration treatments. Today, the reversibility and compatibility of the material implemented have become fundamental points in the modern Code of Ethics and Rules of Conduct to the conservation of artistic works.

Nonetheless, often in the past, the immediate effect was more pursued than long-term survival [1, 2]. The application of reactive materials and their subsequent interactions with the original organic compounds might induce physical and chemical degradation and prevent accurate analyte detection and investigation [3]. The research and definition of the degradation products having an anthropogenic origin could help understand the object's conservation state better and reach the hidden information that might lead to a more comprehensive vision of the artwork and its conservation history.

This chapter will discuss the vast potentialities of a proteomic protocol that combine the classical identification of proteins with the comprehensive examination of chemical and structural modifications to provide a more comprehensive study of artwork previously subjected to invasive restoration treatments. The developed protocol will be detailed in application to a parchment manuscript study.

2. Parchment

2.1 General aspects

Parchment is a thin, flat biomaterial produced from various animals' skins, principally calf, goat and sheep or less commonly, antelope, muflan, and gazelle, appropriately treated to arrest the processes of decomposition [4, 5]. Vellum, from french *vélin* (calf vellum), is frequently implemented to define the fine parchment. Its origins were formerly ascribed to the ancient city of Pergamum in the 2nd century BC; the production date was redefined as earlier following the discovery of Dead Sea Scrolls in Qumran's caves dated 960 - 586 BC [6-8]. By replacing the papyrus, parchment became the most widely used writing medium for many centuries until the advent of paper from the 12th century onward [9-11].

The manufacturing differed depending on the geographical area, historical period or intended final use.

In antiquity, the material was often formerly softened with natural products such as vegetable oils, animal fat and milk, then dehaired and lightly tanned with vegetable matter, wine or other methods [12].

Vegetable tanning agents are composed of natural polyphenols that are principally extracted from plants, roots or bark. Contrarily to more modern tanning products, typically implemented in leather manufacturing (such as aldehyde and chrome compounds), these compounds do not covalently cross-link with collagen proteins, stabilising the protein structure instead with the formation of hydrogen bonds [13]. However, different parchment-making traditions were identified even in the same period; for example, Jewish manuscripts from the Second Temple Era were produced following two traditional productions depending on the geographical area (west and east), which differ as to the presence or absence of tanning [14]. Later, during the Roman empire, parchment was commonly dehaired with urine and ashes, smoothed with pumice powder or chalk and then finished with inorganic lime or with organic compounds such as egg whites milk [4, 5, 15]. A liming step, consisting of an alkaline limestone bath (calcium hydroxide), was introduced in the eighth century AD to weaken hair and epidermis and facilitate their removal. The use of this procedure increased throughout Western Europe during the Early Middle Ages, where it is still in use today [5, 16]. Approximately with the onset of liming, tanning disappeared in the manufacture of parchment, establishing the current fundamental difference with the leather production.

In the final parchment production steps, the skin was stretched on a wooden frame and scraped to remove the outer layers (epidermis and hypodermis), leaving the dermis and corium [17]. These layers are principally composed (approximately 80%) of collagen type I (COL I), a heterometric helix constituted by two collagen alpha-1(I) and one collagen alpha-2(I). The 20% is instead composed of collagen type III (COL III), which is a homometric helix of three collagen alpha-1(III) peptide chains [18]. Traces of collagen type IV, V, VI, and XI [19-21], elastin, and keratins can also be present [22]. The amino acid sequence of collagen proteins is constituted by a repetition of three peptides, glycine-X-X_2 (where X and X_2 are often proline and hydroxyproline). Collagen naturally forms polymers hierarchically structured: collagen molecules are composed of three polypeptidic chains coiled together to form a triple helix that is twisted into fibrils and fibres (Figure V-1) [23-25].

The tensile strength and flexibility features are principally due to the specific alignment and packing of fibrils related to the hierarchical structure and the interactions with non-collagenous compounds, called proteoglycans.

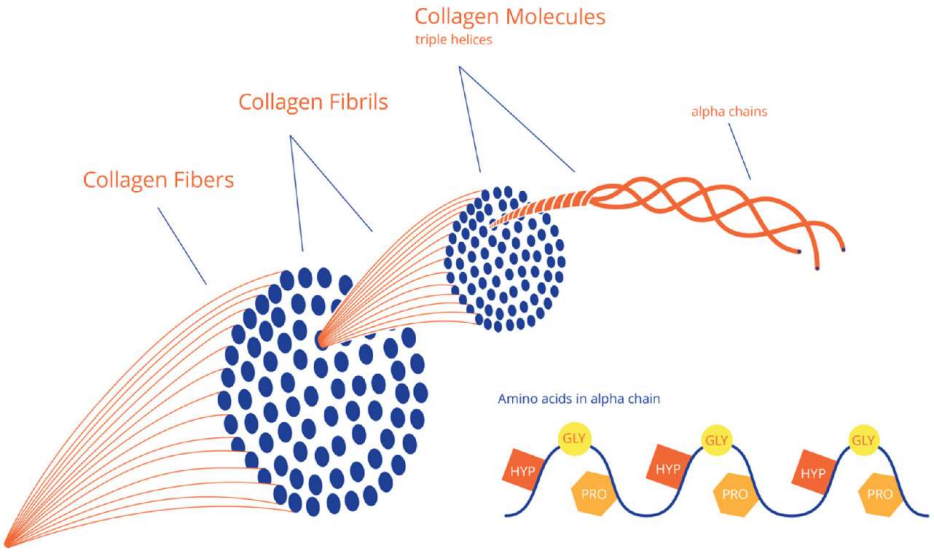


Figure V-1 The hierarchical structure of collagen: the amino acid sequence is composed of a repetition of three amino acid residues (glycine-X-X2, which often are proline and hydroxyproline). The collagen molecules then are composed of three polypeptidic chains coiled together to form a triple helix twisted in fibrils and then in fibres. Figure extracted from Nijhuis et all., 2019.

Collagen molecules are also staggered axially compared to the closer molecules of a specific d spacing, which is a gap between the molecules linearly adjacent at each repetition. The value of 67 nm has been widely adopted for the D-spacing of type I collagen fibril [26].

Parchment-making processes induce several chemical and structural modifications at the molecular level to influence mechanical properties and produce more resistant and durable material than the original skin [27]. The liming leads to a general reduction of the chemical stability, “opening up” the fibre structure at the molecular level, swelling of fibres, and breaking interactions with proteoglycans. Furthermore, d spacing is reduced while the residual distance remains unchanged [28]. During the stretching into the frame, the 3D network characteristic of skin is forced to arrange into a 2D network with layers parallel to the surface, increasing the material's strength and durability [19, 29-31]. During ageing, parchment's principal chemical degradations are oxidation, gelatinisation (denaturation), and hydrolysis. Depending on the storage environment, various physical and biological degradations might also impact collagen-based materials [32].

2.2 Analytical techniques implemented in parchment analysis

A large panel of different analytical techniques has been implemented to investigate parchment in ancient codes and documents to investigate the artefact manufacturing, history, ageing, storage conditions and degradation processes, assisting, in turn, the elaboration of adequate restoration treatments and storage conditions. Visual examinations implementing transmitted and raking lights and microscope magnification are often performed to have a first general overview.

These observations may highlight particular features like manufacturing marks (caused by scraping, shaving and other activities), surface treatment products, and evidence of the animal disease (skin lesions, scars, bruises, holes). These earlier observations could give preliminary information on animal breeding, health and treatments. The skin thickness inspection, together with anatomical feature measurements, can estimate the size and age of the animal and the part of the body the skin was cut, also providing a general idea of the manufacturing [5, 33, 34]. The examination of follicle patterns was widely applied to distinguish the animal species [4, 35-37]; this technique is based on subjective impressions. It cannot always be conducted due to the occasional complete removal of follicles in the parchment fabrication.

Non-invasive and non-destructive analyses have significantly been implemented to study different organic and inorganic components present in parchment artefacts. For instance, UV-VIS, FTIR and Raman spectroscopies have been applied to investigate organic dyes [38-40], inks [41], vegetable materials used for tanning [42] or to distinguish parchment support from other proteinaceous compounds such as egg white and casein glue [43, 44]. XRF (X-ray fluorescence) spectroscopy was also performed to identify particular compounds implemented in parchment making, such as lime or clay [45].

NMR-MOUSE (nuclear magnetic resonance–mobile universal surface explorer) has been used to analyse parchments [46-48]. Furthermore, portable NMR instruments have been used to study historical parchments' degradation (mainly gelatinisation and hydrolytic processes) [49].

Biological information such as accurate species identification is particularly interesting in manuscript studies since it gives insight into livestock exploits or commercial trade. In some instances, it can clarify widely debated manufacturing techniques, such as the previous belief of the use of the hides of small mammals, like rabbit or squirrels, in specific manuscript productions, which now has been proven to be incorrect [50].

Typically spectroscopic techniques cannot provide these details; nonetheless, a recent study based on optical spectroscopy combined with principal component analysis (PCA) successfully identified different biological species in parchments by examining the energy conservation of parchment in the light scattering phenomena, without recourse to molecular level analysis [31].

Species recognition in parchment has also been successfully achieved, implementing PCA with ToF-SIMS data sets [51]. High-throughput biomolecular techniques such as proteomics, genomics and metagenomics are the best techniques to achieve accurate biological details. They were less implemented for parchment investigation until recently because of the need for a large amount of sample hardly achievable in precious artefacts. Instrumental improvements together with the development of non-invasive sampling methods (for example, the analysis of the PVC scraps collected from surface cleaning) boosted proteins [50] and DNA [52] studies, leading to the creation of the new field of biocodicology [33, 53, 54].

The examination of ancient DNA (aDNA) retrieved from parchments demonstrated the ability to learn valuable information on the species and sex of the animal, on breed variations, kinship, and ancient migrations [55, 56]. Successful studies on short fragments of mitochondrial aDNA amplified using PCR have been conducted for the investigation of several writing records from all eras: Dead Sea scrolls [8, 57], from ancient New Zealand [58] and Greek cultures [59], mediaeval [60] and more modern [61] parchments. The need to overcome the principal limit of PCR of choosing the long and less damaged contaminant fragments over endogenous DNA [62, 63] prompted the increasing application of high throughput DNA sequencing techniques (HTS) for ancient studies, such as Next Generation Sequencing (NGS) technologies [64-66], which also proved to produce more and better genetic data. The drawbacks of the analysis are the exogenous DNA contamination, the need for large amounts of sample, the cost and time required for the investigation and the fragile preservation condition of aDNA, the process of which is still not completely clear. Various researches have also been carried out on ancient microbiomes (bacteria and fungi), and DNA from plants grown on parchment supports to gain further information on the artefact history and conservation [52, 67-69].

Proteomics, complementary to DNA studies, is a critical tool for the comprehensive examination of parchment: the study of proteins, which persist longer than DNA, provides unambiguous animal species identification, also ensuring a tissue specificity [33]. ZooMS, based on the observation of peptide markers using peptide mass fingerprinting (PMF) analysis with a MALDI-ToF analyser, has been frequently implemented for parchment studies [9, 70-72]. This analysis recently optimised into eZooMS [45, 50, 72, 73] with the development of a non-invasive sampling with triboelectric protein recovering, also offers the establishment of Parchment Quality Index (PQI). The value is defined through the calculation of deamidation extent induced through the ageing and manufacturing processes (such as lime bath), and it is expressed in percentage where 100% means no deamidation and 0% complete deamidation [33, 50].

The development of an automated approach has recently improved the interpretation of MALDI-ToF spectra concerning the collagen of the three mammals *Ovis aries, Bos taurus* and *Capra hircus* (highly implemented in parchment production) that are closely related one to the other, also taking into account the possible posttranslational modifications [73].

LC-MS / MS studies were also conducted on parchment documents, potentially extending the examinations, in certain circumstances, to the health status of the author of the writings [74-77].

In recent years, an emerging interest is given in understanding the alterations that parchments undergo during manufacturing [78] and later, during the ageing in particular storage conditions, characterised by biological [79], physical [80, 81] and chemical degradations [7, 29, 82].

3. Impact of ancient restoration treatments on the analysis results

Results of analytical studies can be hindered or altered by the formation of various protein PTMs during the ageing and the mixing of original material with external compounds added during later restoration treatments. During parchment conservation, restorers implement a wide variety of collagen-based glues as consolidants, sizes and adhesive (resumed in Table V-1). The specific application of the glue was often left vague in the documents of restoration (for example, Franz Ehrle, whose work will be later discussed, defined in his written the application of a French glue, without specifying the origin [83]).

Furthermore, the composition of the glue itself is rarely known due to “trade secrets”. Additives, like mercuric chloride or aldehydes, were frequently added to these glues to prevent microbiological degradation. The products, although, were often introduced without regard to the potential reactivity with the original material [84, 85].

Table V-1 Principal collagen-based glue implemented in the restoration treatments

Collagen-based glue	Principal characteristics
Parchment glue	Made from small pieces of parchment boiled (different animal origins). Principally implemented to strengthen weak parchment and to patch and join parchments together.
Rabbit skin glue	Theoretically manufactured purely from rabbit connective tissue; however, collagen waste from various small mammals is often added. Still today, this glue is often sold mixed with bovine hide to reach better mechanical properties (rabbit glue is elastic but fragile). The information on the biological source and addition are not always provided.
Hide glue	Produced from the skin of bovine and other smaller mammals. Widely implemented for the consolidation and rebinding of parchment manuscripts.
Bone glue	It is principally manufactured from bones, sinew and cartilage of cattle or pig. Often it refers to unpurified glues having degradation products, such as minerals and fats.
Gelatine	It is the purified form of denatured collagen (from either animal skin or bones) through treatment with lime or acetic acid and subsequently with heat. Final gelatine products from acid and basic treatments differ in properties [16].
Fish glue	Principally there are two types of fish glue: the isinglass glue, produced by the swim fish bladders, is considered better quality. Often obtained from sturgeon (*Acipenseridae sp.*), it is also called sturgeon glue. The second glue is produced from fish waste (head, skin and skeleton), haddock (*Melanogrammus aeglefinus sp.*), cod (*Gadus sp.*), and mackerel (*Scomber scombrus sp.*).The fish glues were widely used to paint and gild codices and as adhesive and consolidant during the restoration treatments, principally due to its lower gelation temperature.
Goldbeater's skin	Manufactured from the outer or peritoneal coat of the cæcum of cattle. It is a thin membrane principally composed of elastine, which is often exploited to repair weak areas of parchments sheets due to its transparency and high tensile strength [86].

3.1 Example of an invasive restoration treatment: gelatin-formol

An example of particularly invasive restoration treatment, highly implemented during the end of the 19th century beginning of the 20th century, consisted of applying gelatine mixed with formol (40% solution of formaldehyde in water) directly on parchment or using Japanese paper. The treatment's objectives were the reinforcement and the consolidation of damaged parchments and the reparation of holes caused by different degradation actions, including the corrosion of iron gall inks and pigments [87, 88]. This technique was principally adopted and promoted by the restorers of Vatican Library, especially by Carlo Marrè, the official restorer of codices in the Library (from 1904), and by Father Franz Ehrle, the Prefect (from 1895 to 1914) and then Cardinal librarian to the Holy Roman Church (from 1929 to 1934).

Franz Ehrle officially presented the treatment during the first "International Conference for Preservation, and Conservation Access of Antique Manuscripts" that occurred between the 30th September and 1st October 1898 in St.Gallen (Swiss) [88].

In the same year, he also published a detailed description of the procedure [83]. Liquid gelatine from animal origin needed to be mixed with a correct amount of formalin to minimise the collagen reactivity to moisture, preventing the biological putrefaction and reducing the influence of temperature changes [89]. The holes were then filled with consecutive brush strokes of liquid gelatine applied on the top of each other (once slight dry) until reaching the thickness of the surrounding parchment. Bigger holes were filled with parchment pieces covered by gelatine, while fissures and cracks were tightened together with parchment strips and consolidated with gelatine layers. Besides the punctual application of gelatine in parchment gaps, the application of a light layer of the mixture to the entire side of the sheet was suggested in the presence of highly corroded parchments.

F. Ehrle was highly confident in the safety and durability of his protocol of consolidation. He ensured the immutability of gelatine with the help of different experts and examining the conservation state of numerous photo negatives left in a layer of gelatine (without formol) for 12-15 years. Furthermore, he was convinced that gelatine/formol was easy to remove from the old scrip without the slightest damage.

However, the technique was highly criticised by other conservators and experts of his time, describing it as highly risky for long-term parchment conservation [90, 91]. It was proven that the gelatin could penetrate irreversibly in the support, migrating into the cavities of the collagen structure and causing a reduction of parchment's solubility as well as an increase of its brittleness [92, 93].

4. Formaldehyde reaction with proteins

4.1 Formaldehyde action on collagen proteins

The parchment rigidity consequent to the application of gelatin-formol treatment most likely is related to the formaldehyde presence: the molecule can induce various cross-linking formations among the collagen proteins.

Supposedly, Franz Herle was unaware of the fixative and cross-linking potentialities of formaldehyde, describing it only as an antimicrobial agent. Only several years later, this molecule's impact started to be extensively investigated on collagen-based materials, particularly in terms of physical and mechanical alterations [94]. Studies showed that the formaldehyde application and the consequent formation of covalent cross-links impact proteins' orientation, increasing the collagen strength and its compressive and Young modulus (i.e. is a measure of a material's resistance to changes in length under longitudinal tension or compression) [20, 95]. The tensile strength is also enhanced due to the higher polymeric matrix cohesion, while the permeability and solubility decrease due to the reduction of free volume in the gelatine matrix [96, 97]. Thermal studies have proven that formaldehyde introduction and the formation of covalent bonds induce endothermic breakages of hydrogen bonds, increasing the melting point and the thermal stability of the material [98].

Formaldehyde cross-links, however, also causes a reduction of elasticity in the biomaterial, rendering it stiff, brittle and easy to fracture [99, 100]. The influence of formaldehyde in collagen materials has been further investigated by FT-IR studies that examined the changes in the collagen spectrum's vibrational bands following a sequential formaldehyde exposure. The observations remarked that the cross-linking induction is relatively rapid, and the prolonged exposition is principally characterised by the aldehyde’s migration into the matrix (with an increment of the chemical modifications). The reactions persist until the saturation of modification products, followed by an accumulation of unreacted formaldehyde in the material [101]. Kinetics studies were done on 13C labelled formaldehyde and monitored by 13C NMR: the findings showed that formaldehyde interacts with gelatin through the initial formation of amine methylol of lysine arginine residues. Cross-linking may then occur between lysine and arginine or between arginine and arginine, but the details of these cross-links are still not fully understood [97, 101-103].

4.2 Formaldehyde as a cross-linker: theory and mechanism

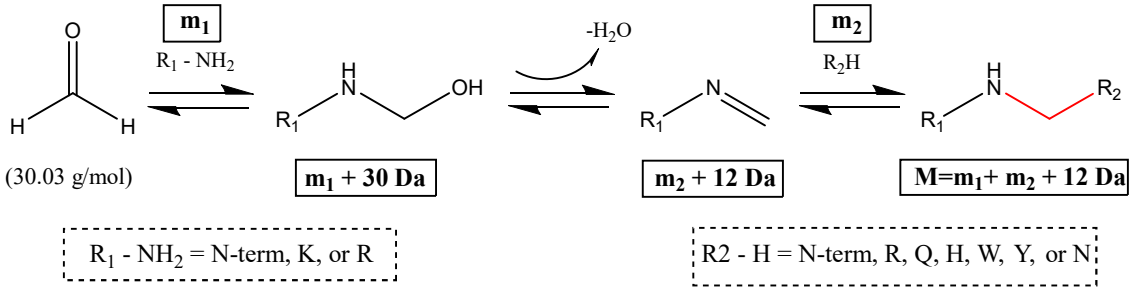
From a molecular perspective, formaldehyde is constituted by two equal hydrogen reactive sites bridged by one carbon atom, which enable its action as a monofunctional cross-linker towards proteins and DNA. Either inter or intramolecular cross-links might be established between molecules in spatial proximity (2.3 to 2.7 Å) [104].

Despite the numerous studies performed on single amino acids [105], model peptides [106, 107] or small proteins [108], the interpretation of cross-linking mechanisms is still controversial. The older and more prevalent theory, illustrated in Figure V-2-A, suggests a two-step reaction resulting in the formation of a methylene bridge (R_1-CH_2-R_2) with the mass increasing equal to 12 Da (1C) [109-111].

In the first stage (Figure V-2-B), formaldehyde reacts with a nucleophilic group (i.e., basic side-chains on a protein), often constituted by a primary amine, such as free N-terminus or ε-amino group of lysine or arginine. This reaction causes the formation of a methylol group (-CH2OH with +30 Da), then, following the loss of a molecule of water (dehydration), a labile Schiff base (an imine) is produced (-N= CH2 with +12 Da). At this stage, the microenvironment of amino acids plays a significant part, encouraging or hampering the formaldehyde reactivity. N-terminal residues and the side chain of lysine and arginine were suggested as the primary reactive site for the formation of an imine. The latter may also be involved in the competitive formation of acetals or hydroxymethyl groups (such as 4-imidazolidinone adduct) [112, 113].

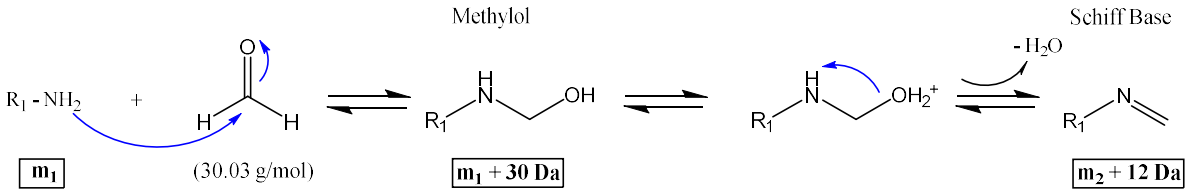
In the second reaction step (Figure V-2- C), the Schiff base's methylene carbon undergoes a nucleophilic attack by a nucleophile on another closer amino acid residue, producing an intra- or inter-protein methylene bridge between two amino acid residues (-CH2- with +12 Da). The formaldehyde's semi-specificity enables the reaction of the molecule with several amino acids residues having amine, amide, guanidyl, phenol, imidazole, or indole groups. The side chains of arginine (R), tyrosine (Y), histidine (H), tryptophan (W), asparagine (N) and glutamine (G) and the N-term were identified as potential second reactive sites in the formation of methylene bridges.

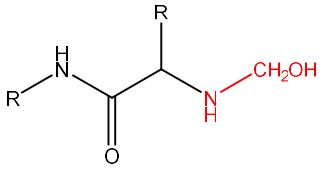
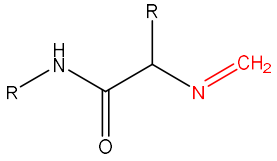
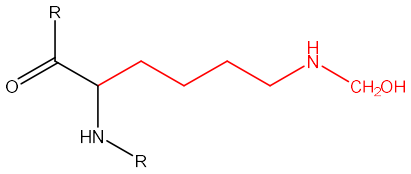
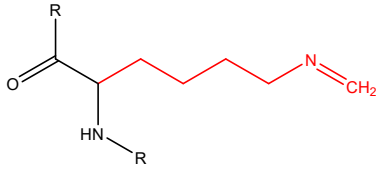
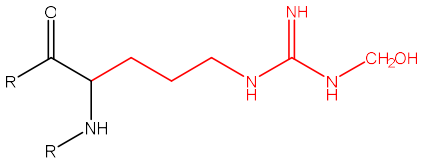
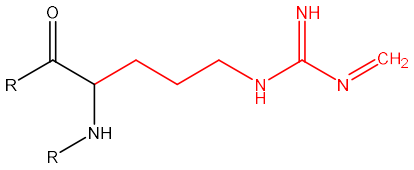
A



1st step

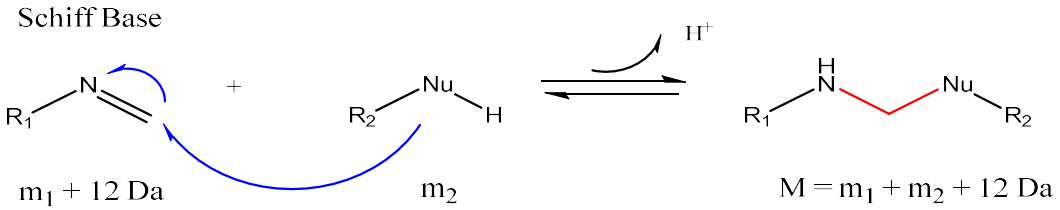
B



m1: R1 – NH2	Chemical structure of Methylol	Chemical structure of Schiff Base
N-term		
Lysine (K)		
Arginine (R)		

2nd step

C



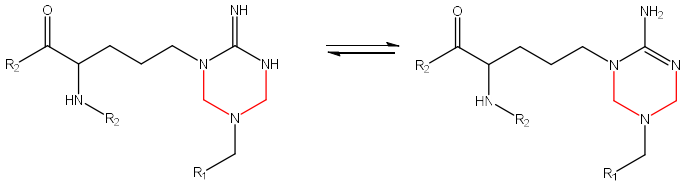
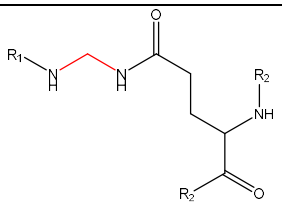
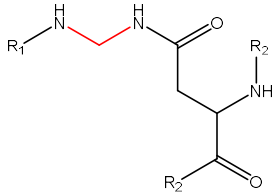
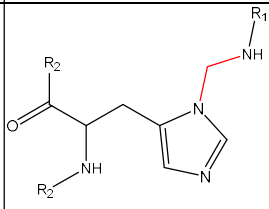
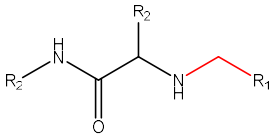
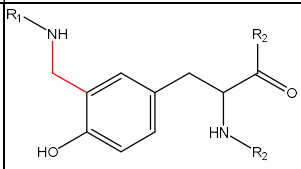
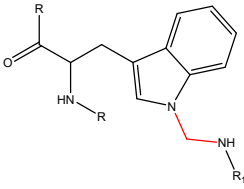
Arginine (R)			
Glutamine (Q)		Asparagine (N)	
Histidine (H)		N-term	
Tyrosine (Y)		Tryptophan (W)	

Figure V-2 Proposed chemistry of the formaldehyde cross-linking reaction: **(A)** entirely resumed and representing the formation of a methylol group (-CH2OH with +30 Da), followed by the production of a labile Schiff base (an imine) due to the loss of a molecule of water (-N=CH2 with +12 Da) and by final cross-link formation with another amino acid residue (R_1-CH_2-R_2); **(B)** with a focus on the first step. Formaldehyde might react with the N-term and with lysine and arginine side chains; (**C**) with a focus on the second step. The imine might cross-link with several amino acids such as arginine, glutamine, asparagine histidine, N-term, tyrosine and tryptophan. Figures elaborated from Sutherland et al., 2008.

More recently, a different chemical mechanism was proposed: the formation of a dimeric interaction of two imines, potentially resulting in a 1,3-diazetidine linkage with a mass shift of 24 Da [114] (Figure V-3).

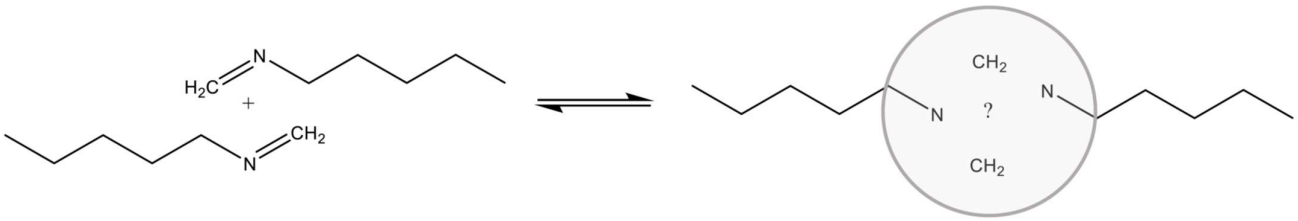


Figure V-3 Alternative proposed chemistry of the formaldehyde cross-linking reaction mechanism. Two side chains modified in imines dimerize among them, inducing an increase of mass of 24 Da. The chemical structure at the covalent bond is although not yet established. Figure extracted from Tayri-Wilk et al., 2020.

The formed cross-links are generally relatively stable [111]; however, the covalent bond is reversible (for example, with heat exposure) [115]. The reactivity of formaldehyde to particular residues and the formation of the subsequent cross-links is related to its concentration and the time of incubation: treatments with a low concentration of formaldehyde and a short period of incubation (often implemented for the *in vivo* interaction studies) show a reasonable specificity of the formaldehyde reaction (principally with lysine side chain). The specificity decreases with the increasing of formaldehyde concentration and time of incubation [116-122].

Over recent years, proteomic studies on formaldehyde-induced modifications conducted on FFPE (Formalin-Fixed Paraffin-Embedded) tissues revealed that the long exposition of formaldehyde on proteinaceous compounds was also leading to peculiar chemical modifications, such as lysine methylation (+14 Da) [123-125]. A minor increment of thiazolidine adduct (+12 Da), methylol (+ 30 Da) [126] and lysine formylation [127, 128] were also observed.

4.3 Structural study of formaldehyde cross-linking in a model protein

A structural investigation was performed on a lysozyme model system treated with formaldehyde to identify potential markers characteristics of formaldehyde-based cross-links and determine the fragmentation patterns of the resulting reticulated peptides. Lysozyme was chosen for its small size and the well-established crystallographic structure (structure of hen egg-white lysozyme, chain A - 1DPX).

4.3.1 Sample preparation and analysis

The model protein, lysozyme (200 μM), was incubated with formaldehyde (2%, w/v) in PBS (37 °C, pH 7.5) for one hour. The reaction was quenched through the addition of 1 M Tris buffer (pH 7.5) with a final concentration of 0.5 M.

The protein was then concentrated using Amicon 3 kDa MW-cut off filters and diluted in 0.1% acid formic in water. Treated lysozyme was tryptic digested following the eFasp protocol (with reduction and alkylation step) and analysed by nano LC-MS/MS procedure. An unbiased cross-link investigation was conducted via MaxQuant software (v1.6.14), looking for reticulated peptides having a mass variation of 12 or 24 Da. The potential covalent bonds were researched setting lysine (K), Arginine (R) and Protein N-terminus as potential first reactive sites, while arginine (R), tyrosine (Y), histidine (H), tryptophan (W), asparagine (N) and glutamine (G) as potential second reactive sites. The detailed protocol for the nano-LC/ MS-MS and the cross-link search is outlined in section 5.6.1 on page 198.

4.3.2 Analysis and discussion of the results

The results achieved suggested the formation of a formaldehyde-based cross-link with 24 Da shift between the arginine (R112) and lysine (K116) lysozyme residues. Both inter-and intra-peptidic formations presenting the covalent bond between these two amino acids were attributed to several values. The list of the results is outlined in Table IV-2.

As an example, Figure V-4 illustrates the MS/MS spectrum of the 5+ charged ion m/z 66.7227, whose fragmentation pattern shows the presence of an inter- peptidic reticulation between the arginine 112 of the residue α ^{98}IVSDGNGmnAWVAW**R**NR^{114} and the lysine 116 of the peptide β ^{115}c**K**GTDVQAWIR^{125}. One oxidation (M105) and one deamidation (N106) were localised in the first peptide; in the second one, instead, a carbamidomethylation was identified (C115).

The HCD fragmentation was proven to occur at both the cross-link bridge and the peptide backbone. Values attributable to each entire peptide plus 12 Da (depicted with an asterisk) were observed (α^{+2*}:987.9654; β*:1345.6664 and β^{+2*}:673.3328), suggesting symmetric fragmentation of the covalent bond.

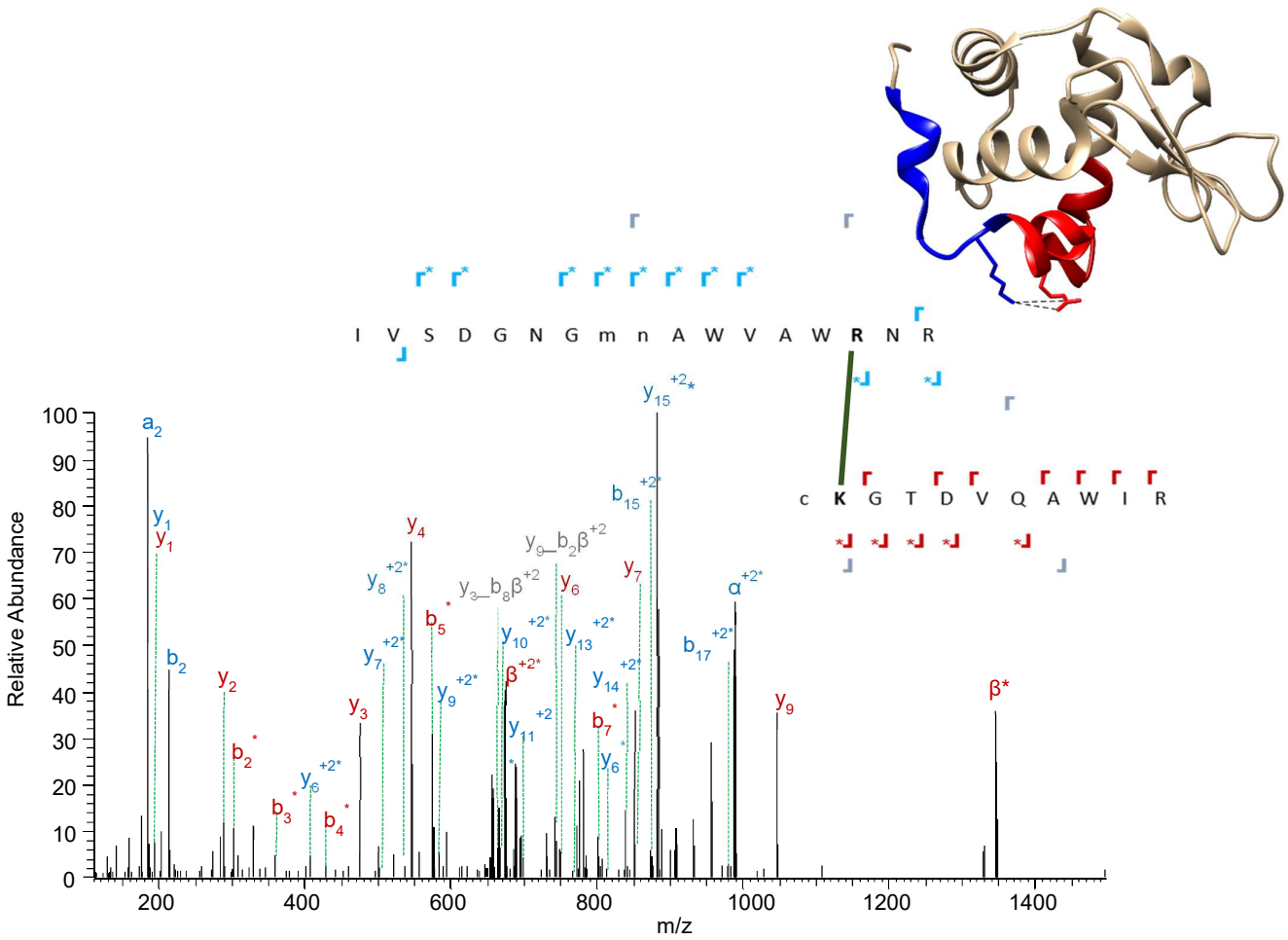


Figure V-4 MS/MS spectrum of the 5+ charged ion m/z 66.7227. The fragmentation pattern strongly suggests the presence of an inter-peptidic cross-link between R112 of the residue α ^{98}IVSDGNGmnAWVAWRNR^{114} with the K116 of the peptide β ^{115}cKGTDVQAWIR^{125}. In the lysozyme atomic structure, the α peptide is highlighted in blue while the β peptide in red; the two side-chains involved in the potential bond are also displayed. The Cα–Cα distance between the two residues results lower than 8.3 Å.

Ion fragments without and with an additional 12 Da mass (this latter depicted with an asterisk) were outlined for peptide α (y_1, b_2 and y_6^{+2}*- y_{11}^{+2}*, y_{14}^{+2}*, y_{16}^{+2}*, b_{15}^{+2}*, b_{17}^{+2}*, in blue in the Figure V-4) and peptide β (y_1- y_4, y_6,y_7, y_9 and b_2*- b_5*, b_7* in red in the Figure V-4), suggesting a further fragmentation on the backbone. Finally, two values were ascribed to fragments of both peptides having the 2C cross-link (reported in grey colour in Figure V-4): (y_3_b_8 β^{+2}) of m/z 664.8284 and (y_9_b_2 β^{+2}) of m/z 743.3622. These data confirmed that the HCD fragmentation also occurred solely at the backbone, leaving the cross-link intact. It is noteworthy that the suggested cross-link well suited the lysozyme atomic structure, presenting a Cα–Cα distance between the two residues lower than 8.3 Å.

Although minimal distance for the formation of a formaldehyde bond was theoretically defined at 2.3- 2.7 Å [104], various researches proved that the cross-links could establish in an average distance between 4.6 and 10.7 Å [111, 114, 116, 129-131].

Intra-peptidic structures containing the same covalent bonds were also likely generated during the digestion, as suggested from the MS/MS interpretation of different values.

For instance, the charged ions 4^+ m/z 858.1727 and 5+ m/z 686.7397 (both with a monoisotopic mass of 3428.6595) were attributed to the miss-cleaved peptide ^{97}KIVSDGNGmnAWVAW**R**NRc**K**GTDVQAWIR^{125} with oxidation on M105, a deamidation on N106, and a carbamidomethylation on C115. The fragmentation of the charged ions 4^+ m/z 858.1727 in Figure V-5 produced values corresponding to the mass of the fragment +24 Da (y_{21}^{+3*}, y_{22}^{+3*}, y_{24}^{+3*}, y_{26}^{+3*}, y_{27}^{+3*}, y_{28}^{+3*}), might suggest the formation of a mono cross-link between R114-K116 (+24 Da).

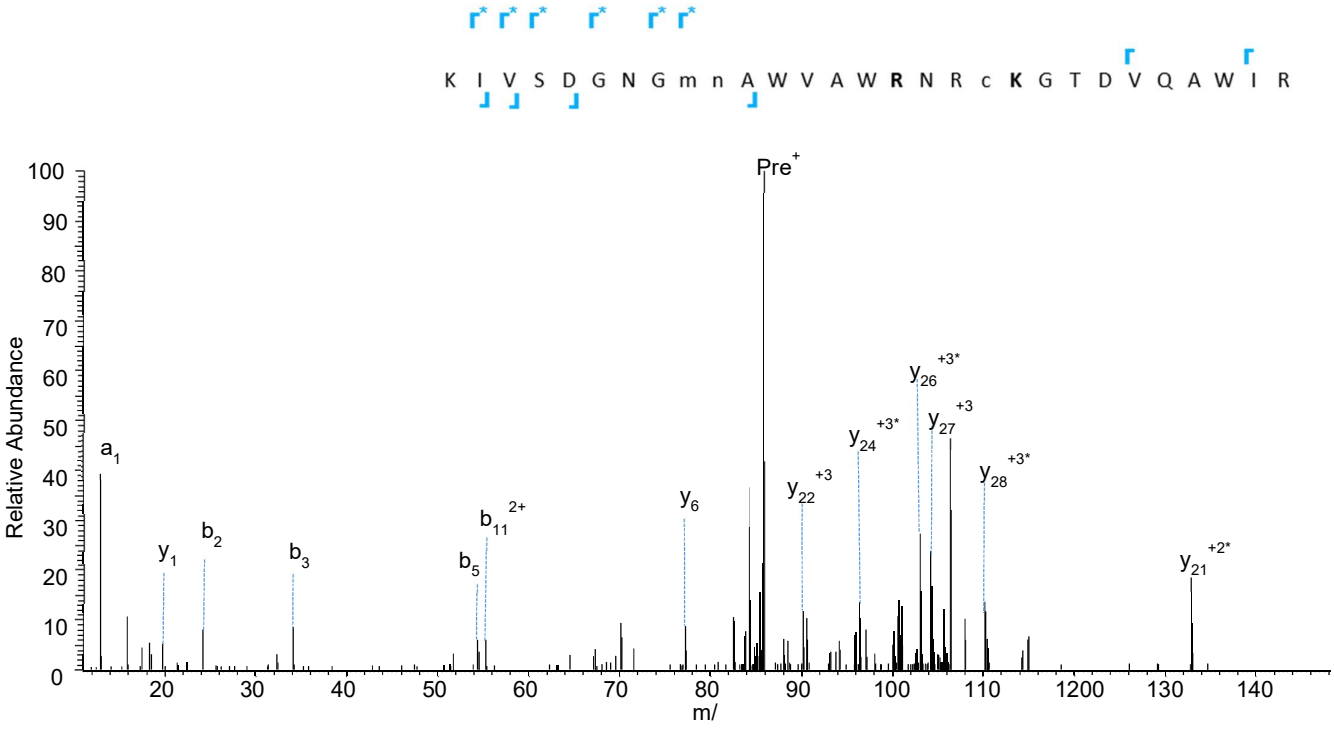


Figure V-5 MS/MS spectrum of the charged ion 4^+ m/z 858.1727 corresponding to the miss-cleaved peptide ^{97}KIVSDGNGmnAWVAWRNRcKGTDVQAWIR^{125} with oxidation on M105, a deamidation on N106, and a carbamidomethylation on C115. The detected values (y_{21}^{+3*}, y_{22}^{+3*}, y_{24}^{+3*}, y_{26}^{+3*}, y_{27}^{+3*}, y_{28}^{+3*}) might suggest the formation of a cross-link between R114-K116 (+24Da).

5. Proteomic analysis of a formaldehyde glue treatment of a Coptic manuscript

Two micro-samples of a glue conservation treatment collected from an ancient Coptic manuscript were investigated in partnership with Dr Julie Arslanoglu and Dr Federica Pozzi (*Department of Scientific Research, The Metropolitan Museum of Art, New York*) and conservators Maria Fredericks and Frank Trujillo (*Thaw Conservation Center, The Morgan Library & Museum, New York*). The codex presented unstable and fragile preservation conditions: its parchment pages were brittle and stiff for unclear reasons and challenging to store. The research's principal aim was to characterise past conservation treatment(s) that could contribute to the understanding of the evident degradation state of the manuscript.

Firstly, a comprehensive proteinaceous investigation was conducted via the classical bottom-up proteomic approach for protein identification and the analysis of their modifications. Following the achieved results, a structural study via cross-linking search was performed to provide a more in-depth understanding at a molecular level of the degradation causes.

5.1 Introduction of Coptic literature

Coptic literature emerged in Egypt in the earlier centuries of the modern era, after the decline of hieroglyphics (in the fourth century) and Demotic scripts (in the fifth century). Coptic is the fifth and final stage of the Egyptian language that Egyptian Christians mainly developed to ensure the reading of sacred scriptures: the idiom consists of the Greek alphabet and new letters adapted from Demotic characters [132]. The majority of the survived manuscripts were written on parchment, in Sahidic dialect and had biblical subjects; although, archaeological founding proves that papyrus and paper were also implemented in the production of literary texts in Coptic. Liturgical books on paper supports were mostly written in Bohairic, the second Coptic dialect, and they were generally of a later date. The parchment decadence period roughly corresponds with the substitution of Coptic as a spoken language with Arabic due to the Muslim conquest of Egypt. Noteworthy, Coptic codices, which were bound in wood or leather covers, are the earliest book-binding texts [133].

5.2 Hamuli collection of Coptic codes

The manuscript under analysis belongs to the so-called Hamuli collection, which is the largest group of intact Coptic codices from a single source. In 1910, the collection consisting of sixty parchment codices (about 3000 folios) was found in a stone cistern nearby the ruin of St Michael's Monastery, located in al- Faiym oasis, along the River Nile and close to the village of el-Hamuli, in the Fayum region, Egypt [134, 135].

Historic researchers assumed that the monastery of Copts was destroyed or disbanded around 1000 due to religious persecution, and the monks themselves buried their manuscripts before leaving the place [136]. The codes were all written on parchment in Sahidic dialect (except one written in the local dialect Fayumic), encompassing both Biblical and magical/documentary arguments. Furthermore, several codes were also presenting the original bindings [137]. Noteworthy, in twenty Hamuli manuscripts, were identified the earliest colophons found in a Coptic manuscript: they consist of short texts that outline the copyists' names, the date and the place of copying. The timespan of these colophons (823- 914 AD) allowed the experts to date the manufacturing to between the 9^{th} and 10^{th} centuries, corroborating the monastery closure's historical assumptions. After their finding, several volumes and leaves of the collection were spread out all over the world. Nonetheless, the collector Pierpont Morgan successfully brought together almost all volumes at the Morgan Library and Museum. The museum is currently the most important depository of Coptic manuscripts outside Egypt [138].

In light of their great fragility and extremely precarious conditions, the codices were sent to the Vatican Restoration Studio in 1912 for conservation treatment under F. Ehrle and A. Castellani's supervision. Both the bookbindings and the manuscripts were described as highly degraded, mostly due to moisture, microbiological attack and pigment corrosion. The principal operations conducted consisted of removing original bookbindings, separation of attached parchment sheets, consolidation of detached fragments and pages whose fragility was compromising the manuscript integrity, and filling of voids with either glue fills or patches and glue. During the restoration, twelve copies of each volume's photographic facsimile were also taken as documentation, whose original scans are still available online at https://archive.org/details/PhantoouLibrary/page/n19/mode/2up.

Finally, it was decided to preserve the original book bindings separately from the manuscripts and cover the pages in a modern dark reddish-brown goatskin. The restoration lasted until 1929, several years more than what initially scheduled, principally due to the First World War outbreak [135].

The manuscripts were then shipped to the Morgan Library in New York, where they are currently displayed. In 1987, Leo Depuydt, professor at Yale University and the Hamuli collection expert, published a complete and detailed description of each codex with an overview of the principal restoration treatments [134].

5.3 Methodology

5.3.1 Samples

Two samples labelled M581F3_S1 and M581F3_S1_A_S1 (shortened to S1 and A_S1) were collected for proteomic analysis from glue repairs on the third page of the Coptic manuscript *"Martyrdom of St. Pteleme"* belonging to the Hamuli Collection. The almost complete codex was bought in 1911 in Paris, while the four missing leaves of the manuscript were purchased one year later in Cairo. The manuscript is a parchment codex of twenty-six pages of literary character and biblical subjects written in Sahidic dialect. The experts conducted the sampling, collecting what appears to be a repair material of the manuscript (glassy fill).

5.3.2 Sample preparation

The two selected samples were digested following the filter-aided protocol (e-Fasp) optimized for trace analysis of ancient and degraded samples (described in detail in Chapter II4.1 on page 35). Sample S1 was digested by incubating the protein solution overnight at 37°C with Trypsin/Lys-C Mix in 50 mM of ammonium bicarbonate pH8.8 (1:50 enzyme to mass protein ratio). In Sample A_S1, proteins were sequentially digested with Glu-C in 50mM phosphate buffer pH 7.8 (overnight) and trypsin in 50 mM ammonium bicarbonate pH8.8. (overnight).

5.3.3 nLC-MS/MS analysis

The peptide mixture was analyzed on an Ultimate 3000 nanoLC system (Dionex, Amsterdam, The Netherlands) coupled to an Electrospray Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA).

The LC and MS/MS analysis settings were kept equal to the ones for the other bottom-up studies outlined in Chapter II4.2 on page 38.

5.3.4 Data Analysis

The research was subdivided into three parts, performing a further data analysis based on the information sought. Firstly, the investigation focused on protein identification. Then, the detection and localisation of the predominant protein modifications were pursued. Following the achieved results, a formaldehyde-based cross-linking study was conducted to examine the potential presence of chemically reticulated peptides.

5.4 Protein/Species Identification

5.4.1 Data analysis

Raw files generated by the nanoLC-MS/MS analysis were searched using PEAKS Studio. At first, data elaboration was performed against a broad database SwissProt (Uniprot) to get an overall look, followed by a more restrict database composed of a complete list of proteins available in NCBI RefSeq of *Ovis Aries* (n° accession 56411), *Capra hircus* (n° accession 51264), *Oryctolagus cuniculus* (n° accession 49038), *Bos Taurus* (n° accession 132278) October 2019 version. A homemade database containing common contaminants from human proteins (like keratins and serpins) and laboratory preparation (typically porcine trypsin) was added to the research. Trypsin was selected as enzyme for the S1 sample, while Glu-C and trypsin were used for the A_S1 sample, with three-missed cleavages allowed for both searches. Error tolerance was set at 10.0 ppm for parental mass and 0.02 Da for fragment mass. Carbamidomethylation was selected as a fixed modification, while oxidation (methionine, proline), acetylation (N-terminal protein), deamidation (asparagine, glutamine), glutamine and glutamic acid to pyroglutamate were set as variable modifications, setting a maximum of 5 modifications per peptide.

Protein false discovery rate (FDR) was set to 0.1%, with -10logP $\geq$ 18 as a score threshold for peptides. A minimum of two non-overlapping peptides was defined as a requirement for protein validation.

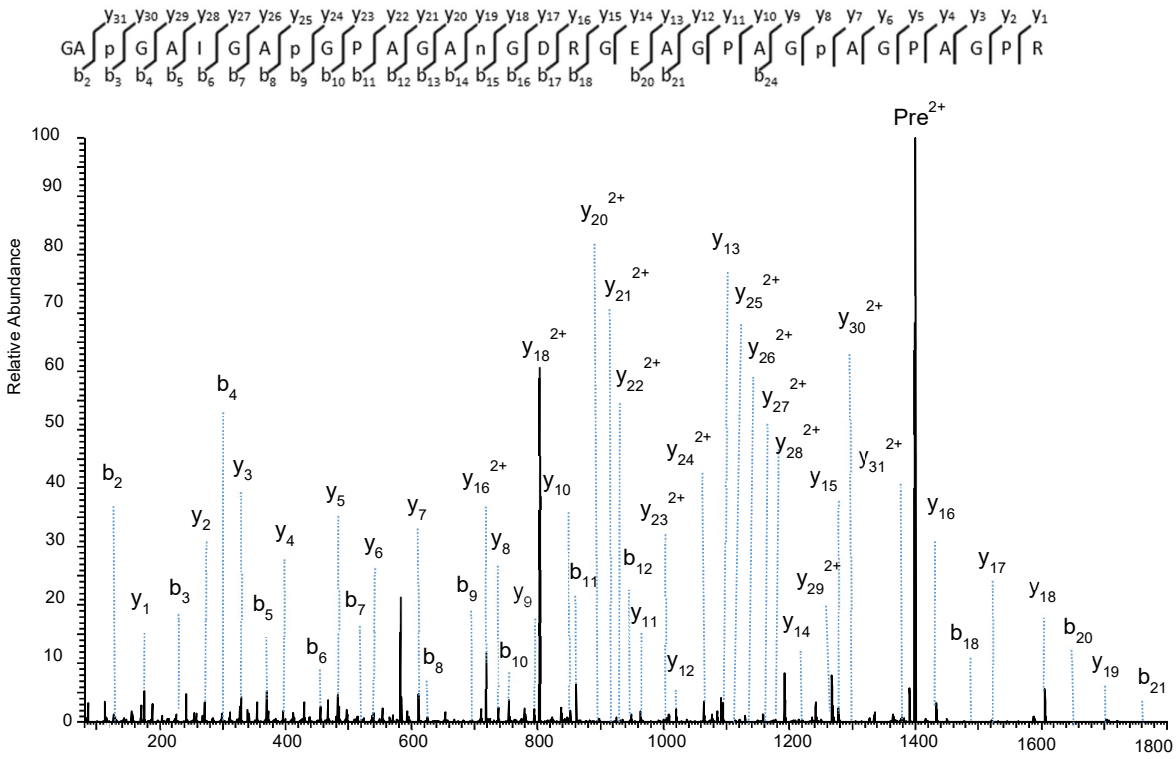
5.4.2 Results

A significant number of different collagen proteins have been detected in both samples with the achievement of remarkable protein sequence coverage (Table V-3). The major collagen proteins detected were collagen α-1(I), collagen α-2(I) and collagen α-1(III), commonly found in animal connective tissue. A manual inspection of the MS/MS spectra ensured the signal-to-noise ratio's quality and the proper attribution of fragment ions. Identifying several species-diagnostic peptides led to the discrimination of three different species sources, specifically from *Bovinae* and *Caprinae* subfamilies and *Oryctolagus Cuniculus* (European rabbit). The list of the unique peptides is reported in Table V-4.

Bovinae subfamily-Bos Taurus

Several peptides were attributed to collagen α-2(I), collagen α-1(I), and collagen α-1(III) from the *Bovinae* subfamily. In sample S1, for example, the three proteins have been detected with a coverage of 75% (393 peptides), 73% (437 peptides) and 61% (211 peptides), respectively. The majority of unique peptides detected were shared among the different species of the *Bovinae* subfamily. Nonetheless, some peptides of collagen α-2(I) and collagen α-1(III) lead to narrowing the protein's attribution to the species of *Bos taurus.*

Figure V-6 represents, as an example, the MS/MS spectrum of the doubly charged ion m/z 1399.1604 attributed to the peptide of collagen 2(I) ^{674}GApGAIGAPGPAGANGDRGEAGPAGPAGPAGPR^{706}, which is unique to the *Bos taurus* species. The sequence was identified with three hydroxyprolines (P3-P9-P26) and a deamidation localised on N15.



Sequences producing significant alignments:

Description	Max Score	Total Score	Query cover	E Value	Per. Ident	Accession
hypothetical protein [Bos mutus]	57.4	57.4	100%	3e-08	100.00	MXQ83349.1
Collagen alpha-2(I) chain [Bos mutus]	55.1	55.1	100%	2e-07	100.00	ELR46121.1
PREDICTED: collagen alpha-2(I) chain [Bos mutus]	55.1	55.1	100%	2e-07	100.00	XP_005909757.1
PREDICTED: collagen alpha-2(I) chain [Bos indicus]	55.1	55.1	100%	2e-07	100.00	XP_019815249.1
collagen alpha-2(I) chain precursor [Bos taurus]	55.1	55.1	100%	2e-07	100.00	NP_776945.1
PREDICTED: collagen alpha-2(I) chain-like isoform X3 [Lipotes...	55.1	55.1	100%	2e-07	93.94	XP_007450706.1
PREDICTED: collagen alpha-2(I) chain-like isoform X2 [Lipotes...	55.1	55.1	100%	2e-07	93.94	XP_007450705.1

Figure V-6 MS/MS of the doubly charged precursor ion at m/z 1399.1604 attributed to the peptide (674 – 706) of collagen 2(I) from sample S1. The peptide sequence was identified through the full y-fragmentation pattern and b_2- b_{18}, b_{20}, b_{21}, b_{24} fragment ions. Three hydroxyprolines (P3-P9-P26) and a deamidation (N15) were localised. The alignment of the experimental peptide sequence with all sequenced species present in the NCBI database provided the sequence's attribution uniquely to *Bos taurus* species.

Caprinae subfamily - Capra hircus

In both investigated samples, proteins attributable to the *Caprinae subfamily* were detected (Table V-3). Contrarily to *B.taurus* species, where the species-specific collagen detected belonged only to type I-II and III, traces of collagen V-VI and XI attributable to the *Caprinae subfamily* were also noted.

Due to the high similarity among collagens of *Capra hircus* and *Ovis aries* species, the majority of discriminating peptides identified refer to both species without the possibility of distinctions (Table V-2).

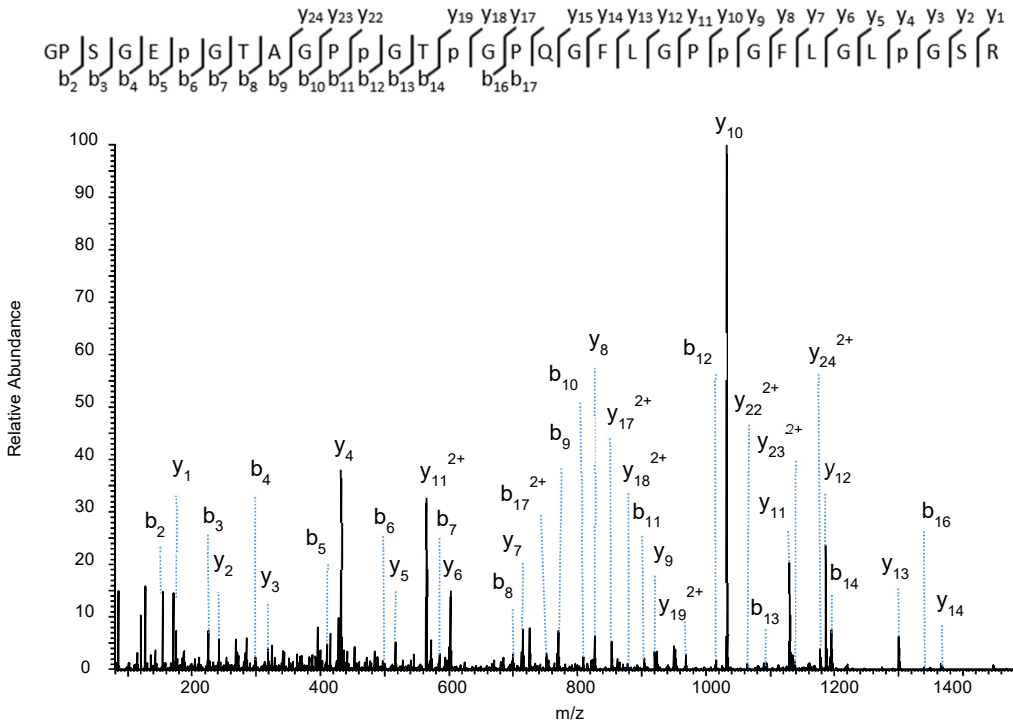
For example, the Collagen α-1(III) sequence from *Capra hircus* has 99.7% of identity with the same protein from *Ovis aries*, presenting 1463 amino acids in common in a total of 1467 peptides. Furthermore, in this case, studying the clear attribution to one species is particularly challenging since all four tryptic peptides discriminating are shared between one of the two species and *Bos taurus*, which was confidently detected in the samples.

Table V-2 Alignment between the peptide sequences of three Collagen α-1(III) from *C. hircus*, *O. aries* and *B. taurus*.

Tryptic peptide sequence	Species
120 NGDPGPPGSPGSPGSPGPPGICESCPTGSQNYSPQYEAYDVK 161 120 NGDPGPPGSPGSPGSPGPPGICESCPTGGQNYSPQYEAYDVK 161 120 NGDPGPPGSPGSPGSPGPPGICESCPTGGQNYSPQYEAYDVK 161	*C.hircus* *O.aries* *B.taurus*
451 GEAGSPGIPGPK 462 451 GEAGSPGIAGPK 462 450 GEAGSPGIAGPK 461	*C.hircus* *O.aries* *B.taurus*
538 GFPGNPGAPGSPGPAGHQGAVGSPGPAGPR 573 538 GFPGNPGPPGSPGPAGHQGAVGSPGPAGPR 573 537 GFPGNPGAPGSPGPAGHQGAVGSPGPAGPR 572	*C.hircus* *O.aries* *B.taurus*
1174 GSEGSPGHPGQPGPPGPPGAPGPCCGAGGVAAISGIGGEK 1213 1174 GSEGSPGHPGQPGPPGPPGAPGPCCGAGGAAAIAGIGGEK 1213 1173 GSEGSPGHPGQPGPPGPPGAPGPCCGAGGVAAIAGVGAEK 1212	*C.hircus* *O.aries* *B.taurus*

Collagen α-2(I) sequence from *Capra hircus* has 99.9% of identity with the same protein from *Ovis aries*, presenting only two amino acids of difference (1362 amino acids in common for a total of 1364), which both belonging to the same tryptic peptide ^{845}GPSGEPGTAGPPGTPGPQGFLGPPGFLGLPGSR^{877}. In detail, the phenylalanine in position 864 and proline in position 867, which are characteristic in *Capra hircus*, are substituted with leucine and alanine in *Ovis aries*. This tryptic peptide is effectively unique for the *Capra hircus* species since it is not shared with any sequenced breed.

In the Hamuli manuscript, the detection of this discriminating tryptic peptide in sample S1 (illustrated in Figure V-7) suggested a potential attribution of the identified collagens to *Capra hircus* species.



Sequences producing significant alignments:

Description	Max Score	Total Score	Query cover	E Value	Per. Ident	Accession
PREDICTED: collagen alpha-2(I) chain [Capra hircus]	59.3	59.3	100%	5e-09	100.00	XP_005678993.1
collagen alpha-2(I) chain [Microtus ochrogaster]	57.4	57.4	100%	2e-08	96.97	XP_005363739.1
collagen alpha-2(I) chain [Odocoileus virginianus texanus]	57.4	57.4	100%	3e-08	96.97	XP_020769440.1
PREDICTED: collagen alpha-2(I) chain [Peromyscus maniculatus...	56.2	56.2	100%	8e-08	93.94	XP_006991260.1
LOW QUALITY PROTEIN: collagen alpha-2(I) chain [Peromyscus...	56.2	56.2	100%	8e-08	93.94	XP_028725629.1
collagen alpha-2(I) chain [Onychomys torridus]	55.8	55.8	100%	9e-08	93.94	XP_036038246.1

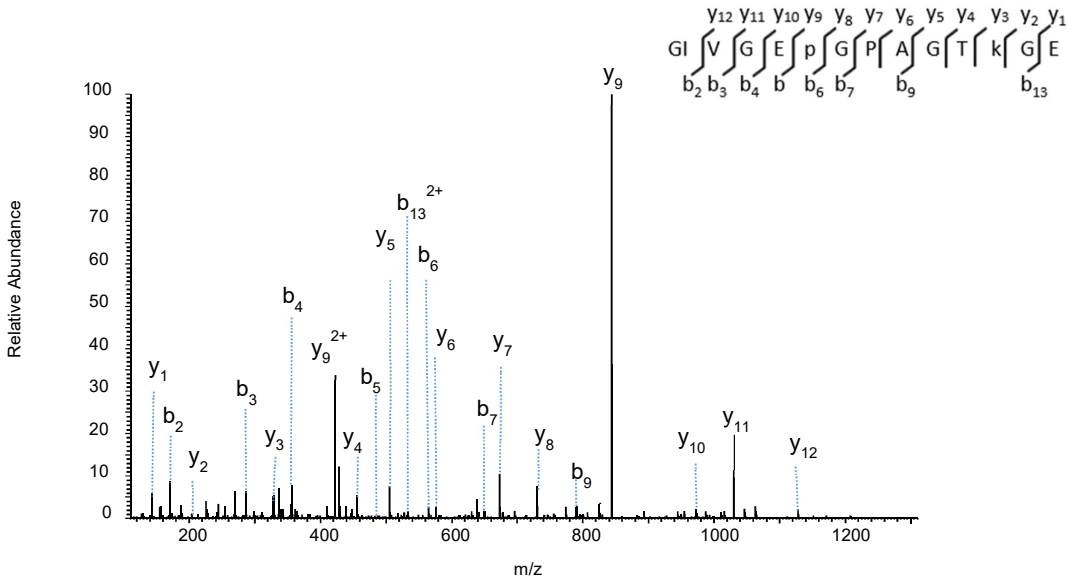
Figure V-7 MS/MS spectrum of the triply charged ion at m/z 1031.8372 (Δm 1.9 ppm) detected in the sample S1. The detection of almost full y- fragment ions and b_2^+ to b_{14}^+, b_{16}^+ b_{17}^+ fragment ions leads to the unambiguous sequence assignments leading to the identification of the peptide 845-877 of collagen 2(I) specific uniquely for the *Capra hircus* species. The fragments from y_{11}^+onwards are of significant importance in the peptide discrimination from its counterparty of *Bos taurus* and *Ovis aries*. The peptide presents five hydroxyprolination in the positions P850-P856-P859-P868-P874.

The assumption of goat-based material might be corroborated by identifying the sample A_S1 of the peptide presenting an unspecific cleavage after leucine ^{866}GPPGFLGLPGSR^{877} (doubly charged ion m/z 593.8153 Δm 1.6 ppm), which is not unique of *Capra hircus,* but it is not attributable to *Ovis aries*.

Oryctolagus cuniculus

Protein sequence alignment supported by the BLAST research suggested collagen proteins from a third biological species, most likely *Oryctolagus cuniculus* (European rabbit). Despite the lack of unique peptides detected, four peptides potentially characteristic for the species were determined in collagen 2(I) (Table V-4). The identified peptidic sequences might be assigned to *O. cuniculus* or a few other species, which were hardly reliable considering the manuscript manufacturing, context, and history.

As an example, Figure V-8 depicts the MS/MS spectrum of the doubly charged ion at m/z 649.8341 that was attributed to the peptide ^{341}GLVGEpGPAGTkGE^{354} of collagen α2(I). The alignment of the experimental peptide sequence with all sequenced species present in the NCBI database provided the attribution to either *O. cuniculus* (European rabbit), *Suricata suricatta* (meerkat), or *Carlito syrichta* (Philippine tarsier). Between the three possible species, the *O. cuniculus* represented the more probable hypothesis.



Sequences producing significant alignments:

Description	Max Score	Total Score	Query cover	E Value	Per. Ident	Accession
collagen alpha-2(I) chain [Suricata suricatta]	43.9	275	100%	0.001	100.00	XP_029785028.1
collagen alpha-2(I) chain isoform X1 [Carlito syrichta]	43.9	207	100%	0.001	100.00	XP_008067045.1
collagen alpha-2(I) chain isoform X2 [Carlito syrichta]	43.9	207	100%	0.001	100.00	XP_008067047.1
collagen alpha-2(I) chain precursor [Oryctolagus cuniculus]	43.9	228	100%	0.001	100.00	NP_001182597.1
collagen alpha-2(I) chain [Fukomys damarensis]	41.4	154	100%	0.008	92.86	XP_010601746.1
Collagen alpha-2(I) chain [Fukomys damarensis]	41.4	153	100%	0.008	92.86	KFO23903.1

Figure V-8 MS/MS spectrum of the doubly charged ion at m/z 649.8341 (Δm 1.8 ppm) detected in the sample A_S1. The full y- fragmentation pattern and b_2 to b_7, b_9 and b_{13} provided the characterisation of the peptide 341-354 of collagen 2(I) with the oxidation of proline in position 347 and methylation of lysine at position 352. The alignment of the experimental peptide sequence with all sequenced species present in the NCBI database provided the attribution of the sequence to either *Oryctolagus cuniculus* (European rabbit), *Suricata suricatta* (meerkat), or *Carlito syrichta* (Philippine tarsier).

In the peptide, proline oxidation was localised in position P(347) and lysine methylation at position K (352). In both samples (S1 and A_S1), a relatively great extent of methylations was unexpectedly detected in several peptides of collagens. The modification is discussed more in-depth in paragraph 5.5.2 on page 195.

Table V-3 List of the identified proteins in the two samples S1 and A_S1 from the Coptic manuscript "Martyrdom of St. Pteleme" of the Hamuli collection

Description	Accession	Species	Samples	-10lgP	Coverage (%)	Peptides	Unique	PSM Sample	Avg. Mass
collagen 2(I)	NP_776945.1	*Bos taurus*	S1	302.44	75	393	74	3736	129064
			A_S1	366.23	70	395	72	1223	
collagen 1(I)	NP_001029211.1		S1	326.05	73	437	27	5121	138939
			A_S1	396.39	65	417	14	1796	
collagen 1(III)	NP_001070299.1		S1	288.53	61	211	37	2245	138439
			A_S1	328.1	56	219	51	721	
collagen 1(II) isoform 2	NP_001106695.1		S1	167.62	35	64	2	403	134427
			A_S1	194.93	34	65	2	169	
collagen 1(II) isoform 1	NP_001001135.2		S1	167.62	33	64	2	403	141829
			A_S1	194.93	32	65	2	169	
collagen 2(I) (predicted)	XP_005678993.1	*Capra hircus*	S1	276.79	73	285	10	2228	129033
			A_S1	335.9	65	305	3	916	
collagen 1(I) isoform X1 (predicted)	XP_017920382.1		S1	319.99	73	404	0	4644	139029
			A_S1	392.98	64	403	2	1725	
collagen 1(III) (predicted)	XP_005675926.1		S1	243.13	55	167	3	1603	138550
			A_S1	293.08	48	163	1	526	
collagen 1(II) (predicted)	XP_017903357.1		S1	164.92	33	63	1	402	141783
			A_S1	192.02	30	63	0	165	
collagen 2(V) (predicted)	XP_005675925.2		S1	127.75	17	19	0	51	144991
			A_S1	134.42	16	22	0	28	
collagen 2(VI) isoform X1 (predicted)	XP_017907820.1		S1	62.92	9	8	5	19	109392
collagen 1(XI) isoform X1 (predicted)	XP_005678082.2		S1	67.25	5	8	1	25	181099
			A_S1	68.12	2	3	0	3	
collagen 1(I) (predicted)	XP_017204320.1	*Oryctolagus cuniculus*	S1	303.13	70	337	35	4069	142873
			A_S1	379.33	57	342	37	1449	
collagen 2(I)	NP_001182597.1		S1	225.56	59	177	19	1336	128748
			A_S1	243.6	45	148	30	454	
collagen 1(III) (predicted)	XP_002712379.1		S1	181.94	42	95	0	661	134925
			A_S1	247.92	27	70	0	192	
collagen 1(II)	NP_001182600.1		S1	146.62	30	59	2	345	141683
			A_S1	188.67	27	56	0	149	

Table V-4 Table of the specie characteristic peptides identified in the two samples S1 and A_S1 from the Coptic manuscript "Martyrdom of St. Pteleme" of the Hamuli collection.

Proteins	Accession number	peptide	position	samples	m/z	z	ppm	-10lgP	RT (min)	Specie
collagen 2(I)	NP_776945.1	GApGAIGApGPAGAnGDRGEAGPAGPAGPAGPR	674-706	S1	1391.1688	2	0.2	84.54	77.68	*B.taurus, B. mutus, B. indicus, Toxodon*
				A_S1	927.7828	3	1.4	62.69	78.06	
		GYpGnAGPVGAAGApGPqGPVGPVGK	947-972	S1	1132.053	2	-0.4	72.12	106.37	*B. taurus, B. mutus, B.bubalis*
		GYpGnAGPVGAAGApGPQGPVGPVGK		A_S1	1131.563	2	1.8	66.83	100.7	
		GPAGPSGPAGKDGRIGqPGAVGPAGIR	1052-1078	S1	600.3196	4	-0.7	65.36	76.44	*B.taurus, B.mutus, B. bubalis, B.indicus*
		GPAGPSGPAGKDGRIGqpGAVGPAGIR		A_S1	604.3184	4	-0.6	53.68	70.32	
		GEpGnIGFpGPkGPSGDpGkAGEK	485-508	S1	776.3351	3	1.9	22.84	58.24	*B.taurus, B.mutus, B.indicus*
collagen 1(III)	NP_001070299.1	GPTGPIGPpGPAGqpGDKGESGAPGVpGIAGPR	762-794	S1	984.4883	3	0.4	82.72	101.21	*B.taurus, B.bubalis, B.indicus x B.taurus, B.b.bison, B.indicus, B.mutus*
		GPTGPIGPpGPAGqpGDKGESGApGVpGIAGPR		A_S1	989.8212	3	1.7	92.42	97.01	
		GEGGPpGAAGPAGGSGPAGPpGPqGVKGER	834-863	S1	847.4089	3	0.0	94.97	56.74	*B.taurus, B.bubalis, B.indicus x B.taurus, Bison bison bison, Bos indicus, Bos mutus*
		GEGGPpGAAGPAGGSGPAGPpGPQGVK	834-860	A_S1	1107.033	2	0.5	84.28	66.24	
		GAMGIKGhR	1101-1109	S1	471.7502	2	-0.7	28.23	33.76	*Bos taurus, Bos indicus x Bos taurus, B.b.bison, B. indicus, B.mutus*
		GPAGAnGLpGEKGPpGDR	492-509	S1	840.4028	2	0.1	55.66	45.89	*B.taurus, O.v.texanus, B.b.bison, B.indicus, B.mutus*
				A_S1	560.6051	3	1.5	38.5	46.14	
		GVpGFRGPAGAnGLpGEK	486-503	S1	571.9568	3	0.2	33.75	74.72	*B.taurus, O.v.texanus, B.indicus x B.taurus, B.b.bison, B. indicus, B. mutus*
		GEnGVpGEnGApGPMGPR	285-302	S1	863.8701	2	-0.4	49.67	71.71	*B.taurus, B.indicus x B. taurus, B.b.bison, B.indicus, B.mutus*
		GFpGPpGmK	243-251	A_S1	468.2186	2	0.9	22.67	38.61	*B.taurus, B.bubalis, B.b.bison, B.indicus, B.mutus,*
		GPpGPpGTnGVpGqR	414-428	A_S1	719.335	2	1.7	41.03	47.92	*B.taurus, M.davidii, B.indicus x B. taurus, B.b.bison, B. indicus, B.mutus*
collagen 2(I)	XP_005678993.1	GPSGEpGTAGPpGTpGPQGFLGPpGFLGLpGSR	845-877	S1	1031.837	3	1.9	69.48	177.17	*C.hircus*
		GPpGFLGLpGSR	866-877	A_S1	593.8153	2	1.6	25,31	119,65	*C.hircus, H.equinus, O. v. texanus*
		GEpGPVGAVGPAGAVGPR	977-994	S1	780.9112	2	1.5	54.19	88.83	*C.hircus, O.aries*
		TGPpGpAGISGPpGPpGPAGK	793-813	S1	610.6381	3	-1.7	50.15	71.6	*C.hircus, O.aries,*
		TGEpGAAGPpGFVGE	829-843	A_S1	687.8138	2	2.4	38.13	94.93	*C.hircus, O. aries*
		GEpGPVGAVGPAGAVGPR	977-994	A_S1	780.9115	2	1.9	40.86	86.89	*C.hircus, O.aries*
collagen 1(III)	XP_005675926.1	GPTGPIGPpGPAGqpGDKGEGGApGVpGIAGPR	763-795	S1	979.8184	3	2.4	78.3	95.65	*C.hircus, O. aries*
		DAGQpGERGApGPqGPPGAPGPLGIAGLTGAR	926-957	S1	970.4900	3	1.9	58.24	139.23	*C. hircus, O. aries*
		DAGQpGERGAPGpqGPpGApGPLGIAGLTGAR		A_S1	981.1489	3	-2.5	30.98	117.98	*C. hircus, O.aries*
collagen2(I)	*NP_001182597.1*	GLVGEpGPAGTkGE	341-354	A_S1	649.8341	2	1.8	34.15	50.51	*O.cuniculus, S.suricatta, C.syrichta*
		AGVMGPpGSRGSTGPAGVR	412-430	S1	576.2916	3	-1.2	22.17	66.57	*O. cuniculus, O.princeps*
		GHnGLQGLpGLAGqHGDqGApGAVGpAGPR	1022-1051	S1	932.1084	3	-1.2	83.87	98.32	*O. cuniculus, O.princeps,C.hoffmanni*
		GLpGLAGqHGDQGApGAVGpAGPR	1028-1051	A_S1	729.3581	3	0.7	39.65	83.52	*O.cuniculus, O.princeps, C. hoffmanni*

5.5 Investigation of proteins modifications

5.5.1 Data analysis

The investigation of the protein modifications was conducted unbiased using the PEAKS PTM tool [139]. The general parameters set for the analysis were analogous to the PEAK DB search implemented for protein identification. Once the analysis was concluded, the data were further filtered: protein false discovery rate (FDR) was set to 0.1% with de novo average-local-confidence (ALC) score of ≥ 50% and –log ≥ 20 score threshold for proteins. A semi-quantitative approach provided an assessment of the most abundant modifications present in the two samples. The computing was performed on the ensemble of unique sequences of identified proteins with a prior exclusion of contaminations and false positives. The incidence of a specific PTM was calculated by dividing the peptide-to spectrum match (PSM) by the corresponding amino acid residue's total frequency, except in the case of overlapping sequences. The resulting value estimates the modification extent and depicts an amino acid residue's probability of having particular PTMs.

The deamidation rates have been calculated for asparagine (N) and glutamine (Q) for both samples using deamiDATE 1.0 [140]. The program measures the relative level of deamidation: the value 1 corresponds to no deamidation, and 0 is the complete deamidation.

5.5.2 Results from the analysis

The incidence of specific modifications resulted in strong similarities between the two samples, without remarkable differences, except for oxidation (Table V-5).

Table V-5 Most abundant modifications detected in the two samples. The incidence of a specific modification was calculated, dividing the peptide-to spectrum match (PSM) by the total frequency of the corresponding amino acid residue (%).

Modifications	PSM modified / PSM tot (%)	
	S1	**A_S1**
Hydroxyprolination (P)	81	87
Deamidation (Q)	63	59
Deamidation (N)	84	90
Methylation (K)	8	7
Oxidation (M)	35	80

Besides the high number of hydroxyproline modifications (81% and 87%), which is a standard indicator of collagen proteins, methionine oxidation was also observed, especially on A_S1, where the detected rate (80%) was more significant than S1 (35%). Considering the samples were taken from the same page of the manuscript, methionine oxidation may result from the digestion protocol used (trypsin/LysC for S1 versus trypsin/GluC for A_S1) [141].

The calculation of the PSM modified / PSM tot (%) showed a relatively high deamidation rate in asparagine (84% and 90%) and glutamine (63% and 59%) for the sample S1 and A_S1, respectively.

The bulk deamidation was then calculated for the two amino acids residues (using deamiDATE 1.0), focusing on collagen type 1 (COL1A1 and COL1A2). In accordance with the observations achieved from the PSM calculation, an average of 10% of asparagine and 50% glutamine residues were observed undamaged in the collagens (Figure V-9), confirming significant levels of deamidation.

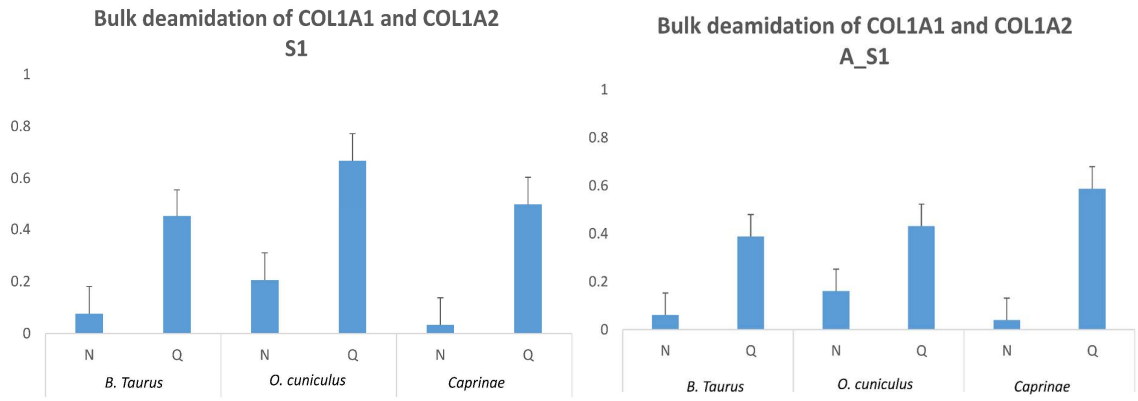


Figure V-9 The calculated rate of non deamidated asparagine and glutamine per sample in collagen type 1 (COL1A1 and COL1A2). The vertical axis shows the relative "non-deamidated" portion: the value 1 correspond to no deamidation, and 0 is the complete deamidation. The error bars represent the standard deviation.

An unexpected rate of lysine methylations emerged from the unbiased investigation of protein modifications via the PEAKS PTM algorithm showing 7% and 8% for each of the two samples (Table V-5). This specific modification at this concentration is hardly ever observed in artistic samples (e.g., to give an idea, the methylation rate detected in three investigated mediaeval manuscripts presented an average of 0.8%).

The unusual amount of lysine methylations in the investigated proteins presented a remarkable analogy with the proteinaceous scenario in the Formalin-Fixed Paraffin-Embedded FFPE tissue in which this modification is reported as the most recurrent [123]. In these clinical tissue specimens, the modification was explained with the interaction between the amino acid residues and the formaldehyde added as a fixative. Gathering this information with historical studies on the Hamuli collections and the hard and brittle texture of the glue in the sampling area, the supposition that the methylations observed were mainly induced by formaldehyde arose. Indeed, the description of the restored codex [134] reported a possible application of gelatine-formol made by Franz Herle during the extensive restoration treatment conducted on the Coptic codes at the Vatican library.

5.6 Cross-links investigation and localisation

5.6.1 Data analysis

Raw data were processed using MaxQuant software (v1.6.14) to identify and localize potential formaldehyde-based cross-links. The investigation focused on three types of collagens retrieved from the NCBI database: Collagen α-1(I), Collagen α-2(I) and Collagen α-1(III) were selected for each one of the three species under analysis, *Capra hircus*, *Bos Taurus* and *Oryctolagus Cuniculus*. Initially, each collagen type was separately investigated to reduce analysis variability; however, the three species were also gathered to observe cross-links among different species. In a second moment, the three collagen proteins of *Bos Taurus* have been investigated together to detect inter-protein interactions in the same species.

The following parameters were set for the search: Trypsin (S1) or Glu-C/Trypsin (A_S1) as enzymes, based on the sample digestion conditions; missed cleavage allowed to a max of 2. MS accuracy of 10 ppm; MS/MS accuracy of 10 ppm. Carbamidomethylation of cysteine was selected as a fixed modification, while oxidations of proline and methionine and deamidation of asparagine and glutamine were set as variable modifications. The maximum of modifications per peptide was five.

Formaldehyde was defined as a cross-linker, with lysine (K), arginine (R) and protein N-terminus as a potential first reactive site while arginine (R), tyrosine (Y), histidine (H), tryptophan (W), asparagine (N) and glutamine (G) as potential second reactive sites in the formation of methylene bridge. Several runs were performed, setting either 12 Da (1C) or 24 Da (2C) as the cross-link's mass variation. The minimum length for a paired peptide sequence was four. One mono-peptide and one di-peptide unsaturated link were selected, while mono-peptide saturated links were set to zero. Two minimum matches were established.

The obtained results were manually filtered, choosing cross-links detected in both samples or detected in one sample when fragmented multiple times, especially with different modifications and miscleavage. Furthermore, the cross-links were reported if: The mass of the precursor ion has a ppm mass lower than 5 ppm, and no other linear peptide was found to match the data (also denovo research through PEAKS software).

Then, each peptide of the cross-linked pair was searched against the whole nrNCBI database by using the BLAST to control the presence of unique-species peptides.

The cross-linked peptides' description has been done using the widely accepted nomenclature reporting the longer chain as alpha, α and the smaller chain beta, β (Schilling nomenclature) [142].

5.6.2 Result analysis

In all three proteins (collagen 1(I), 1(III) and 2(I)), a vast extent of formaldehyde-based cross-links was highlighted. The totality of the cross-links identified is listed in Table V-8 (collagen 2(I)), Table V-9 (collagen 1(III)) and Table V-10 (collagen 1(I)) at the end of the chapter.

Interestingly, the different types of formaldehyde-induced cross-links have been identified between the three collagens with a similar proportion (Figure V-10). The covalent bonds occurred mostly between Lys-Arg and Arg-Arg, all with the variation mass of 24 Da (illustrated in grey and dark blue in the pie chart). These data are in accordance with previous observations in collagen [102] and in other proteins [114, 129] in which the predominant cross-links observed were lysine-arginine. The shift of 24 Da in the presence of arginine, which was also observed in the lysozyme model, might correspond to two methylene bridges of 12 Da each (Figure V-2 - C) or one dimeric interaction between two formed imines (24 Da) (Figure V-3).

Other cross-links were identified involving lysine-asparagine or lysine-glutamine depicted in the pie chart with light blue and orange, respectively. Only a few cross-links with the mass shift of 24Da in the absence of arginine were detected (five K-Q and two K-N). The presence of 24Da shift might be explained with different assumptions: the addition of a Schiff base (+12 Da) on another residue of the cross-linked pair, The presence of two methylene bridges or a dimeric interaction between two formed imines (+24 Da), as assumed in the recently proposed mechanism [114]. Lysine-histidine (yellow) was detected with a lower extent (formation of only two different reticulated peptidic pairs) with a mass shift of 12 Da.

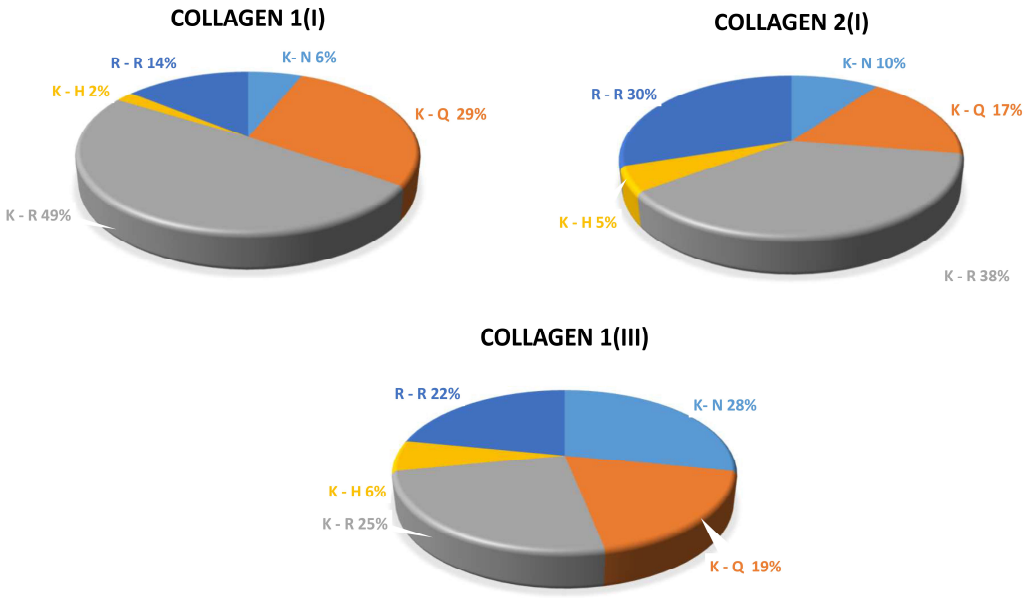


Figure V-10 Pie charts representing the extent of different types of formaldehyde-based cross-links in the three collagen proteins under investigation. The number of cross-links detected in different amino acid residues is reported in brackets.

Identification of reticulated peptides characteristic for a particular biological species

Numerous cross-linked pairs detected could not be discriminated among the three species under analysis. The great extent of these not-specific peptides detected, particularly in collagen 1(I), reflected the almost total sequence resemblance in the different species (Figure V-11).

Nonetheless, several species-diagnostic peptides for all three species (*Bos Taurus, Caprinae subfamily, Oryctolagus Cuniculus*) have also been documented, suggesting that all collagenous compounds present in the artwork established cross-links via formaldehyde. Interestingly, different reticulated peptides characteristic for *O.cuniculus* and *B.taurus* were observed in all three collagens, with a greater extent in collagen 2(I); conversely, peptides attributed to *Caprinae subfamily* were observed almost uniquely in collagen 1(III), with only two cross-link types detected in collagen 2(I).

Not unique peptides were also considered when discriminating for only one of the three species under analysis.

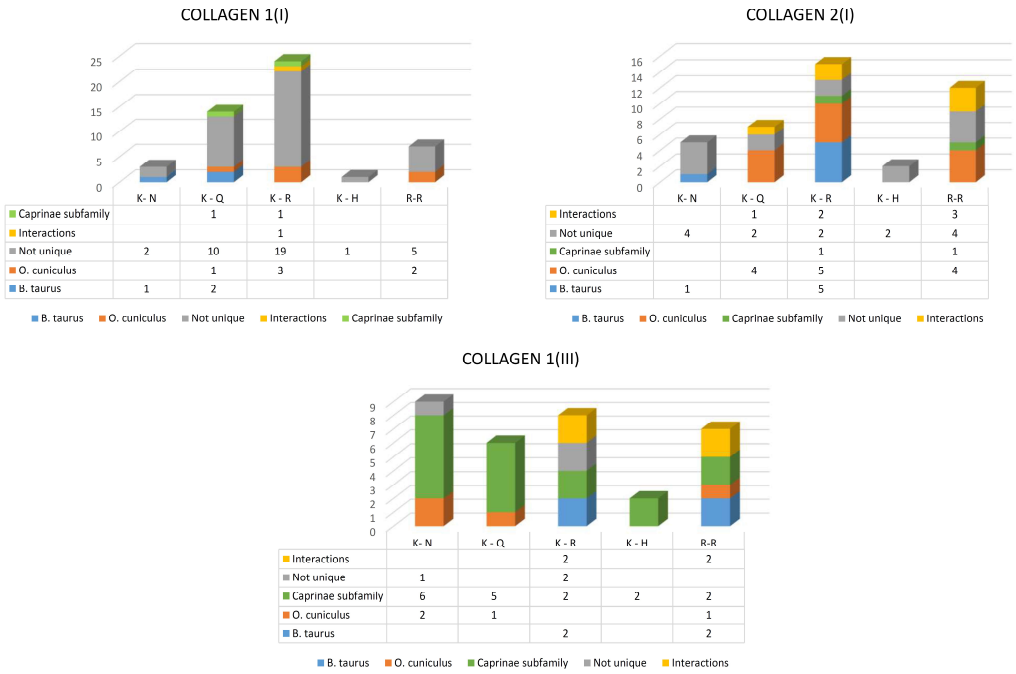


Figure V-11 Histograms representing the source of the cross-linked peptides identified in the three collagens under investigation.

The MS/MS spectrum of the charged ion 4^+ at m/z 717.6157 (Δm: -1.39 ppm) in Figure V-12 is reported as an example of unique reticulated peptides pairs. The fragmentation pattern's interpretation suggested the attribution of the value to a formaldehyde-based reticulation between lysine in position 652 of the peptide α ^{662}GAAGIpGGKGE**K**GETGLR^{673} and the arginine in position 669 of the peptide β ^{644}GDIGSpG**R**DGAR^{661}, with a mass shift of 24 Da. Both peptides refer to collagen 2 (I) protein, peptide α (662-673) is characteristic of the Bovinae subfamily (*Bos taurus, Bos indicus, Bos mutus and Bubalus bubalis*), while the peptide β is unspecific. One hydroxyprolination is detected in each peptide (P667 and P649, respectively).

The MS/MS spectrum displays a heterogeneous extent of values from various fragmentations. The fragment ions that have assisted in the hydroxyproline localisation are: y_1 and b_2 to b_6, b_8, b_9 in α peptide (in blue in Figure V-12) and y_1, y_3,y_4 and b_2-b_4, b_6 in β peptide (in red in Figure V-12). The attribution of the m/z value 593.2815 to $(\beta +12)^{+2}$ (in red in the figure) and m/z 841.9427 to $(\alpha +12)^{+2}$ (in blue in the figure) instead suggested symmetric fragmentation of the cross-linking bridge caused by HCD with the consequent formation of an ion composed by the single reticulated pair component plus 12 Da.

Numerous b- and y-ions were also identified, presenting a shift of +12Da (annotated with an asterisk); they are potentially formed by the symmetric fragmentation of the covalent bond and backbone.

Specifically, b_{13}^{+2*}(554.7863), b_{14}^{+2*}(619.3097), y_{12}^{+2*}(657.3429) were attributed to the peptide α (in blue in the figure), while y_8^{+2*}(422.2031), y_9^{+2*}(450.7153), y_{10}^{+2*}(507.2559) and b_9^{+2*}(442.2031) to the peptide β (in red in the figure).

Contrary to other studies in which the fragmentation completely broke the cross-link [114], in our case the assumption of the location of the bond at K652 – R669 was further supported by the observation of values corresponding to a unique fragmentation at the peptide backbone. Six values were ascribed to fragments of both peptides with 2C cross-link (in grey colour in the figure), e.g. 534.2742 (y_9_$y_6\beta^{+3}$), m/z 600,9640 (y_9_$y_8\beta^{+3}$), m/z 732,6908 (y_{13}_$b_9\beta^{+3}$), m/z 776.7228 (y_{12}_$y_{11}\beta^{+3}$), m/z 1041.021 (y_{13}_$b_8\beta^{+2}$), m/z 1098.5327 (y_{13}_$b_9\beta^{+2}$).

Furthermore, five identified values reflected different fragments of the ^{662}GAAGIpGGKGE**K**GETGLR^{673} peptide summed with the entire β peptide, preserving the bond (annotated with a dark blue in the figure): m/z 715.0188 (y_9_β^{+3}), m/z 757.717 (y_{10}_β^{+3}), m/z 776.7228 (y_{11}_β^{+2}), m/z 795.7306 (y_{12}_β^{+3}), m/z 833.4135 (y_{13}_β^{+3}).

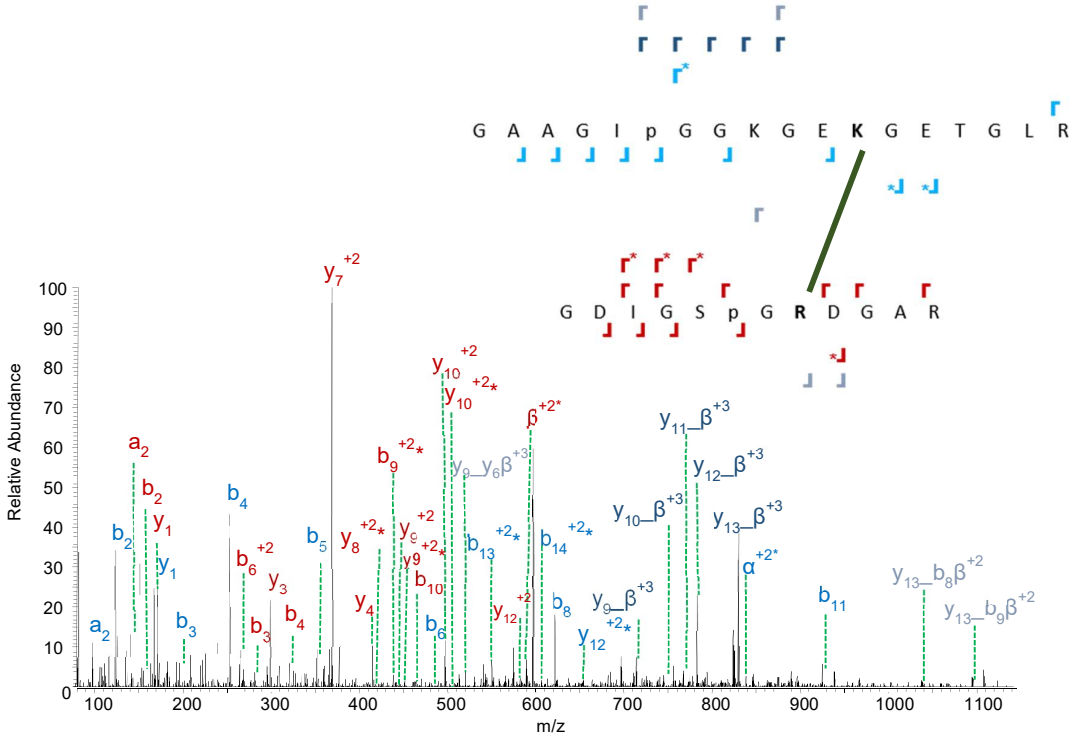


Figure V-12 The MS fragmentation spectrum of the 4+ charged ion at m/z 717.6157 (Δm: -1.39 ppm) has been attributed to a formaldehyde-based cross-link between two peptides: ^{662}GAAGIpGGKGEKGETGLR^{673} and ^{644}GDIGSpGRDGAR^{661} presenting both one hydroxyprolination (P667 and P649). The peptide α (662-673) refers to the *Bovinae* subfamily (*Bos taurus, Bos indicus, Bos mutus and Bubalus bubalis*). The peptide β instead is not specific.

Finally, the identification of three y- fragments of the peptide β (y_7^{+2}, y_9^{+2}, y_{10}^{+2}) without any mass shift suggests a complete fragmentation of cross-link (-24 Da) also occurred.

Interestingly, numerous peptides identified in their cross-linked form were not previously detected with the classical and linear bottom-up investigation, some of whom were also species-specific. For instance, the spectrum in Figure V-13 shows the fragmentation of the triply charged ion m/z 820.4151 (Δm -1.22 ppm) attributed to two peptides ^{490}GPAGPSGTPGE**K**GPAGER^{507} and ^{967}GSPGP**Q**GVK^{975} bonded by a methylene bridge lysine 501 – glutamine 972 (+12 Da). The two peptides refer to collagen 1(III). The peptide α (490-507) is characteristic of *Oryctolagus cuniculus* (European rabbit); the peptide β (967-975) is not unique, but it is not attributed to proteins from *Bovinae* (V 976-> I 976) or *Caprinae* subfamily (V 976-> I 977).

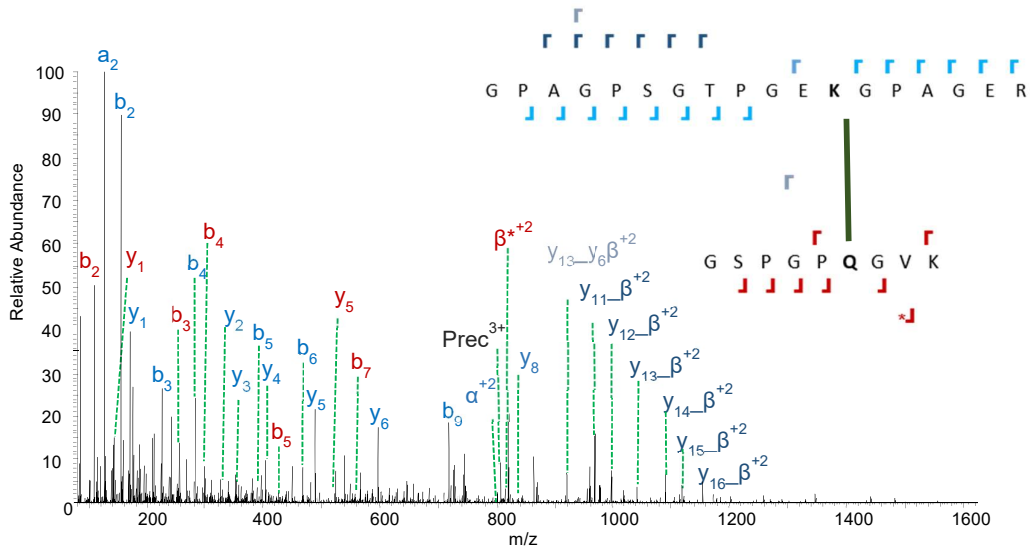


Figure V-13 Fragmentation of the triply charged ion m/z 820.4151 (Δm= -1.22 ppm) attributed to two peptides ^{490}GPAGPSGTPGEKGPAGER^{507} and ^{967}GSPGPQGVK^{975} bonded by a methylene bridge (+12 Da). The two peptides refer to collagen 1(III). The peptide α (490-507) is characteristic of *Oryctolagus cuniculus* (European rabbit) specie; the peptide β (967-975) is not unique but it does not refer to proteins from *Bovinae* (V 976-> I 976) or *Caprinae* subfamily (V 976-> I 977).

The pairs of ions of the component peptides resulting from the fragmentation of the covalent bond were identified in the MS/MS spectrum: β peptide was detected with the methylene bridge attached $(\beta + 12)^{+2}$ at m/z 837.4345, while α peptide without $(\alpha)^{+2}$ at m/z 811.4001. The assignment of y_1 to y_6 and b_2 to b_9 ion fragments to the peptide α and y_1 and b_2 to b_5 ion fragments to peptide β provided accurate identification of the two sequences.

The hypothesis of the proposed cross-link (K501 – Q972) was supported by the determination of one value attributed to fragments of both peptides covalently bonded represented in grey in the spectrum (m/z 919.9755 $y_{13}\_y_6\beta^{+2}$), and six values are presenting y-fragments of different measures bonded with the entire peptide β, represented in dark blue: m/z 968.5017 ($y_{11}\_\beta^{+2}$), m/z 997.0122 ($y_{12}\_\beta^{+2}$), m/z 1040.5284 ($y_{13}\_\beta^{+2}$), m/z 1089.0540 ($y_{14}\_\beta^{+2}$), m/z 1117.5651 ($y_{15}\_\beta^{+2}$), and m/z 1153.0832 ($y_{16}\_\beta^{+2}$).

Besides ^{490}GPAGPSGTPGEKGPAGER^{507}, five other peptides unique for *O. cuniculus* species were identified only in the cross-linking investigation as components of cross-linked pairs. Specifically, three peptides of collagen 2 (I) protein and three peptides of collagen 1(III) were highlighted in addition to the classical investigation. (Table V-6). These results strongly reinforced the assumption that a rabbit-derived proteinaceous material was used to restore the manuscript.

Table V-6 Unique peptides identified in the samples only as reticulated pair.

Peptide unique	Cross-linking pair	Cross-link	m/z	z	Ppm Error	Species
Collagen 2(I)						
GLPGEFGLPGPAGPRGER GAPGESGAAGpPGPIGSR	GPRGER	R589-R41	798.8084	5	2.00	*O.cuniculus*
GVmGpQGAR	GPTGDpGKnGDK	Q160-K504	696.32	3	1.74	*O.cuniculus*
AqpEnISVKnWYK	GPSGDRGpR Bovinae/Caprinae	K1212-R38	845.07	3	-0.70	*O.cuniculus*
Collagen 1(III)						
GPAGPSGTPGEKGPAGER	GSPGPQGVK	K50-Q972	820.4151	3	-1.22	*O.cuniculus,*
GESGKPGVnGQNGER	PGAKGER	N949-K784	743.03	3	-2.68	*O.cuniculus*
GDAGQPGEKGSPGAqGpP GAPGPLGLAGITGAR	KNPARNCR Bovinae/Caprinae	K930-R1260	1317.70	3	2.40	*O.cuniculus*
Collagen 1(III)						
pGERGFPGPpGmK	GPpGPqGPRGDK	R242-R1091	854.74	3	3.43	*B.taurus,B.bubalis, B.indicus x B.taurus, B.b.bison, B.indicus,B. mutus*
GENGVPGENGAPGPmGP RGAPGERGR	GDSGAPGERGpp GAGGPPGpR	R308-R683	881.41	5	-0.42	*B.taurus, B.indicus xB.taurus,B.b.bison B.indicus, B.mutus*

Two peptides of collagen 1(III) specific for the *Bovinae* subfamily were also detected solely in their reticulated form. The data well depicts the difficulties in the detection of reticulated fraction from the standard proteomic analysis.

Interactions among collagens from different biological breeds

The significance of reticulated pairs composed of peptides not shared among the three distinct species suggested the formation of inter-protein bonds among proteins of different species (illustrated in yellow in Figure V-11 and reported in the tables Table V-8, Table V-9, and Table V-10).

Specifically, five reticulated peptidic pairs were identified in collagen 2(I): three pairs were attributed to *O. cuniculus* peptide (not unique) bonded to a *B. Taurus* peptide (in one pair unique); one pair attributed to *Caprinae* (not unique) and *O. cuniculus* (unique), and one pair attributed to a peptide unique of *O. cuniculus* bonded to a peptide ascribable to either *B. taurus* or *Caprinae*. Four reticulated peptide pairs were detected in collagen 1(III): two pairs attributable to *O. cuniculus* (not unique) and *B. taurus* (unique in one pair), and two pairs attributed to *O. cuniculus* (unique in one pair) cleaved to a peptide ascribable to either *B. taurus* or *Caprinae*. For collagen 1(I), one peptidic pair was observed with a peptide ascribable to *O. cuniculus* (not unique) linked with a peptide ascribable to *B. taurus* (not unique).

Figure V-14 shows an example of the fragmentation spectrum of the 4^+ charged ion m/z 915.4427 (Δm 0.11ppm) attributed to the formaldehyde-based cross-link (K261-R1091) between ^{241}GLpGppGIKGpAGMpGFpGM**K**GHR^{264} and ^{1083}GppGpQGp**R**GDK^{1094}. The two peptides (presenting all proline oxidized) are attributable to collagen 1(III) from two biological species. Even though none of these peptidic residues is unique for a particular species, the peptide α (241-264) is assignable to *O. cuniculus* but not to *Bovinae* (L242->F244 and I248->M250) or *Caprinae* (I248-> M251). The peptide β (1083-1094) instead might be attributed to *Bovinae* subfamily collagen but not to *Caprinae* (^{1084}GAPGPQGPRGDK^{1095}) or *O. Cuniculus* (^{1081}GAPGPQGPRGDK^{1092}).

The detection of y- (y_1-y_3) and b- (b_2,b_4,b_6,b_7,b_{12},b_{14},b_{15},b_{17},b_{19},b_{20}) fragment ions attributable to peptide α and 10 ion fragments (y_1,y_3 and b_2,b_3,b_5,b_6,b_8) ascribable to peptide β assists the characterization of the residues.

Similarly to the previously described K-R bridge spectrum, both y- and b- values of both peptides were highlighted, presenting a shifting mass of 12 (asterisk) related to a fragmentation of the bridge. Specifically, 509.293 (y_4^{+2}*), 957.470 (y_8^{+2}*), 1127.539 (y_{10}^{+2*}), 1386.641 (y_{13}^{+2*}) , 1979.959 (y_{19}^{+2*}), 2122.115 (b_{22}^{+2*}) were attributed to the peptide α, and 487.264 (y_4*), 657.333 (y_6*), 785.400 (y_7*), 898.437 (y_8*), 955.460 (y_9*), 534.759 (b_{10}^{+2}*) to the peptide β. Only three (y_8^{+2}, y_{16}^{+2}, y_{23}^{+2}) and two (y_9^{+2}, b_{10}^{+2}) ion fragments were identified respectively for α and β peptides, caused by the complete fragmentation of the bond.

The accurate localisation of the covalent bond was possible due to the detection of values matching with both peptides' fragments covalently bonded (reported in grey in the Figure): m/z 866.4023 (y_5_$b_{11}\beta^{+2}$), m/z 882.9380 (y_7_$y_9\beta^{+2}$), m/z 910.943 (y_5_$y_{11}\beta^{+2}$), m/z 927.947 (y_8_$y_8\beta^{+2}$), m/z 1152.551 (y_{23}_$b_{11}\beta^{+2}$). Furthermore, four-valued identified reflected fragments of peptide α together with the entire peptide β, plus 2C: m/z 1012.4833 (y_{17}_β^{+3}), m/z 1031.4935 (y_{18}_β^{+3}), m/z 1075.1712 (y_{19}_β^{+3}), m/z 1183.0665 (b_{10}_β^{+2}).

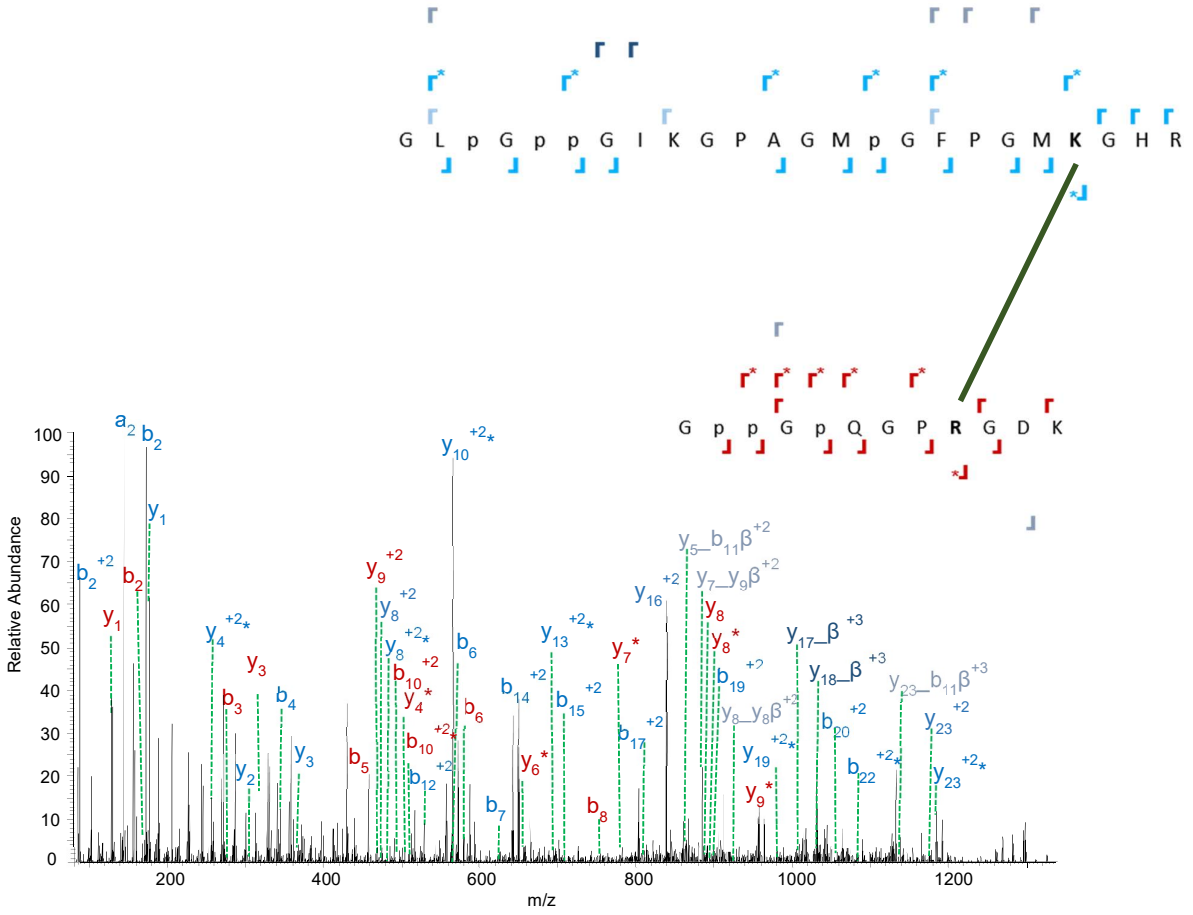


Figure V-14 Fragmentation of the ion m/z 915.4427 with a charge 4^+ (Δm 0.11ppm) has been attributed to ^{241}GLpGppGIKGpAGMpGFpGMKGHR^{264} and ^{1083}GppGpQGpRGDK^{1094}, presenting all proline oxidized and covalently bonded by a formaldehyde-based bridge. The two peptides are both from collagen 1(III) but not from the same species; even though none of the peptidic residues is unique for a particular specie, the peptide α (241-264) is attributable to *Oryctolagus cuniculus* but not to *Bovinae* (L242->F244 and I248->M250)or *Caprinae* (I248-> M251). The peptide β (1083-1094) instead might be attributed to *Bovinae* subfamily collagen but not to *Caprinae* (^{1084}GAPGPQGPRGDK^{1095}) or *O. Cuniculus* (^{1081}GAPGPQGPRGDK^{1092}).

Interactions among different collagens from the same biological breed

A further investigation was performed focusing on the interactions between collagen 1(I), collagen 1(III) and collagen 2(I) from *Bos Taurus* species to detect the presence of reticulations between different collagen proteins. Several cross-links with a mass shift of either 24 Da or 12 Da formed between collagen 1(III) and collagen 1(I) and between collagen 2(I) and collagen 1(I), only one bond was highlighted between 2(I) and 1(III). The list of reticulated peptide pairs is detailed in Table V-7.

Table V-7 Interactions among different collagens from the same biological breed

Cross-link	Sample	RT	Scan	Sequence	shift	Matches 1 and 2	Charge	m/z	Mass	Mass error ppm
Collagenalpha-1(I)- Collagen 1(III)										
K585-R509	S1	142.38	30002	^{574}GQAGVMGFPGPKGAAGEPGKAGER$^{597(COL1AI)}$-^{504}GPpGDRGGpGPAGPRGVAGEpGR$^{526(COL1AIII)}$	24	3 - 13	4	1092.5	4364.1	-1.01
	A_S1	74.57	14580	^{574}GQAGVMGFPGPKGAAGEPGKAGER$^{597(COL1AI)}$-^{504}GPpGDRGGpGPAGPRGVAGEpGR$^{526(COL1AIII)}$		3 - 10	4	1092	4364.1	-1.42
K276-Q988	S1	219.78	46378	^{268}GFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQmGPR$^{303(COL1AI)}$-^{983}pGPSGQNGER$^{992(COL1AIII)}$	24	4 - 6	4	1083.5	4330	-0.62
	A_S1	97.53	19543	^{268}GFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQmGPR$^{303(COL1AI)}$-^{983}pGPSGQNGER$^{992(COL1AIII)}$		7 - 3	4	1083.5	4330	-0.62
K551-Q988	S1	47.81	9911	^{538}GLTGSpGSpGPDGKTGPPGPAGQDGR$^{563(COL1AI)}$-^{983}PGPSGQNGER$^{992(COL1AIII)}$	24	21 - 2	4	843.4	3369.6	4.11
		51.22	10632	^{538}GLTGSpGSPGPDGKTGPPGPAGQDGR$^{563(COL1AI)}$-^{983}PGPSGQNGER$^{992(COL1AIII)}$		20 - 4	3	1118.9	3353.6	-3.88
		51.78	10751	^{538}GLTGSpGSPGPDGKTGPPGPAGQDGR$^{563(COL1AI)}$-^{983}PGPSGQNGER$^{992(COL1AIII)}$		23 - 3	3	1118.9	3353.6	-3.88
		52.35	10872	^{538}GLTGSpGSPGPDGKTGPPGPAGQDGR$^{563(COL1AI)}$-^{983}PGPSGQNGER$^{992(COL1AIII)}$		23 - 3	3	1118.9	3353.6	-3.88
		52.99	11007	^{538}GLTGSpGSPGPDGKTGPPGPAGQDGR$^{563(COL1AI)}$-^{983}PGPSGQNGER$^{992(COL1AIII)}$		22 - 4	3	1118.9	3353.6	-3.88
Collagen 2(I) – Collagen 1(III)										
N688-K1094	S1	58.91	12265	^{674}GApGAIGAPGPAGAnGDR$^{691(COL2AI)}$-^{1083}GPPGPqGPRGDKGETGER$^{1100(COL1AIII)}$	24	17 - 4	4	835.4	3337.6	-0.96
		61.28	12766	^{674}GApGAIGAPGPAGAnGDR$^{691(COL2AI)}$-^{1083}GPPGPqGPRGDKGETGER$^{1100(COL1AIII)}$		14 - 4	4	835.4	3337.6	-1.16
Collagen 2(I) – Collagen 1(I)										
N748-R1256	S1	111.39	23427	^{746}GENGPVGPTGPVGAAGPSGPnGpPGPAGSR$^{775(COL2AI)}$-^{1253}nPARTCR$^{1259(COL1AI)}$	24	19 - 2	4	870.91	3479.6	2.03
		111.98	23551	^{746}GEnGPVGPTGPVGAAGPSGPnGpPGPAGSR$^{775(COL2AI)}$-^{1253}NPARTCR$^{1259(COL1AI)}$		10 - 2	4	870.91	3479.6	2.03
	A_S1	90.11	17926	^{746}GEnGPVGPTGPVGAAGPSGPnGpPGPAGSR$^{775(COL2AI)}$-^{1253}NPARTCR$^{1259(COL1AI)}$		17 - 2	4	870.91	3479.6	-1.91
K1095-H942	S1	62.21	12963	^{1084}GPAGpqGPRGDKGETGEQGDR$^{1104(COL1AI)}$-^{925}DGnPGnDGPPGRDGQPGHK$^{943(COL2AI)}$	12	9 - 4	4	992.69	3966.7	-3.67
		62.26	12973	^{1084}GPAGpqGPRGDKGETGEQGDR$^{1104(COL1AI)}$-^{925}DGnPGnDGPPGRDGQPGHK$^{943(COL2AI)}$		11 - 5	3	1323.3	3966.7	-2.21
		64.34	13414	^{1084}GPAGpqGPRGDKGETGEQGDR$^{1104(COL1AI)}$-^{925}DGnPGnDGPPGRDGQPGHK$^{943(COL2AI)}$		8 - 3	4	992.69	3966.7	-2.55
N912-K906	S1	56.17	11683	^{905}GPpGnVGNPGVnGAPGEAGR$^{924(COL2AI)}$-^{903}EGSKGPR$^{909(COL1AI)}$	12	9 - 3	3	845.07	2532.2	-4.35
		65.70	13705	^{905}GPpGnVGNpGVnGAPGEAGR$^{924(COL2AI)}$-^{903}EGSKGPR$^{909(COL1AI)}$		11 - 6	3	850.4	2548.2	3.86
Q1068-K236	S1	150.24	31666	^{1052}GPAGpSGpAGKDGRIGQpGAVGPAGIR$^{1078(COL2AI)}$-^{220}GpPGPPGKnGDDGEAGKPGR$^{239(COL1AI)}$	12	9 - 5	4	1084	4332.1	-3.95
		150.83	31791	^{1052}GPAGPSGpAGKDGRIGQpGAVGpAGIR$^{1078(COL2AI)}$-^{220}GpPGPPGKnGDDGEAGKPGR$^{239(COL1AI)}$		9 - 5	4	1084.1	4332.1	-3.95
	A_S1	69.28	13463	^{1052}GPAGpSGPAGKDGRIGQPGAVGPAGIR$^{1078(COL2AI)}$-^{220}GpPGppGKnGDDGEAGKPGR$^{239(COL1AI)}$		6 - 5	4	1084	4332.1	-2.43
Q249-K508	S1	46.05	9539	^{237}PGRpGERGppGpQGAR$^{252(COL1AI)}$-^{505}AGEKGHAGLAGAR$^{517(COL2AI)}$	12	8 - 7	3	952.48	2854.4	3.87

		46.20	9572	^{237}PGRpGERGppGpQGAR$^{252(COL1AI)}$- ^{505}AGEKGHAGLAGAR$^{517(COL2AI)}$		7 - 6	3	952.48	2854.4	3.87
Q456-K352	S1	50.48	10475	^{448}GEPGPTGIqGppGPAGEEGKR$^{467(COL1AI)}$- ^{341}GLVGEPGpAGSKGESGnK$^{358(COL2AI)}$	12	4 - 8	5	738.76	3688.7	0.64
		50.65	10512	^{448}GEPGPTGIqGPpGpAGEEGKR$^{467(COL1AI)}$- ^{341}GLVGEpGPAGSKGESGnK$^{358(COL2AI)}$		12 - 19	4	923.19	3688.7	0.12
	A_S1	47.37	8765	^{448}GEPGPTGIqGPpGpAGEEGKR$^{467(COL1AI)}$- ^{341}GLVGEpGPAGSKGESGnK$^{358(COL2AI)}$		10 - 12	5	738.76	3688.7	0.64

5.7 Discussion of the results achieved from the three investigations

This study is the first example of the identification of formaldehyde cross-linked glue in art conservation. Both cross-linked peptides and protein chemical modifications signing the use of formaldehyde treatment (methylations) were accurately characterized in the studied glue repairs material, informing on the restoration practice at the Vatican library. Unique peptides of three biological origins (*O. cuniculus, B. taurus* and *Caprinae* subfamilies, most likely *C. hircus*) were found presenting these modifications (i.e. methylations and formaldehyde-based cross-links); therefore, formaldehyde was applied in conditions that provided its reaction with all proteinaceous compounds. The application of warm and liquid gelatin mixed to formalin on the parchment (as written in the historical documents) certainly contributed to its diffusion (and possible penetration into the pages) and formaldehyde interaction with different collagen-based materials. This conservation treatment is likely the reason that the parchments pages are brittle and fragile.

Considering the three identified species, it can be remarked that the glue produced from rabbit skin is often mixed with other glues to reduce its fragility preserving its characteristic flexibility, which is not obtainable in other cases. Collagens from the *Caprinae* and *Bovinae* subfamilies might be associated with conservation actions. In his writings, Franz Ehlre referred to gelatin (or gelatin-formol) without specifying the biological origin [83]. Further analyses on samples taken from the manuscript (the same page from an area without visible restoration and other pages) are currently investigated.

6. Conclusions

In this study, the research extended beyond classical protein identification, also investigating the chemical and structural modifications of the proteinaceous compounds. Even though the invasive conservation treatment on the manuscript prevented a straightforward interpretation, the analysis succeeded in understanding the object's conservation history. The detection of lysine methylations and most of the specific markers of cross-links induced by formaldehyde in the manuscript provided the first analytical proofs of the formaldehyde-based treatment during parchment conservation at the Vatican library. The identification of the formaldehyde cross-links moves the conversation of gelatin/formol treatment from the anecdotal to the scientifically confirmed. Having a better understanding of the repair material's composition may help determine future conservation approaches. Furthermore, the definition of the fragmentation pattern specific for formaldehyde cross-linking could assist in future studies of other manuscripts, where pages are visibly rigid and brittle without apparent reason.

Additionally, it has been proven that the standard bottom-up approach alone might fail to detect several peptides when chemically modified or structurally unreachable, hindering the correct interpretation of the results. Therefore an unbiased modification search and the detection of reticulated forms (naturally-induced or chemically induced) could enhance the data acquired and improve the data interpretation.

Table V-8 List of the cross-links induced by formaldehyde detected in the proteins of collagen 2(I) of different biological breeds (*B. taurus*, *C. subfamily*, *O. cuniculus*, not specific and interactions).

Cross-link	Sample	RT	Scan	Sequence	shift	Matches 1 and 2	Charge	m/z	Mass	Mass error ppm
				Bos taurus						
K504-N433	S1	80.4	16845	^{485}GEpGnIGFPGpKGPSGDPGK^{504}-^{431}GPNGDSGR^{438}	12	4-2	3	890.08	2667.2	0.71
	A_S1	33.7	6923	^{497}GPSGDpGKAGE^{507}-^{431}GPNGDSGR^{438}		7-6	3	586.59	1756.8	4.41
K652-R669	S1	48.87	10135	^{644}GAAGIpGGKGEKGETGLR^{661}-^{662}GDIGSpGRDGAR^{673}	24	12-7	5	574.29	2866.4	1.96
		48.64	10087			8-7	4	717.61	2866.4	1.39
		46.34	9602			7-8	3	956.48	2866.4	1.42
K736-R154	S1	58.05	12081	^{713}GEVGPAGpNGFAGpAGAAGqPGAKGER^{739}-^{151}PGERGVVGpqGAR^{163}	24	6-5	3	1243.9	3728.8	1.42
		159.39	33599	^{713}GEVGPAGPNGFAGPAGAAGQPGAKGER^{739}-^{151}pGRPGERGVVGPQGAR^{163}		5-3	5	802.01	4005	2.72
R154-K504	S1	41.49	8573	^{151}pGERGVVGpQGAR^{163}-^{497}GPSGDPGKAGEK^{508}	24	5-6	4	609.31	2433.2	-3.03
		42.44	8775	^{151}PGERGVVGpQGAR^{163}-^{497}GPSGDPGKAGEK^{508}		9-13	3	806.74	2417.2	-4.05
R41-K1318	S1	55.47	11534	^{33}GPSGDRGPRGER^{44}-^{1318}KTNEWQK^{1324}	24	6-3	4	550.03	2196.1	-3.24
		56.05	11658	^{33}GpAGDRGPRGER^{44}-^{1318}KTNEWQK^{1324}		6-4	4	550.03	2196.1	-3.24
R691-K1062	S1	144.56	30463	^{670}DGARGAPGAVGAPGPAGAnGDRGEAGPAG PAGPAGPR^{706}-^{1052}GPAGPSGPAGKDGR^{1065}	24	6-3	4	1096	4380.1	-1.11
	A_S1	125.01	25301	^{670}DGARGAPGAVGAPGPAGANGDRGEAGPAG PAGPAGPR^{706}-^{1052}GPAGpTGPAGKDGR^{1065}		7-6	5	882.83	4409.1	-1.45
				Caprinae subfamily						
K813-R820	S1	69.78	14582	^{793}TGppGpAGISGpPGpPGPAGKEGLR^{817}-^{818}GPRGDQGPVGR^{828}	24	15-9	5	684.75	3418.7	0.09
	A_S1	76.02	14920			13-8	4	855.67	3418.7	0.51
R52-R41	S1	140.94	29696	^{45}GPpGpPGRDGDDGIPGPPGPPGPPGPPGLGGN FAAQFDGK^{84}-^{39}GPRGER^{44}	24	5-2	4	1100	4396.1	0.96
	A_S1	86.01	17059	^{45}GPpGPPGRDGDDGIPGPPGPPGPPGPPGLGGN FAAQFDGK^{84}-^{39}GPRGER^{44}		5-2	4	1096	4380.1	2.00
				Oryctolagus cuniculus						
K504-Q160	S1	36.80	7582	^{497}GPTGDpGKnGDK^{508}-^{155}GVmGpQGAR^{163}	24	4-5	3	696.32	2085.9	- 4.03
		41.97	7741			3-4	4	522.47	2085.9	- 4.13
R589-R41	S1	158.96	33508	^{572}GLPGEFGLPGPAGPRGERGAPGESGAAGpPG PIGSR^{607}-^{39}GPRGER^{44}	24	3-3	4	998.25	3989	2.41
		159.60	33644			4-2	5	798.81	3989	2.4176
K1212-R38	S1	168.78	35581	^{1204}AqpEnISVKnWYK^{1216}-^{33}GPSGDRGpR^{41}	24	2-6	3	845.07	2532.2	-0.70
		130.21	27408	^{1204}AqPEnISVKnWYK^{1216}-^{33}GpSGDRGPR^{41}		2-5	3	839.74	2516.2	2.60
	A_S1	60.60	11683	^{1204}AqpENISVKnWYK^{1216}-^{33}GpSGDRGpR^{41}		2-3	3	850.07	2547.2	1.49
R669-R38	S1	56.26	11702	^{662}GEIGnPGRDGAR^{673}-^{33}GPAGDRGPR^{41}	24	8-4	3	702.35	2104	4.38
		73.81	15439	^{662}GEIGNpGRDGARGAPGAVGApGpAGATGDR ^{691}-^{33}GPAGDRGPR^{41}		6-3	4	907.69	3626.7	1.59
Q123-K352	A_S1	43.00	7957	^{121}PGQTGPAGAR^{130}-^{346}PGPAGTKGE^{354}	12	4-2	3	579.3	1734.9	-2.66
		78.45	15438	^{118}PGEPGQTGpAGARGpPGPPGK^{138}-^{346}PGPAGTKGE^{354}		2-3	4	689.34	2753.3	4.46
Q367-K1012	A_S1	104.63	21087	^{361}pGSAGPQGppGpSGEEGKR^{379}-^{1009}PGDKGpR^{1015}	12	2-2	4	645.55	2578.2	-2.11
		112.30	22780			4-3	3	860.41	2578.2	-1.94
Q123-K1012	A_S1	101.03	20287	^{98}GPpGAAGAPGpqGFQGpAGEPGEPGQTGPAG AR^{130}-^{1007}GEpGDKGPR^{1015}	12	10-4	4	971.7	3882.8	-2.88
		101.23	20329			11-4	3	1295.3	3882.8	-2.88
		102.16	20532			9-3	5	777.56	3882.8	-2.88
K196-R38	A_S1	52.57	9993	^{188}GqPGAPGVKGE^{198}-^{33}GPAGDRGpR^{41}	24	3-4	3	640.32	1917.9	2.19
		52.67	10015			4-3	3	640.32	1917.9	2.19
K352-R150	A_S1	19.31	2691	^{346}PGPAGTKGE^{354}-^{148}PGRPGER^{154}	24	2-3	4	401.96	1603.8	0.76
		29.69	5000	^{346}pGPAGTKGE^{354}- ^{148}PGRPGE^{153}		5-5	3	488.91	1463.7	2.67
R438-R38	S1	41.71	8621	^{431}GPNGDSGRPGEPGLmGPR^{448}-^{33}GPAGDRGPR^{41}	24	4-4	3	891.43	2671.3	2.11
		43.25	8946	^{431}GPNGDSGRPGEpGLmGpR^{448}-^{33}GpAGDRGPR^{41}		3-9	3	907.42	2719.2	-1.19
K352-R379	A_S1	76.28	14975	^{346}PGpAGTKGESGnK^{358}-^{377}GKRGSpGE^{384}	24	5-6	4	511.5	2042	2.27
		78.43	15434			4-6	3	681.66	2042	2.01
R1003-K352	A_S1	45.98	8581	^{995}GpSGpqGIRGDK^{1006}-^{346}pGpAGTKGE^{354}	24	6-9	4	518.25	2069	-4.49
		47.33	8871			4-5	3	690.66	2069	-4.05
R691-R150	A_S1	78.98	15548	^{674}GApGAVGAPGPAGATGDRGEAGAAGpAGp AGPR^{706}-^{148}PGRpGE^{153}	24	5-4	4	849.41	3393.6	1.96
		80.08	15787			8-2	5	679.72	3393.6	1.96
				Not unique						
Q366-K352	S1	48.21	9996	^{359}GEpGAVGQpGPpGPSGEEGKR^{379}-^{341}GLVGEpGPAGSKGESGnK^{358}	12	12-18	4	919.94	3675.7	-0.12
		49.47	10263	^{359}GEpGAVGQpGPpGPSGEEGKR^{379}-^{341}GLVGEpGpAGSKGESGNK^{358}		20-13	4	923.69	3690.7	-3.75
		50.05	10385	^{359}GEpGAVGQpGPpGPSGEEGKR^{379}-^{341}GLVGEpGpAGSKGESGNK^{358}		10-15	5	739.14	3690.7	-3.75
	A_S1	81.50	16084	^{359}GEPGAVGQPGPPGPSGEE^{376}-^{346}pGPAGSKGESGNKGE^{354}		2-4	3	1006.5	3016.4	-3.42

		73.06	14297	^{361}PGAVGQPGPPGPSGE^{375}-^{346}pGPAGSKGE^{354}		6-8	3	710.67	2129	-3.18
N927-K1012	S1	84.44	17701	^{925}DGNPGnDGPPGRDGqpGHK^{943}-^{1007}GEpGDKGPR^{1015}	12	4-5	4	708.07	2828.2	-1.48
	A_S1	96.50	19319	^{925}DGNpGNDGPpGR^{936}-^{1009}PGDKGpR^{1015}		2-2	2	969.44	1936.9	-2.88
R438-R41	S1	89.0	18669	^{431}GPnGDSGRpGEPGLMGPRGFPGSPGnIGPAGK^{462}-^{39}GPRGER^{44}	24	8-3	3	1233.9	3698.8	-3.27
	A_S1	97.23	19481	^{431}GpnGDSGRPGEPGLMGPR^{448}-^{33}GPAGDRGpRGE^{43}		7-5	3	959.11	2874.3	4.05
R44-R1003	S1	63.50	13236	^{39}GPRGERGppGpPGR^{52}-^{995}GPSGPQGIRGDK^{1006}	24	3-4	4	657.34	2625.3	1.12
		70.28	14687			7-5	5	526.06	2625.3	1.31
		72.21	15098	^{39}GPRGERGppGppGR^{52}-^{995}GPSGpQGIRGDK^{1006}		6-5	5	532.4	2657.3	1.70
		79.90	16739			5-4	4	665.33	2657.3	1.39
K1006-R1159	S1	57.33	11929	^{995}GPSGPqGIRGDKGEPGDK^{1012}-^{156}NpARTCR^{1162}	24	7-5	5	534.06	2665.3	-1.51
		58.43	12162	^{995}GPSGpqGIRGDKGEPGDK^{1012}-^{156}NPARTCR^{1162}		7-4	4	667.32	2665.3	-1.46
		59.08	12300	^{995}GPSGPqGIRGDKGEPGDK^{1012}-^{156}NpARTCR^{1162}		9-5	3	889.43	2665.3	-1.12
K378-R820	S1	51.36	10663	^{335}GEpGAVGqpGPPGPSGEEGKR^{379}-^{818}GPRGDQGPVGR^{828}	24	8-7	3	1037.8	3110.5	-2.28
		105.55	22181	^{353}GESGnKGEPGSAGpQGPPGPSGEEGKR^{379}-^{814}EGLRGpRGDQGPVGR^{828}		10-6	3	1376.3	4125.9	-2.32
R1003-R44	S1	50.46	10472	^{995}GpSGPqGIRGDK^{1006}-^{42}GERGppGpPGR^{52}	24	9-6	3	778.38	2332.1	0.44
		62.72	13071	^{995}GpSGPqGIRGDK^{1006}-^{42}GERGPPGppGR^{54}		7-5	3	773.05	2316.1	-3.23
K1006-N1239	A_S1	125.70	25435	^{995}GPSGPqGIRGDKGEPGDK^{1012}-^{1238}YNVE^{12141}	12	2-5	3	763.36	2287.1	2.09
		125.70	25451			2-2	4	572.77	2287.1	2.08
H144-K138	A_S1	14.19	1780	^{142}DGHpGKpGRpGE^{153}-^{131}GPpGPPGKAGE^{141}	12	4-10	4	561.27	2241	-1.01
		18.28	2482	^{142}DGHpGKpGRPGE^{153}-^{131}GPpGPPGKAGE^{141}		4-7	3	742.69	2225	1.89
		84.36	16699	^{139}AGEDGHpGKpGRpGE^{153}-^{131}GPpGpPGKAGE^{141}		2-6	3	839.05	2514.1	-3.92
K352-H144	A_S1	91.69	18261	^{346}PGpAGSKGESGNK^{358}-^{142}DGHPGKPGRPGER^{154}	12	8-2	4	643.82	2571.3	1.23
		88.56	17595			7-2	3	858.09	2571.3	1.25
K352-N205	A_S1	51.05	9665	^{199}PGAPGENGTPGqTGAR^{214}-^{345}pGPAGSKGE^{354}	12	2-4	3	765.36	2293.1	-0.36
		77.92	15326	^{190}NGTpGQTGARGLpGE^{204}-^{345}pGpAGSKGE^{354}		10-10	3	762.69	2285.1	1.93
		88.90	17684	^{199}PGAPGENGTPGqTGAR^{214}-^{345}PGPAGSKGE^{354}		2-2	3	760.03	2277.1	0.47
K352-Q160	S1	72.84	15233	^{346}pGPAGSKGESGnK^{358}-^{155}GVVGpqGAR^{163}	12	7-2	3	691.01	2070	1.80
	A_S1	79.14	15582	^{346}PGpAGSKGESGnK^{358}-^{155}GVVGpqGAR^{163}		8-2	3	691.01	2070	2.74
N912-K1246	A_S1	91.66	18255	^{905}GPpGnVGNpGVNGAPGE^{921}-^{1272}GVTTKE^{1274}	12	5-2	3	723.34	2167	4.48
		130.74	26351	^{905}GPpGNVGNPGVNGApGE^{921}-^{1238}YNVEGVTTKE^{1247}		6-2	3	891.42	2671.2	-4.39
R820-R571	A_S1	50.63	9574	^{818}GpRGDqGpVGR827-^{558}pGERGLpGE^{576}	24	3-4	3	699	2094	2.28
		98.04	19657	^{818}GpRGDqGpVGRSGE^{831}-^{568}pGERGLPGE^{576}		2-2	3	784.7	2351.1	1.79
	Interactions									
R661-R38	S1	105.44	22158	^{656}GETGLRGDIGSpGR^{669}-^{33}GPAGDRGPR^{41}	24	2-4	3	765.05	2292.1	-3.22
	A_S1	115.11	23384	^{656}GETGLRGDIGSPGRDGAR^{673}-^{33}GpAGDRGPRGER^{44}		2-2	3	1012.2	3033.5	0.18
R924-R661	S1	42.51	8789	^{905}GPpGAVGSpGVnGAPGEAGRDGNPGSDGpPGR^{936}-^{656}GETGLRGEIGnPGR^{669}	24	5-2	4	1074.8	4295	-2.02
		51.17	10621	^{905}GPpGAVGSPGVnGAPGEAGRDGnpGSDGPPGR^{936}-^{656}GETGLRGDIGSPGR^{669}		9-2	4	1060.5	4238	1.60
K736-R669	S1	48.61	10081	^{713}GEVGPAGPnGFAGPAGAAGQpGAKGER^{739}-^{662}GEIGnPGRDGAR^{673}	24	6-7	4	904.93	3615.7	1.94
		49.16	10196	^{713}GEVGpAGPnGFAGpAGAAGQpGAKGER^{739}-^{662}GEIGNpGRDGAR^{673}		3-3	5	733.55	3662.7	-0.69
		50.91	10567	^{713}GEVGpAGpNGFAGPAGAAGQpGAKGER^{739}-^{662}GEIGnPGRDGAR^{673}		6-3	5	730.35	3646.7	-1.27
		51.21	10630	^{713}GEVGpAGPnGFAGpAGAAGQpGAKGER^{739}-^{662}GEIGNPGRDGAR^{673}		14-8	3	1216.6	3646.7	-1.19
Q123-K1012	A_S1	101.03	20287	^{98}GPpGAAGAPGpqGFQGpAGEPGEPGQTGPAGAR^{130}-^{1007}GEpGDKGPR^{1015}	12	10-4	3	1295.3	3882.8	-2.88
		102.16	20532	^{98}GppGAAGAPGPqGFQGPAGEPGEPGQTGPAGAR^{130}-^{1007}GEpGDKGpR^{1015}		9-3	3	1295.3	3882.8	-2.88
R41-K1318	S1	55.47	11534	^{33}GPSGDRGPRGER^{44}-^{1318}KTNEWQK^{1324}	24	6-3	4	550.03	2196.1	-3.24
		56.06	11658	^{33}GpAGDRGPRGER^{44}-^{1318}KTNEWQK^{1324}		6-4	4	550.03	2196.1	-3.24
R154-R1127	S1	92.13	19333	^{148}PGRpGERGVmGpQGAR^{163}-^{1122}SpASLRPK^{1129}	24	6-3	3	855.44	2563.3	-0.88
		91.15	19124	^{148}pGRPGERGVmGpqGAR^{163}-^{1122}SpASLRPK^{1129}		3-2	3	855.77	2564.3	2.91

Table V-9 List of the cross-links induced by formaldehyde detected in the proteins of collagen 1(III) of different biological breeds (*B. taurus*, *C. subfamily*, *O. cuniculus*, not specific and interactions)

Cross-link	Sample	RT	Scan	Sequence	shift	Matches 1 and 2	Charge	m/z	Mass	Mass error ppm
				Bos taurus						
R242-R1091	S1	127.75	26888	^{239}pGERGFPGPpGmK^{251}-^{1083}GPpGPqGPRGDK^{1094}	24	4-5	3	859.74	2576.2	3.43
	A_S1	86.74	17210	^{239}PGERGFPGPPGMK^{251}-^{1083}GPpGpQGPRGDK^{1094}		4-7	3	848.75	2543.2	-3.60
R308-R683	S1	140.04	29508	^{285}GENGVPGENGAPGPmGPRGAPGERGR^{308}-^{675}GDSGAPGERGppGAGGPPGpR^{695}	24	2-3	5	881.41	4402	-0.42
	A_S1	105.83	21357			2-2	4	1101.5	4402	1.86
R428-K350	A_S1	42.22	7795	^{414}GPpGppGTNGVPGqRGAAGE^{433}-^{347}pGAKGE^{352}	24	5-3	4	605.52	2418.1	-3.17
		42.82	7921			6-3	3	807.04	2418.1	-2.18
K440-R1091	S1	97.84	20546	^{434}PGKnGAKGDpGPR^{446}-^{1083}GPPGPqGPRGDK^{1094}	24	3-8	3	818.74	2453.2	1.55
	A_S1	103.28	20784	^{434}PGKnGAKGDPGPR^{446}-^{1083}GPPGPqGpRGDK^{1094}		4-4	3	818.74	2453.2	1.55
				Caprinae subfamily						
K759-Q1089	S1	119.52	25148	^{745}GEpGTSGVDGApPGKDGPR^{763}-^{1084}GApGPQGPR^{1092}	12	3-4	2	1275.1	2548.2	-3.04
		119.09	25058	^{745}GEpGTSGVDGApGKDGPR^{763}-^{1084}GAPGpQGpR^{1092}		4-4	3	855.74	2564.2	-0.43
R429-R1035	A_S1	104.11	20971	^{415}GpPPGppGTnGAPGqRGAAGE^{434}-^{1033}GDRGE^{1037}	24	7-4	3	784.34	2350	-2.14
		101.85	20465	^{415}GppPGpPGTnGAPGqRGAAGE^{434}-^{1033}GDRGE^{1037}		5-4	4	588.5	2350	-2.31
K1095-Q929	S1	51.87	10770	^{1066}GETGPAGPSGAPGPAGSRGApGPQGpRGDKGETGER^{1101}-^{925}GDAGQpGER^{933}	12	8-3	5	843.79	4213.9	2.86
	A_S1	77.90	15322	^{1084}GAPGpqGPRGDKGE^{1101}-^{925}GDAGQPGE^{932}		6-4	4	520.97	2079.9	3.23
N423-K1032	S1	84.61	17737	^{415}GPpGpPGTNGAPGqR^{429}-^{1026}DGApGTKGDR^{1035}	12	8-2	3	798.37	2392.1	4.34
		157.06	33108	^{415}GpPGPPGTNGAPGqRGAAGEPGK^{437}-^{1026}DGAPGTKGDR^{1035}		2-2	3	1010.2	3027.4	4.28
		79.29	16609	^{415}GPpGPpGTNGAPGQR^{429}-^{1026}DGApGTKGDR^{1035}		18-2	3	798.05	2391.1	-3.96
		60.56	12614	^{415}GPpGPPGTNGAPGqRGAAGEPGK^{437}-^{1014}DGNpGSDGLpGRDGAPGTKGDR^{1035}		6-2	4	1046.5	4181.9	-0.16
	A_S1	99.36	19937	^{415}GPPGPPGTNGApGqRGAAGE^{434}-^{1026}DGAPGTKGDRGE^{1037}		2-2	3	978.12	2931.3	-0.69
K504-N423	S1	173.68	36612	^{493}GPAGANGLPGEKGPPGER^{510}-^{415}GpPGpPGTNGAPGQR^{429}	24	5-4	3	1025.8	3074.5	0.05
	A_S1	95.01	18986	^{493}GpAGANGLpGEKGPPGE509-^{415}GPpGPPGTNGApGQR^{429}		2-5	3	984.47	2950.4	-2.93
K504-N288	S1	118.07	24843	^{493}GPAGAnGLpGEKGPPGER^{510}-^{286}GENGIpGENGApGPmGPR^{303}	12	7-4	3	1148.5	3442.6	-1.76
		118.52	24939	^{493}GPAGANGLPGEKGpPGER^{510}-^{286}GENGIpGEnGApGPmGPR^{303}		3-2	3	1148.5	3442.6	-1.76
		125.26	26364	^{493}GPAGANGLPGEKGPPGER^{510}-^{286}GENGIPGEnGApGpmGpR^{303}		7-3	3	1148.5	3442.6	-0.15
		125.51	26416	^{493}GPAGAnGLpGEKGPPGER^{510}-^{286}GENGIPGEnGApGPMGPR^{303}		5-2	3	1138.2	3411.6	0.36
	A_S1	88.62	17607	^{493}GpAGANGLpGEKGppGER^{510}-^{288}NGIPGENGAPGPMGPR^{303}		3-2	3	1086.2	3255.5	-2.30
K1151-Q428	S1	176.16	37134	^{1141}GPVGpSGPPGKDGTSGHPGpIGPPGpR^{1167}-^{415}GppGPPGTNGAPGQR^{429}	12	3-4	4	970.72	3878.9	4.27
		171.60	36175	^{1141}GPVGpSGPPGKDGTSGHpGPIGPPGPR^{1167}-^{415}GpPGPPGTNGApGQR^{429}		4-3	3	1288.6	3862.9	1.34
H1157-K1032	S1	141.69	29855	^{1152}DGTSGHpGPIGppGPR^{1167}-^{1026}DGApGTKGDR^{1035}	12	7-2	3	849.73	2546.2	-4.43
		142.37.	30000	^{1152}DGTSGHPGpIGpPGpR^{1167}-^{1026}DGAPGTKGDR^{1035}		4-2	3	844.4	2530.2	-2.27
N288-K1107	S1	48.94	10150	^{286}GENGIPGENGAPGPMGPRGAPGER^{309}-^{1102}GAnGIKGHR^{1110}	12	8-3	4	799.64	3194.5	0.31
		169.19	35666	^{286}GENGIpGEnGApGPmGPR^{303}-^{1102}GAnGIKGHR^{1110}		5-3	3	893.08	2676.2	-3.02
K504-N989	S1	90.69	19027	^{493}GPAGAnGLPGEKGPPGER^{510}-^{984}pGpSGqNGER^{993}	12	5-3	3	902.09	2703.2	3.88
		117.67	24760	^{493}GPAGANGLpGEKGPpGER^{510}-^{984}pGPSGqNGER^{993}		10-3	3	907.09	2718.3	3.42
K285-Q999	S1	147.39	31065	^{277}GETGApGLKGENGIPGENGAPGPMGPR^{303}-^{994}GPpGpQGLpGLAGAAGEPGR^{1013}	12	4-5	4	1087.8	4347.1	2.19
	A_S1	177.80	33659	^{277}GETGApGLKGEnGIPGENGAPGPMGPR^{303}-^{994}GPpGPQGLPGLAGAAGEPGR^{1013}		5-3	4	1080	4316.1	4.39
N288-K1032	S1	142.95	30123	^{286}GENGIPGENGAPGPMGPR^{303}-^{1026}DGApGTKGDR^{1035}	12	5-2	3	903.09	2706.2	3.87
		151.15	31859	^{286}GENGIPGENGAPGPMGPR^{303}-^{1026}DGAPGTKGDR^{1035}		2-2	3	897.76	2690.2	1.10
	A_S1	113.40	23020	^{286}GENGIPGENGAPGPMGPR^{303}-^{1026}DGApGTKGDR^{1035}		3-3	3	903.09	2706.2	-1.97
K498-H1109	S1	64.65	13481	^{493}GPAGAnGLPGEKGPPGER^{510}-^{1102}GAnGIKGHR^{1110}	12	6-2	3	861.77	2582.3	2.94
		168.88	35601	^{493}GPAGANGLPGEKGPPGER^{510}-^{1096}GETGERGAnGIKGHR^{1110}		5-2	3	1071.2	3210.6	-0.66
		160.44	33821	^{487}GVPGFRGPAGANGLPGEKGppGERGGpGPAGpR^{519}-^{1102}GAnGIKGHR^{1110}		10-2	3	1336	4005	1.47
		187.64	39557	^{487}GVpGFRGPAGANGLpGEKGPPGERGGPGPAGpR^{519}-^{1102}GAnGIKGHR^{1110}		4-3	4	998.26	3989	-0.20

Q999-K1032	S1	83.67	17539	^{994}GPpGpQGLpGLAGAAGEpR^{1013}-^{1026}DGAPGTKGDR^{1035}	24	2-2	4	704.8	2815.3	3.45
		159.14	33547	^{994}GpPGpQGLPGLAGAAGEPGR^{1013}-^{1026}DGApGTKGDR^{1035}		4-2	3	934.1	2799.4	0.64
R243-K440	S1	96.64	20292	^{241}pGERGLPGPPGmK^{252}-^{438}nGAKGDpGPR^{447}	24	5-13	3	778.38	2332.1	2.44
		100.20	21048	^{241}pGERGLPGPPGmK^{252}-^{438}nGAKGDPGpR^{447}		5-9	4	584.02	2332.1	-0.25
R510-R243	S1	156.31	32948	^{493}GpAGANGLpGEKGPPGERGGPGPAGPR^{519}-PGRPGERGLPGPPGMK^{252}	24	2-2	4	1017	4064.1	-0.94
		97.42	20457	^{505}GPPGERGGPGPAGPRGVAGEPGR^{527}-^{241}PGERGLPGPPGMK^{252}		2-2	4	850.19	3396.7	-3.31
R243-K350	A_S1	90.91	18091	^{241}PGERGLpGppGmK^{252}-^{247}pGAKGE^{352}	24	3-5	3	651.98	1952.9	0.64
		90.98	18106			6-5	4	489.22	1952.9	0.89
	Oryctolagus cuniculus									
K501-Q972	S1	65.87	13741	^{490}GPAGPSGTPGEKGPAGER^{507}-^{967}GSPGPQGVK^{975}	12	21-5	3	820.41	2458.2	-1.22
		65.48	13658	^{490}GPAGPSGTPGEKGPAGER^{507}-^{967}GSPGPQGVK^{975}		15-5	4	615.56	2458.2	3.39
N949-K784	A_S1	67.86	13208	^{976}GESGKpGVNGQNGER^{990}-^{781}PGAKGE^{786}	12	2-3	3	743.03	2226.1	-3.32
		75.65	14842	^{967}GESGKPGVNGQNGER^{990}-^{781}PGAKGE^{786}		2-3	3	738.03	2211.1	-2.68
R560-R444	S1	120.20	25292	^{549}PGppGSQGDSGRpGpPGpSGPR^{570}-^{439}GEPGPRGER^{447}	24	7-3	4	767.86	3067.4	-1.52
		123.39	25968	^{549}pGpPGSQGDSGRPGpPGpSGPR^{570}-^{439}GEpGPRGER^{447}		7-3	3	1023.46	3067.4	0.36
N474-K438	A_S1	74.63	14629	^{471}PGANGLPGAAGER^{483}-^{432}PGKnGAKGEPGPR^{444}	12	7-6	4	611.57	2442.2	1.58
		77.34	15204	^{460}GEDGKDGSPGEPGANGLpGAAGE^{482}-^{432}pGKNGAKGE^{444}		4-4	3	980.45	2938.3	2.70
	Not unique									
K263-R827	S1	150.26	31671	^{250}GPAGmpGFpGmKGHR^{264}-^{823}GERGApGEK^{831}	24	4-3	3	834.05	2499.1	-2.56
		151.61	31956			4-2	4	625.77	2499.1	-2.25
K550-R446	S1	51.24	10637	^{537}GIpGSPGGPGSDGKPGPPGSqGETGR^{562}-^{441}GDPGPRGER^{449}	24	31-3	3	1095.2	3282.5	-1.82
		50.95	10575	^{537}GIpGSPGGPGSDGKPGPPGSqGETGR^{562}-^{441}GDPGPRGER^{449}		17-3	4	821.64	3282.5	-0.61
		40.98	8466	^{537}GIpGSpGGpGSDGKPGPPGSQGETGR^{562}-^{441}GDpGPRGER^{449}		17-2	4	833.39	3329.5	-1.35
N1114-K1031	S1	50.68	10519	^{1110}GFPGNPGApGSpGPAGHqGAVGSpGPAGPR^{1139}-^{1025}DGAPGAKGDR^{1034}	12	42-2	3	1203.6	3607.7	2.13
		50.95	10574			42-4	4	902.93	3607.7	2.39
	Interaction									
K261-R1091	S1	54.09	11242	^{241}GLpGppGIKGpAGMpGFpGMKGHR^{266}-^{1083}GppGpQGPRGDK^{1094}	24	12-10	4	915.4	3657.7	0.11
		211.96	44715	^{241}GLpGppGIKGpAGMpGFpGMKGHRGFDGRNGEK^{273}-^{1083}GppGpQGPRGDK^{1094}		11-8	4	1155.5	4618.1	1.40
K930-R1260	S1	76.29	15967	^{922}GDAGQPGEKGSPGAqGpPGAPGPLGLAGITGAR^{954}-^{1252}KNPARNCR^{1259}	24	8-4	3	1317.7	3950	2.40
		81.68	17118			3-2	4	988.5	3950	3.94
R1022-R240	S1	109.24	22963	^{1011}DGnpGSDGqPGRDGAPGGKGDR^{1032}-^{237}pGERGLPGPPGIK^{249}	24	3-2	4	850.4	3397.6	2.51
		109.75	23072			4-2	3	1133.5	3397.6	0.42
		112.33	23625			2-2	5	680.52	3398.6	-1.68
R570-R507	S1	144.50	30449	^{549}pGPPGSqGETGRPGPPGSPGPRGQPGVMGFPGPK^{582}-^{502}GPAGERGGPGPAGpR^{516}	24	3-6	4	1145.8	4579.2	-0.68
		205.94	43435	^{561}PGpPGSPGPRGqPGVmGFPGPK^{582}-^{502}GPAGERGGpGpAGPR^{516}		2-5	3	1164.6	3490.7	-2.45

Table V-10 List of the cross-links induced by formaldehyde detected in the proteins of collagen 1(I) of different biological breeds (*B. taurus*, *O. cuniculus*, not specific and interactions)

Cross-link	Sample	RT	Scan	Sequence	shift	Matches 1 and 2	Charge	m/z	Mass	Mass error ppm
Bos taurus										
K1032-N1253	S1	128.60	27068	^{1014}EGAPGAEGSpGRDGSpGAKGDR^{1035}-^{1252}KnpARTCR^{1259}	12	2-5	3	1029.8	3086.4	-3.20
		129.18	27189			2-3	4	772.6	3086.4	-2.89
K1151-Q366	S1	153.82	32424	^{1141}GpPGSAGSPGKDGLnGLpGPIGpPGpR^{1167}–^{361}GSEGPQGVR^{369}	12	3-2	4	842.18	3364.7	2.37
	A_S1	127.02	25678	^{1141}GPPGSAGSPGKDGLnGLpGPIGpPGpR^{1167}-^{352}GEAGPQGPRGSEGPQGVR^{369}		10-4	4	1058.8	4231.1	0.90
K351-Q560	S1	236.92	50029	^{322}GNDGATGAAGPPGPTGpAGPPGFPGAVGAKGE GGPQGPR^{360}-^{552}TGpPGPAGQDGR^{563}	12	3-5	4	1130.3	4517.1	-0.79
		241.67	51052	^{322}GNDGATGAAGppGPTGPAGPPGFPGAVGAKGEG GPQGPR^{360}-^{552}TGpPGPAGQDGR^{563}		2-3	5	904.42	4517.1	1.60
Caprinae subfamily										
K983-R1321	A_S1	83.29	16474	^{981}PGKQGPSGASGE^{992}-^{1319}DKRHVWYGE^{1327}	24	16-7	3	762.03	2283.1	3.90
		78.10	15364	^{981}pGKQGPSGASGE^{992}–^{1319}DKRHVWYGE^{1327}		16-9	3	767.36	2299.1	4.08
K429-Q357	S1	45.51	9350	^{415}GpSGpQGPSGPPGPKGNSGEPGAPGSK^{441}-^{352}GEAGpQGpR^{360}	12	8-4	4	825.39	3297.5	-0.92
		47.32	9808	^{415}GpSGPQGPSGPPGpKGNSGEPGApGSK^{441}-^{352}GEAGPQGPR^{360}		4-7	4	821.39	3281.5	0.79
		47.95	9840	^{415}GpSGPqGPSGPPGPKGNSGEpGAPGSK^{441}-^{352}GEAGpQGpR^{360}		10-8	4	825.63	3298.5	0.93
	A_S1	53.37	10203	^{415}GpSGPqGPSGPPGPKGNSGEpGAPGSK^{441}-^{352}GEAGpQGPR^{360}		13-4	4	821.64	3282.5	2.47
Oryctolagus cuniculus										
K524-R1459	S1	48.21	9996	^{511}GLTGSPGSPGpDGKTGPPGPAGQDGR^{536}-^{1459}RqTTqTETppK^{1469}	24	13-2	4	919.94	3675.7	1.52
		50.65	10512	^{511}GLTGSpGSpGPDGKTGpPGpAGqDGR^{536}–^{1459}RQTTQTETPPK^{1469}		12-4	4	923.69	3690.7	-3.33
R225-R1459	S1	128.27	26997	^{210}pGRpGERGPPGPQGAR^{225}-^{1457}NRRQTTQTETPPK1^{469}	24	5-3	3	1066.5	3196.6	-0.88
		131.70	27723	^{210}pGRpGERGPPGPQGAR^{225}-^{1457}NRRQTTQTETPPK1^{469}		4-3	4	800.16	3196.6	-0.31
R1459-R657	A_S1	33.12	5775	^{1457}NRRQTTqTE^{1465}-^{649}GERGFPGE^{656}	24	4-6	3	669.32	2004.9	0.46
		39.22	7148	^{1459}RqTTqTETppK^{1469}–^{649}GERGFpGE^{656}		2-3	3	736.67	2207	0.79
Q429-K566	S1	219.02	46217	^{421}GEpGpTGVQGpPGPAGEEGKR^{441}-^{559}GAAGEPGKAGER^{570}	12	2-2	3	1044.8	3131.5	-3.14
	A_S1	66.01	12819	^{423}PGPTGVqGPPGPAGE^{437}-^{564}PGKAGER^{570}		3-4	3	682.01	2043	-0.91
K405-R212	S1	87.66	18386	^{403}GNSGEpGAPGNKGDTGAK^{420}-^{210}PGRPGER^{216}	24	2-7	3	803.72	2408.1	-2.97
		80.20	16803	^{403}GnSGEPGAPGnKGDTGAK^{420}-^{210}PGRpGER^{216}		3-5	3	804.38	2410.1	0.74
	A_S1	94.11	18789	^{388}GPSGPQGPSGPpGpKGNSGEpGAPGnK^{414}–^{210}pGRPGER^{216}		12-4	3	1076.2	3225.5	3.40
		100.80	20237	^{388}GpSGpQGPSGPPGPKGNSGEpGAPGnK^{414}–^{210}pGRPGER^{216}		4-4	3	1076.2	3225.5	4.19
K402-R282	S1	55.49	11538	^{388}GpSGPQGpSGpPGPKGNSGEPGAPGnK^{414}-^{277}GLPGERGR^{284}	24	13-4	5	659.92	3294.6	-1.80
		57.75	12018	^{388}GpSGpQGPSGPPGPKGNSGEPGAPGnK^{414}–^{277}GLpGERGR^{284}		11-5	4	824.65	3294.6	-4.35
	A_S1	91.75	18274	^{388}GPSGpqGpSGPPGPKGnSGEpGAPGNKGDTGAK^{420}-^{277}GLPGERGR^{284}		4-3	3	1275.9	3824.8	-1.18
Not unique										
K285-Q249	S1	79.99	16759	^{250}GDAGpAGPKGEPGSpGEnGApGqMGPR^{276}–^{217}GPPGPQGAR^{225}	24	14-2	4	839.4	3353.5	-4.00
	A_S1	84.46	16719	^{250}GDAGpAGpKGE^{260}-^{217}GpPGpqGAR^{225}		2-4	3	627.26	1878.8	2.85
		89.91	17879	^{250}GDAGpAGpKGE^{260}–^{217}GPPGpqGAR^{225}		2-3	3	621.93	1862.8	2.01
K519-Q249	S1	51.81	10757	^{478}GPAGERGApGpAGPKGSPGEAGR^{527}–^{217}GPPGPQGAR^{225}	12	8-6	3	970.48	2908.4	-4.17
	A_S1	58.57	11259	^{478}GpAGERGAPGPAGpKGSpGE^{524}-^{217}GPpGPQGAR^{225}		6-3	3	886.43	2656.3	-3.64
Q660-K593	S1	51.70	10733	^{657}PGEQGVPGDLGAPGpSGAR^{675}-^{586}GAAGEPGKAGER^{597}	12	13-9	4	712.1	2844.4	4.33
	A_S1	50.72	9592	^{660}QGVpGDLGApGPSGAR^{675}–^{591}pGKAGE^{596}		12-4	3	685	2052	0.04
		53.74	10237	^{660}QGVpGDLGApGPSGAR^{675}–^{591}PGKAGE^{596}		10-4	3	679.67	2036	-0.74
K519-R1092	S1	36.93	7609	^{505}GPAGERGAPGPAGPKGSPGEAGR^{527}-^{1084}GPAGPqGPRGDK^{1095}	24	6-4	4	798.4	3189.6	-1.94
	A_S1	47.36	8878	^{511}GApGPAGpKGSpGE^{524}–^{1084}GPAGpqGPRGDK^{1095}		7-10	3	801.71	2402.1	-3.23
K519-R909	S1	93.27	19578	^{511}GApGPAGPKGSPGEAGR^{527}-^{907}GPRGETGPAGR^{917}	24	7-8	3	852.76	2555.3	-0.96
	A_S1	107.03	21627	^{511}GAPGpAGpKGSpGEAGR^{527}-^{904}GSKGpRGE^{911}		4-4	3	779.71	2336.1	-3.17
R303-R239	S1	52.63	10930	^{286}GEpGSPGENGAPGQMGPRGLPGER^{309}-^{237}PGRPGER^{243}	24	3-4	4	778.62	3110.5	-1.12
	A_S1	103.20	20766	^{288}PGSpGENGAPGQmGpRGLPGE^{308}-^{237}PGRPGE^{242}		5-2	3	882.41	2644.2	-4.08
	S1	52.17	10834		12	13-6	4	693.11	2768.4	0.27

K741-Q687		52.20	10840	^{733}GAAGLpGpKGDRGDAGPK^{750}-^{685}GVQGPpGPAGPR^{696}		12-12	5	554.69	2768.4	-0.61
K285-R239	S1	40.81	8430	^{277}GDAGPAGpKGEpGSpGEnGAPGqMGPR^{303}-^{237}PGRPGER^{243}	24	2-2	4	822.37	3285.5	-0.11
		46.72	9681	^{277}GDAGpAGPKGEPGSPGEnGAPGQMGpR^{303}-^{237}PGRPGER^{243}		9-5	4	818.13	3268.5	1.42
R303-K1252	S1	86.68	18177	^{286}GEpGSPGENGAPGqMGPRGLPGER^{309}-^{1252}KnPAR^{1256}	24	4-2	3	977.47	2929.4	3.46
		87.26	18299			2-2	4	733.35	2929.4	3.46
R510-R1092	S1	47.51	9849	^{493}GFPGADGVAGPKGPAGERGAPGpAGPK^{519}-^{1084}GpAGpqGPRGDK^{1095}	12	5-3	5	716.96	3579.7	-0.43
		91.09	19112	^{505}GPAGERGApGPAGpK^{519}-^{1084}GpAGpQGPRGDK^{1095}		5-6	3	848.08	2541.2	1.49
Q249-K906	A_S1	106.49	21503	^{244}GPpGPQGAR^{252}-^{904}GSKGPR^{909}	24	9-4	3	975.5	1475.8	-2.31
		116.43	23673	^{244}GPPGPQGAR^{252}-^{904}GSKGPR^{909}		8-3	3	487.6	1459.8	-2.47
N294-K593	A_S1	76.67	15063	^{294}NGAPGqMGPR^{303}-^{591}PGKAGE^{596}	12	6-6	2	777.87	1553.7	3.77
		107.72	21779	^{288}PGSpGENGAPGQMGPRGLPGER^{309}-^{591}PGKAGER^{596}		5-2	3	953.8	2858.4	1.39
		47.17	8838	^{294}NGAPGqMGPRGLPGE^{308}-^{591}pGKAGE^{595}		9-4	3	675.32	2022.9	4.18
		62.55	12092	^{288}PGSPGENGApGQMGpR^{303}-^{591}PGKAGER^{596}		8-4	3	756.03	2265.1	-0.41
Q420-K519	A_S1	92.82	18512	^{415}GPSGPqGPSGppGpK^{429}-^{511}GAPGpAGPKGSPGE^{524}	12	6-4	3	857.74	2570.2	-0.53
		93.94	18754	^{415}GPSGPQGPSGppGpK^{429}-^{511}GAPGpAGpKGSPGE^{524}		7-5	3	862.74	2585.2	-4.14
K825-R909	A_S1	77.23	15181	^{804}pGpPGPAGFAGppGADGQpGAKGE^{827}-^{907}GPRGE^{911}	24	13-2	3	902.08	2703.2	-0.19
		87.15	17297	^{804}pGppGPAGFAGPPGADGqPGAKGE^{827}-^{907}GPRGETGpAGR^{917}		6-4	4	807.88	3227.5	2.59
		69.14	13476	^{804}pGPpGpAGFAGpPGADGQpGAKGE^{827}-^{907}GPRGE^{911}		4-2	4	676.8	2703.2	3.94
K1095-R1092	S1	82.29	17247	^{1084}GPAGPqGpRGDKGETGEQGDR^{1104} - ^{1084}GpAGPQGPRGDK^{1095}	24	8	3	1086.8	3257.5	1.48
		83.40	17483	^{1084}GPAGPqGPRGDKGETGEQGDR^{1104}-^{1084}GpAGPQGPRGDK^{1095}		12-9	4	811.37	3241.5	2.35
		88.09	18477			13-9	5	649.3	3241.5	1.80
	A_S1	105.01	21173	^{1084}GpAGPqGpRGDKGE^{1097}-^{1084}GPAGpqGPRGDKGE^{1094}		6-4	3	906.75	2717.2	2.58
R801-R239	S1	83.86	17579	^{796}GAPGDRGEPGPPGpAGFAGPPGADGqPGAKGEP GDAGAK^{834}-^{237}PGRpGER^{243}	24	7-5	4	1062.3	4245	4.48
	A_S1	18.35	2498	^{796}GApGDRGE^{803}-^{237}PGRpGER^{242}		8-6	3	452.54	1354.6	0.24
K227-R239	S1	31.89	6540	^{220}GPpGPpGKnGDDGEAGK236-^{237}PGRpGER^{243}	24	14-7	4	595.28	2377.1	-0.61
		31.91	6543	^{220}GpPGPpGKnGDDGEAGK^{236}-^{237}PGRpGER^{243}		18-12	3	793.37	2377.1	-0.72
		32.76	6724	^{220}GPpGPpGKnGDDGEAGK^{236}-^{237}PGRPGER^{243}		15-8	4	591.28	2361.1	-0.06
		32.82	6736	^{220}GPpGPpGKnGDDGEAGK^{236} – ^{237}PGRPGER^{243}		22-8	3	788.04	2361.1	-0.62
		33.92	6970	^{220}GpPGPpGKnGDDGEAGK^{236}-^{237}PGRpGER^{243}		6-6	3	793.37	2377.1	3.39
		90.69	19027	^{220}GPpGpPGKnGDDGEAGKpGR^{236}-^{237}PGRpGER^{243}		5-2	3	902.09	2703.3	-1.17
	A_S1	25.50	4083	^{220}GPpGPpGKnGDDGEAGK^{233}-^{237}PGRPGE^{242}		9-5	4	552.26	2205	0.54
K429-N1253	S1	85.90	18010	^{415}GPSGPqGPSGPPGPKGnSGEPGAPGSK^{441}-^{1253}NPARTCR^{1259}	24	13-4	3	1085.5	3253.5	-2.72
		47.57	9860	^{415}GpSGPQGPSGPPGpKGNSGEPGAPGSKGDTGAK^{441}-^{1253}NPARTCRDLK^{1262}		6-2	5	834.8	4169	-0.13
R454-K714	S1	56.18	11684	^{442}GARGEPGpTGLpGPPGER^{459}-^{706}GAAGLpGpKGDR^{717}	24	7-4	5	577.69	2883.4	-1.37
		60.39	12579	^{442}GARGEPGPTGLPGPPGER^{459}-^{706}GAAGLpGpKGDR^{717}		8-5	5	571.29	2851.4	-2.01
K1385-R243	S1	110.01	23129	^{1385}KALLLqGSnEIEIRAEGNSR^{1404}-^{240}PGERGPPGPQGAR^{252}	24	3-9	3	1166.9	3497.8	0.07
		110.59	23256			2-8	4	875.45	3497.8	-1.49
		110.75	23290			2-10	5	700.56	3497.8	-1.52
K741-Q1089	S1	64.71	13493	^{733}GAAGLPGPKGDRGDAGpK^{750}-^{1084}GPAGpQGpR^{1092}	24	6-11	4	629.82	2515.3	0.76
		66.48	13870			4-7	3	839.43	2515.3	1.50
		73.55	15383	^{733}GAAGLPGPKGDRGDAGpK^{750}-^{1084}GPAGPQGpR^{1092}		10-8	4	625.82	2499.3	1.25
H266-K227	S1	152.62	32171	^{240}pGERGpPGPqGARGLpGTAGLPGMKGHR^{267}-^{220}GPPGPPGKnGDDGEAGK^{236}	12	8-3	3	1439.4	4315.1	3.97
		182.46	38465	^{253}GLpGTAGLpGMKGHRGFSGLDGAK^{276}-^{220}GPPGPPGKnGDDGEAGKPGR^{239}		6-3	4	1047	4184	0.69
K429-Q357	S1	45.15	9350	^{415}GpSGpQGPSGPPGPKGNSGEPGAPGSK^{441}-^{352}GEAGpQGpR^{360}	12	8-4	4	825.39	3297.5	-0.92
		47.32	9808	^{415}GpSGPQGPSGPPGpKGNSGEPGApGSK^{441}-^{352}GEAGPQGPR^{360}		4-7	4	821.39	3281.5	0.79
		47.47	9840	^{415}GpSGPqGPSGPPGPKGNSGEpGAPGSK^{441}-^{352}GEAGpQGpR^{360}		10-8	4	825.63	3298.5	0.93
	A_S1	53.57	10203	^{415}GpSGPqGPSGPPGPKGNSGEpGAPGSK^{441}-^{352}GEAGpQGPR^{360}		13-4	4	821.64	3282.5	2.47
K429-R1256	S1	208.17	43911	^{388}GpSGPqGpSGPPGPKGNSGEPGApGSK^{441}-^{1224}KnPARTCR^{1231}	24	3-4	3	1144.2	3429.6	1.95
		208.29	43937	^{388}GPSGpqGpSGPPGPKGNSGEpGAPGSK^{441}-^{1224}KnPARTCR^{1231}		3-4	3	1144.2	3429.6	1.95
	S1	144.32	30411		24	2-3	4	1088	4348.1	-2.85

K447-R252		145.66	30696	^{442}GDTGAKGEPGPTGIQGpPGPAGEEGK^{467}-^{244}GPpGPqGARGLPGTAGLPGmK^{264}		2-3	5	870.62	4348.1	-3.12
		149.74	31561	^{442}GDTGAKGEPGPTGIQGPPGpAGEEGK^{467}-^{244}GPpGPqGARGLPGTAGLPGmK^{264}		2-3	4	1088	4348.1	0.56
R1083-K227	S1	172.88	36445	^{1062}SGDRGETGPAGpAGPIGPVGARGpAGPQGPR^{1092}-^{220}GppGPPGKnGDDGEAGK^{236}	24	2-3	4	1108.5	4430.1	3.00
		79.98	16756	^{1066}GETGPAGPAGpIGPVGARGPAGPqGPR^{1092}-^{220}GPpGPPGKnGDDGEAGK^{236}		8-4	3	1329	3983.9	0.52
K611-Q618	A_S1	75.74	14861	^{597}RGVpGpPGAVGPAGKDGE^{614}-^{615}AGAQGPPGPAGPAGE^{629}	24	5-4	3	969.5	2905.4	-3.39
		76.75	15080	^{597}RGVpGPPGAVGPAGKDGE^{614}-^{615}AGAqGPpGPAGPAGE^{629}		5-11	4	727.6	2906.4	-0.09
		79.37	15632	^{597}RGVpGPPGAVGPAGKDGE^{614}-^{615}AGAQGPPGPAGPAGE^{629}		6-19	3	964.1	2889.4	-2.83
Q794-K1080	A_S1	74.35	14570	^{775}GEPGPPGpAGFAGppGADGQpGAKGE^{800}-^{1078}GIKGHR^{1083}	12	7-2	3	1005.5	3013.4	3.49
		89.16	17719	^{775}GEPGPPGpAGFAGPpGADGQPGAKGE^{800}-^{1078}GIKGHR^{1083}		5-2	3	994.82	2981.4	3.78
K1068-R882	A_S1	30.57	5201	^{1057}GPAGPQGPRGDKGE^{1070_880}GpRGE^{884}	24	11-7	3	626.3	1875.9	-0.28
		33.11	5772	^{1057}GPAGPqGPRGDKGE^{1070_880}GpRGE^{884}		10-5	3	626.63	1876.9	0.09
		33.24	5803			10-6	5	376.38	1876.9	0.12
K227-R801	A_S1	94.45	18863	^{220}GPpGpPGKNGDDGE^{233_796}GAPGDRGE^{803}	24	11-7	3	702.97	2105.9	3.59
		91.28	18173	^{220}GPPGPPGKnGDDGE^{233_796}GAPGDRGE^{803}		5-3	3	692.63	2074.9	-1.40
K227-R882	A_S1	21.21	3116	^{220}GPpGPpGKNGDDGEAGK^{236_877}GGKGPRGE^{884}	24	14-2	4	591.28	2361.1	-0.57
		92.68	18479	^{220}GPpGPPGKnGDDGE^{233_877}GGKGPRGE^{884}		10-3	2	1046	2089.9	-0.62
		85.85	17025	^{220}GPpGPPGKnGDDGE^{233_877}GGKGPRGE^{884}		6-3	3	697.65	2089.9	1.16
K983-R123	A_S1	51.98	9868	^{981}pGKQGpSGASGE^{992}-^{121}GpRGPAGppGR^{131}	24	3-5	3	731.68	2192	1.16
		53.27	10140	^{981}pGKQGPSGASGE^{992}-^{121}GpRGpAGppGR^{131}		2-3	4	549.03	2192	-0.77
R486-R917	A_S1	65.03	12614	^{474}pGPAGLpGpPGERGGpGSR^{492_912}TGPAGRPGE^{920}	24	3-5	3	881.1	2640.3	-4.43
		67.58	13149	^{474}PGpAGLPGPPGERGGpGSR^{492_912}TGPAGRpGE^{920}		4-12	3	875.76	2624.3	-4.48
R630-R369	A_S1	101.58	20405	^{630}RGEQGpAGSPGFqGLPGPAGPpGE^{653}-^{364}GpqGVRGE^{371}	24	5-4	4	773.1	3088.4	2.29
		103.35	20802			6-4	3	1030.5	3088.4	4.40
R311-K741	S1	46.03	9535	^{304}GLPGERGRpGApGPAGAR^{321}-^{733}GAAGLpGpKGDR^{744}	24	9-16	4	714.62	2854.5	-3.10
		46.20	9572			10-11	3	952.49	2854.5	-3.89
Interactions										
K429-R123	S1	200.06	42191	^{415}GPSGPQGpSGpPGPKGnSGEPGAPGnK^{441}-^{121}GpRGPAGPPGR^{131}	24	2-5	3	1158.6	3472.6	-1.94
		212.13	44750	^{415}GPSGpQGPSGpPGPKGnSGEPGAPGnK^{441}-121GPRGPAGpPGR^{131}		2	4	865.15	3472.6	-3.10
		217.94	45997	^{415}GPSGpQGPSGPPGPKGnSGEPGAPGnK441-^{121}GPRGpAGPPGR^{131}		2-2	3	1153.2	3456.6	-3.41

Reference

1. BRANDI, C., *Teoria del restauro, Ed.* Einaudi, Torino. Lorusso et al Il restauro architettonico: le diverse concezioni nel corso dei secoli, 1977.
2. Stoner, J.H., *Changing approaches in art conservation: 1925 to the present.* Scientific Examination of Art: Modern Techniques in Conservation and Analysis, 2005: p. 40-57.
3. Orsini, S., A. Yadav, M. Dilillo, L.A. McDonnell ,I. Bonaduce, *Characterization of degraded proteins in paintings using bottom-up proteomic approaches: new strategies for protein digestion and analysis of data.* Analytical chemistry, 2018. **90**(11): p. 6403-6408.
4. Ryder, M.L., *Parchment—its history, manufacture and composition.* Journal of the Society of Archivists, 1964. **2**(9): p. 391-399.
5. Reed, R., *Ancient skins, parchments and leathers*. 1972.
6. Rabin, I., *Building a Bridge from the Dead Sea Scrolls to Mediaeval Hebrew Manuscripts.* 2014.
7. Sobel, H. ,H. Ajie, *Modification in amino acids of dead sea scroll parchments.* Free Radical Biology and Medicine, 1992. **13**(6): p. 701-702.
8. Parry, D.W. ,S.D. Ricks, *Current Research and Technological Developments on the Dead Sea Scrolls: Conference on the Texts from the Judean Desert, Jerusalem, 30 April, 1995*. Vol. 20. 1996: Brill.
9. van der Werf, I.D., C.D. Calvano, G. Germinario, T.R. Cataldi ,L. Sabbatini, *Chemical characterization of medieval illuminated parchment scrolls.* Microchemical Journal, 2017. **134**: p. 146-153.
10. Reed, R., *The nature and making of parchment.* 1975.
11. Roberts, M.T. ,D. Etherington, *Bookbinding and the Conservation of Books. A Dictionary of Descriptive Terminology*. 1982: ERIC.
12. Levey, M., *Chemistry and chemical technology in ancient Mesopotamia.* 1959.
13. Harlan, J. ,S. Feairheller, *Chemistry of the crosslinking of collagen during tanning*, in *Protein Crosslinking*. 1977, Springer. p. 425-440.
14. Wandrey, I., *Jewish Manuscript Cultures: New Perspectives*. Vol. 13. 2017: Walter de Gruyter GmbH & Co KG.
15. Forbes, R.J., *Studies in ancient technology. 5*. Vol. 2. 1965: Brill Archive.
16. Haines, B., *The manufacture of parchment*, in *Conservation of leather and related materials*. 2006, Routledge. p. 220-221.
17. Lampe, G.W.H., *The Cambridge History of the Bible: Volume 2, The West from the Fathers to the Reformation*. Vol. 2. 1975: Cambridge University Press.
18. Fleischmajer, R., E.D. MacDonald, J.S. Perlish, R.E. Burgeson ,L.W. Fisher, *Dermal collagen fibrils are hybrids of type I and type III collagen molecules.* Journal of structural biology, 1990. **105**(1-3): p. 162-169.
19. Kennedy, C.J. ,T.J. Wess, *The structure of collagen within parchment–a review.* Restaurator, 2003. **24**(2): p. 61-80.
20. Brodsky, B. ,J.A. Ramshaw, *The collagen triple-helix structure.* Matrix Biology, 1997. **15**(8-9): p. 545-554.
21. Haines, B.M., *Parchment: The physical and chemical characteristics of parchment and the materials used in its conservation*. 1999.
22. Heidemann, E., *Fundamentals of leather manufacture*. 1993: Roether.
23. Bozec, L., G. van der Heijden ,M. Horton, *Collagen fibrils: nanoscale ropes.* Biophysical journal, 2007. **92**(1): p. 70-75.
24. Sommer, D.V. ,R. Larsen, *Detection of COL III in parchment by amino acid analysis.* Amino acids, 2016. **48**(1): p. 169-181.

25. Nijhuis, W., D. Eastwood, J. Allgrove, I. Hvid, H. Weinans, R. Bank ,R. Sakkers, *Current concepts in osteogenesis imperfecta: bone structure, biomechanics and medical management.* Journal of children's orthopaedics, 2019. **13**(1): p. 1-11.
26. Su, H.-N., L.-Y. Ran, Z.-H. Chen, Q.-L. Qin, M. Shi, X.-Y. Song, X.-L. Chen, Y.-Z. Zhang ,B.-B. Xie, *The ultrastructure of type I collagen at nanoscale: large or small D-spacing distribution?* Nanoscale, 2014. **6**(14): p. 8134-8139.
27. Bicchieri, M., M. Monti, G. Piantanida, A. Sodo ,M.T. Tanasi, *Inside the parchment.* Proceedings of Art, 2008: p. 25-30.
28. Maxwell, C.A., T.J. Wess ,C.J. Kennedy, *X-ray diffraction study into the effects of liming on the structure of collagen.* Biomacromolecules, 2006. **7**(8): p. 2321-2326.
29. Axelsson, K.M., R. Larsen, D.V. Sommer ,R. Melin, *Degradation of collagen in parchment under the influence of heat-induced oxidation: preliminary study of changes at macroscopic, microscopic, and molecular levels.* Studies in conservation, 2016. **61**(1): p. 46-57.
30. Možir, A., I.K. Cigić, M. Marinšek ,M. Strlič, *Material properties of historic parchment: A reference collection survey.* Studies in conservation, 2014. **59**(3): p. 136-149.
31. Alvarez, A.M.F., J. Bouhy, M. Dieu, C. Charles ,O. Deparis, *Animal species identification in parchments by light.* Scientific reports, 2019. **9**(1): p. 1-10.
32. Gustavson, K.H., *Chemistry and reactivity of collagen.* 1956.
33. Fiddyment, S., M.D. Teasdale, J. Vnouček, É. Lévêque, A. Binois ,M.J. Collins, *So you want to do biocodicology? A field guide to the biological analysis of parchment.* Heritage Science, 2019. **7**(1): p. 35.
34. Cains, A. *A facing method for leather, paper and membrane*. in *The Institute of paper conservation: conference papers Manchester 1992*. 1992.
35. Ryder, M., *Follicle arrangement in skin from wild sheep, primitive domestic sheep and in parchment.* Nature, 1958. **182**(4638): p. 781-783.
36. Ryder, M.L., *Follicle remains in some British parchments.* Nature, 1960. **187**(4732): p. 130-132.
37. Clarkson, C., *Rediscovering parchment: The nature of the beast.* The Paper Conservator, 1992. **16**(1): p. 5-26.
38. Aceto, M., A. Idone, A. Agostino, G. Fenoglio, M. Gulmini, P. Baraldi ,F. Crivello, *Non-invasive investigation on a VI century purple codex from Brescia, Italy.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2014. **117**: p. 34-41.
39. Aceto, M., E. Calà, A. Agostino, G. Fenoglio, M. Gulmini, A. Idone, C. Porter, C. Hofmann, S. Rabitsch ,C. Denoël, *Mythic dyes or mythic colour? New insight into the use of purple dyes on codices.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019. **215**: p. 133-141.
40. Grazia, C., D. Buti, A. Amat, F. Rosi, A. Romani, D. Domenici, A. Sgamellotti ,C. Miliani, *Shades of blue: non-invasive spectroscopic investigations of Maya blue pigments. From laboratory mock-ups to Mesoamerican codices.* Heritage Science, 2020. **8**(1): p. 1-20.
41. Ghigo, T., O. Bonnerot, P. Buzi, M. Krutzsch, O. Hahn ,I. Rabin, *An Attempt at a Systematic study of inks from Coptic manuscripts.* 2018.
42. Bicchieri, M., M. Monti, G. Piantanida, F. Pinzari ,A. Sodo, *Non-destructive spectroscopic characterization of parchment documents.* Vibrational Spectroscopy, 2011. **55**(2): p. 267-272.
43. Miguel, C., C. Barrocas-Dias, T. Ferreira ,A. Candeias, *The comparative study of four Portuguese sixteenth-century illuminated Manueline Charters based on spectroscopy and chemometrics analysis.* Applied Physics A, 2017. **123**(1): p. 72.

44. Orna, M.V., P.L. Lang, J.E. Katon, T.F. Mathews ,R.S. Nelson, *Applications of infrared microspectroscopy to art historical questions about medieval manuscripts*. 1989, ACS Publications.
45. Calà, E., A. Agostino, G. Fenoglio, V. Capra, F. Porticelli, F. Manzari, S. Fiddyment ,M. Aceto, *The Messale Rosselli: Scientific investigation on an outstanding 14th century illuminated manuscript from Avignon.* Journal of Archaeological Science: Reports, 2019. **23**: p. 721-730.
46. Badea, E., L. Miu, P. Budrugeac, M. Giurginca, A. Mašić, N. Badea ,G. Della Gatta, *Study of deterioration of historical parchments by various thermal analysis techniques complemented by SEM, FTIR, UV-Vis-NIR and unilateral NMR investigations.* Journal of Thermal Analysis and Calorimetry, 2008. **91**(1): p. 17-27.
47. Della Gatta, G., E. Badea, A. Mašić ,R. Ceccarelli, *Structural and thermal stability of collagen within parchment: a mesoscopic and molecular approach.* Improved Damage Assessment of Parchment (IDAP) Collection and Sharing of Knowledge (Research Report No 18). EU-Directorate-General for Research, 2007: p. 89-98.
48. Sendrea, C., C. Carsote, E. Badea, A. Adams, M. Niculescu ,H. Iovu, *Non-invasive characterisation of collagen-based materials by NMR-MOUSE and ATR-FTIR.* UPB Sci Bull, 2016. **78**(3): p. 27-38.
49. Masic, A., M.R. Chierotti, R. Gobetto, G. Martra, I. Rabin ,S. Coluccia, *Solid-state and unilateral NMR study of deterioration of a Dead Sea Scroll fragment.* Analytical and bioanalytical chemistry, 2012. **402**(4): p. 1551-1557.
50. Fiddyment, S., B. Holsinger, C. Ruzzier, A. Devine, A. Binois, U. Albarella, R. Fischer, E. Nichols, A. Curtis ,E. Cheese, *Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting.* Proceedings of the National Academy of Sciences, 2015. **112**(49): p. 15066-15071.
51. Vilde, V., M.L. Abel ,J.F. Watts, *A surface investigation of parchments using ToF-SIMS and PCA.* Surface and Interface Analysis, 2016. **48**(7): p. 393-397.
52. Teasdale, M.D., S. Fiddyment, J. Vnouček, V. Mattiangeli, C. Speller, A. Binois, M. Carver, C. Dand, T.P. Newfield ,C.C. Webb, *The York Gospels: a 1000-year biological palimpsest.* Royal Society open science, 2017. **4**(10): p. 170988.
53. Piñar, G., H. Tafer, M. Schreiner, H. Miklas ,K. Sterflinger, *Decoding the biological information contained in two ancient Slavonic parchment codices: an added historical value.* Environmental Microbiology, 2020.
54. Bower, M.A., M.G. Campana, C. Checkley-Scott, B. Knight ,C.J. Howe, *The potential for extraction and exploitation of DNA from parchment: a review of the opportunities and hurdles.* Journal of the Institute of Conservation, 2010. **33**(1): p. 1-11.
55. Cassidy, L.M., M.D. Teasdale, S. Carolan, R. Enright, R. Werner, D.G. Bradley, E.K. Finlay ,V. Mattiangeli, *Capturing goats: documenting two hundred years of mitochondrial DNA diversity among goat populations from Britain and Ireland.* Biology letters, 2017. **13**(3): p. 20160876.
56. Librado, P., C. Gamba, C. Gaunitz, C. Der Sarkissian, M. Pruvost, A. Albrechtsen, A. Fages, N. Khan, M. Schubert ,V. Jagannathan, *Ancient genomic changes associated with domestication of the horse.* Science, 2017. **356**(6336): p. 442-445.
57. Woodward, S.R., G. Kahila, P. Smith, C. Greenblatt, J. Zias ,M. Broshi, *Analysis of parchment fragments from the Judean Desert using DNA techniques*, in *Current research and technological developments on the Dead Sea Scrolls*. 1996, Brill. p. 215-238.
58. Shepherd, L.D., P. Whitehead ,A. Whitehead, *Genetic analysis identifies the missing parchment of New Zealand's founding document, the Treaty of Waitangi.* PloS one, 2019. **14**(1): p. e0210528.

59. Poulakakis, N., A. Tselikas, I. Bitsakis, M. Mylonas ,P. Lymberakis, *Ancient DNA and the genetic signature of ancient Greek manuscripts.* Journal of archaeological science, 2007. **34**(5): p. 675-680.
60. Pangallo, D., K. Chovanova ,A. Makova, *Identification of animal skin of historical parchments by polymerase chain reaction (PCR)-based methods.* Journal of archaeological science, 2010. **37**(6): p. 1202-1206.
61. Campana, M.G., M.A. Bower, M.J. Bailey, F. Stock, T.C. O'Connell, C.J. Edwards, C. Checkley-Scott, B. Knight, M. Spencer ,C.J. Howe, *A flock of sheep, goats and cattle: ancient DNA analysis reveals complexities of historical parchment manufacture.* Journal of archaeological science, 2010. **37**(6): p. 1317-1325.
62. Gilbert, M.T.P., H.-J. Bandelt, M. Hofreiter ,I. Barnes, *The nine criteria for authenticity.* Trends in Ecology & Evolution, 2005. **10**(20): p. 541-544.
63. Hofreiter, M., D. Serre, H.N. Poinar, M. Kuch ,S. Pääbo, *ancient DNA.* Nature Reviews Genetics, 2001. **2**(5): p. 353-359.
64. Brandt, L.Ø., K. Haase ,M.J. Collins, *Species identification using ZooMS, with reference to the exploitation of animal resources in the medieval town of Odense.* Danish Journal of Archaeology, 2018. **7**(2): p. 139-153.
65. Teasdale, M.D., N.L. van Doorn, S. Fiddyment, C.C. Webb, T. O'Connor, M. Hofreiter, M.J. Collins ,D.G. Bradley, *Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2015. **370**(1660): p. 20130379.
66. Bi, K., T. Linderoth, D. Vanderpool, J.M. Good, R. Nielsen ,C. Moritz, *Unlocking the vault: next-generation museum population genomics.* Molecular ecology, 2013. **22**(24): p. 6018-6032.
67. Cicero, C., F. Pinzari ,F. Mercuri, *18th Century knowledge on microbial attacks on parchment: analytical and historical evidence.* International biodeterioration & biodegradation, 2018. **134**: p. 76-82.
68. Migliore, L., M.C. Thaller, G. Vendittozzi, A.Y. Mejia, F. Mercuri, S. Orlanducci ,A. Rubechini, *Purple spot damage dynamics investigated by an integrated approach on a 1244 AD parchment roll from the Secret Vatican Archive.* Scientific reports, 2017. **7**(1): p. 1-12.
69. Beata, G., *The use of-omics tools for assessing biodeterioration of cultural heritage: A review.* Journal of Cultural Heritage, 2020.
70. Buckley, M., M. Collins, J. Thomas-Oates ,J.C. Wilson, *Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry, 2009. **23**(23): p. 3843-3854.
71. Kirby, D.P., K. Phillips ,N. Khandekar, *Implementation of MALDI mass fingerprinting in a museum laboratory for the identification of proteins in works of art*. 2016.
72. Kirby, D.P., M. Buckley, E. Promise, S.A. Trauger ,T.R. Holdcraft, *Identification of collagen-based materials in cultural heritage.* Analyst, 2013. **138**(17): p. 4849-4858.
73. Hickinbotham, S., S. Fiddyment, T.L. Stinson ,M.J. Collins, *How to Get Your Goat: Automated Identification of Species from MALDI-ToF Spectra.* Bioinformatics, 2020.
74. Toniolo, L., A. D'Amato, R. Saccenti, D. Gulotta ,P.G. Righetti, *The Silk Road, Marco Polo, a bible and its proteome: A detective story.* Journal of proteomics, 2012. **75**(11): p. 3365-3373.
75. D'Amato, A., G. Zilberstein, S. Zilberstein, M.I. Golovan, A.A. Zhuravleva ,P.G. Righetti, *Anton Chekhov and Robert Koch cheek to cheek: a proteomic study.* Proteomics, 2018. **18**(9): p. 1700447.

76. Zilberstein, G., S. Zilberstein, U. Maor ,P.G. Righetti, *Surface analysis of ancient parchments via the EVA film: The Aleppo Codex.* Analytical biochemistry, 2020: p. 113824.
77. Zilberstein, G., R. Zilberstein, S. Zilberstein, G. Fau, A. D'Amato ,P.G. Righetti, *Il n'y a pas d'amour heureux pour Casanova: Chemical-and bio-analysis of his Memoirs.* Electrophoresis, 2019. **40**(23-24): p. 3050-3056.
78. Fourneau, M., C. Canon, D. Van Vlaender, M.J. Collins, S. Fiddyment, Y. Poumay ,O. Deparis, *Histological study of sheep skin transformation during the recreation of historical parchment manufacture.* Heritage Science, 2020. **8**(1): p. 1-7.
79. Pinzari, F., *Microbial processes involved in the deterioration of paper and parchment.* Biodeterioration and preservation in art, archaeology and architecture. London: Publications Ltd, 2018: p. 33-56.
80. Kennedy, C.J. ,T.J. Wess, *The use of X-ray scattering to analyse parchment structure and degradation*, in *Physical techniques in the study of art, archaeology and cultural heritage*. 2006, Elsevier. p. 151-172.
81. Vadrucci, M., C. Cicero, P. Parisse, L. Casalis ,G. De Bellis, *Surface evaluation of the effect of X-rays irradiation on parchment artefacts through AFM and SEM.* Applied Surface Science, 2020: p. 145881.
82. Gonzalez, L.G. ,T.J. Wess, *The effects of hydration on the collagen and gelatine phases within parchment artefacts.* Heritage Science, 2013. **1**(1): p. 14.
83. Ehrle, F., *Della conservazione e del restauro dei manoscritti antichi.* Rivista delle Biblioteche e degli Archivi, ix, 1898. **5**.
84. Bicchieri, M. ,F. Pinzari, *Discoveries and oddities in library materials.* Microchemical Journal, 2016. **124**: p. 568-577.
85. Bicchieri, M., P. Biocca, P. Colaizzi ,F. Pinzari, *Microscopic observations of paper and parchment: the archaeology of small objects.* Heritage Science, 2019. **7**(1): p. 47.
86. Kite, M., *Collagen products: glues, gelatine, gut membrane and sausage casings*, in *Conservation of leather and related materials*. 2006, Routledge. p. 214-219.
87. Stijnman, A., *Iron-gall ink and ink corrosion.* Archivum Lithuanicum, 2002(04): p. 171-178.
88. Furia, P., *Storia del restauro librario dalle origini ai nostri giorni*. 1992.
89. Cutter, C., *The restoration of paper documents and manuscripts.* College & research libraries, 1967. **28**(6): p. 387-397.
90. Casanova, E., *Archivistica*. 1928: Stab. arti grafiche Lazzeri.
91. Bicchieri, M., *Hard Science and History*. 2018.
92. Vaisey, D., *EWB Nicholson and the St. Gall conference, 1898.* The Bodleian Library Record, 1974. **9**(2): p. 101-13.
93. Zervos, S. ,I. Alexopoulou, *Paper conservation methods: a literature review.* Cellulose, 2015. **22**(5): p. 2859-2897.
94. Fakirov, S., Z. Sarac, T. Anbar, B. Boz, I. Bahar, M. Evstatiev, A. Apostolov, J. Mark ,A. Kloczkowski, *Mechanical properties and transition temperatures of crosslinked-oriented gelatin.* Colloid and Polymer Science, 1997. **275**(4): p. 307-314.
95. Usta, M., D. Piech, R. MacCrone ,W. Hillig, *Behavior and properties of neat and filled gelatins.* Biomaterials, 2003. **24**(1): p. 165-172.
96. De Carvalho, R. ,C. Grosso, *Characterization of gelatin based films modified with transglutaminase, glyoxal and formaldehyde.* Food hydrocolloids, 2004. **18**(5): p. 717-726.
97. Albert, K., E. Bayer, A. Wörsching ,H. Vögele, *Investigation of the hardening reaction of gelatin with 13C labeled formaldehyde by solution and solid state 13C NMR spectroscopy.* Zeitschrift für Naturforschung B, 1991. **46**(3): p. 385-390.

98. Usha, R. ,T. Ramasami, *Effect of crosslinking agents (basic chromium sulfate and formaldehyde) on the thermal and thermomechanical stability of rat tail tendon collagen fibre.* Thermochimica Acta, 2000. **356**(1-2): p. 59-66.
99. Currey, J.D., K. Brear, P. Zioupos ,G.C. Reilly, *Effect of formaldehyde fixation on some mechanical properties of bovine bone.* Biomaterials, 1995. **16**(16): p. 1267-1271.
100. Bigi, A., G. Cojazzi, S. Panzavolta, K. Rubini ,N. Roveri, *Mechanical and thermal properties of gelatin films at different degrees of glutaraldehyde crosslinking.* Biomaterials, 2001. **22**(8): p. 763-768.
101. Tengroth, C., U. Gasslander, F.O. Andersson ,S.P. Jacobsson, *Cross-linking of gelatin capsules with formaldehyde and other aldehydes: an FTIR spectroscopy study.* Pharmaceutical development and technology, 2005. **10**(3): p. 405-412.
102. Salsa, T., M. Pina ,J. Teixeira-Dias, *Crosslinking of gelatin in the reaction with formaldehyde: An FT-IR spectroscopic study.* Applied spectroscopy, 1996. **50**(10): p. 1314-1318.
103. Ofner, C.M., 3rd, Y.E. Zhang, V.C. Jobeck ,B.J. Bowman, *Crosslinking studies in gelatin capsules treated with formaldehyde and in capsules exposed to elevated temperature and humidity.* J Pharm Sci, 2001. **90**(1): p. 79-88.
104. Klockenbusch, C. ,J. Kast, *Optimization of formaldehyde cross-linking for protein interaction analysis of non-tagged integrin ? 1.* BioMed Research International, 2010. **2010**.
105. Kamps, J.J.A.G., R.J. Hopkinson, C.J. Schofield ,T.D.W. Claridge, *How formaldehyde reacts with amino acids.* Communications Chemistry, 2019. **2**(1).
106. Metz, B., G.F. Kersten, P. Hoogerhout, H.F. Brugghe, H.A. Timmermans, A. De Jong, H. Meiring, J. ten Hove, W.E. Hennink ,D.J. Crommelin, *Identification of formaldehyde-induced modifications in proteins reactions with model peptides.* Journal of Biological Chemistry, 2004. **279**(8): p. 6235-6243.
107. Toews, J., J.C. Rogalski, T.J. Clark ,J. Kast, *Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions.* Anal Chim Acta, 2008. **618**(2): p. 168-83.
108. Metz, B., G.F. Kersten, G.J. Baart, A. de Jong, H. Meiring, J. ten Hove, M.J. van Steenbergen, W.E. Hennink, D.J. Crommelin ,W. Jiskoot, *Identification of formaldehyde-induced modifications in proteins: reactions with insulin.* Bioconjugate chemistry, 2006. **17**(3): p. 815-822.
109. Fraenkel-Conrat, H., M. Cooper ,H.S. Olcott, *The reaction of formaldehyde with proteins.* Journal of the American Chemical Society, 1945. **67**(6): p. 950-954.
110. Sutherland, B.W., J. Toews ,J. Kast, *Utility of formaldehyde cross-linking and mass spectrometry in the study of protein–protein interactions.* Journal of mass spectrometry, 2008. **43**(6): p. 699-715.
111. Hoffman, E.A., B.L. Frey, L.M. Smith ,D.T. Auble, *Formaldehyde crosslinking: a tool for the study of chromatin complexes.* J Biol Chem, 2015. **290**(44): p. 26404-11.
112. Gustafsson, O.J., G. Arentz ,P. Hoffmann, *Proteomic developments in the analysis of formalin-fixed tissue.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2015. **1854**(6): p. 559-580.
113. Magdeldin, S. ,T. Yamamoto, *Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues.* Proteomics, 2012. **12**(7): p. 1045-1058.
114. Tayri-Wilk, T., M. Slavin, J. Zamel, A. Blass, S. Cohen, A. Motzik, X. Sun, D.E. Shalev, O. Ram ,N. Kalisman, *Mass spectrometry reveals the chemistry of formaldehyde cross-linking in structured proteins.* Nature communications, 2020. **11**(1): p. 1-9.
115. Fox, C.H., F.B. Johnson, J. Whiting ,P.P. Roller, *Formaldehyde fixation.* Journal of Histochemistry & Cytochemistry, 1985. **33**(8): p. 845-853.

116. Hoffman, E.A., H. Zaidi, S.J. Shetty, S. Bekiranov ,D.T. Auble, *An Improved Method for Measuring Chromatin-binding Dynamics Using Time-dependent Formaldehyde Crosslinking.* Bio Protoc, 2018. **8**(4).
117. Schmiedeberg, L., P. Skene, A. Deaton ,A. Bird, *A temporal threshold for formaldehyde crosslinking and fixation.* PLoS One, 2009. **4**(2): p. e4636.
118. Sinz, A., *Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here?* Angew Chem Int Ed Engl, 2018. **57**(22): p. 6390-6396.
119. Solomon, M.J. ,A. Varshavsky, *Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures.* Proc Natl Acad Sci U S A, 1985. **82**(19): p. 6470-4.
120. Strutt, H. ,R. Paro, *Mapping DNA target sites of chromatin proteins in vivo by formaldehyde crosslinking.* Methods Mol Biol, 1999. **119**: p. 455-67.
121. Toews, J., J.C. Rogalski, T.J. Clark ,J. Kast, *Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions.* Analytica chimica acta, 2008. **618**(2): p. 168-183.
122. Vasilescu, J., X. Guo ,J. Kast, *Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry.* Proteomics, 2004. **4**(12): p. 3845-3854.
123. Zhang, Y., M. Muller, B. Xu, Y. Yoshida, O. Horlacher, F. Nikitin, S. Garessus, S. Magdeldin, N. Kinoshita ,H. Fujinaka, *Unrestricted modification search reveals lysine methylation as major modification induced by tissue formalin fixation and paraffin embedding.* Proteomics, 2015. **15**(15): p. 2568-2579.
124. Nirmalan, N.J., P. Harnden, P.J. Selby ,R.E. Banks, *Development and validation of a novel protein extraction methodology for quantitation of protein expression in formalin-fixed paraffin-embedded tissues using western blotting.* The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 2009. **217**(4): p. 497-506.
125. Coscia, F., S. Doll, J.M. Bech, L. Schweizer, A. Mund, E. Lengyel, J. Lindebjerg, G.I. Madsen, J.M. Moreira ,M. Mann, *A streamlined mass spectrometry–based proteomics workflow for large-scale FFPE tissue analysis.* The Journal of pathology, 2020. **251**(1): p. 100-112.
126. Dapic, I., L. Baljeu-Neuman, N. Uwugiaren, J. Kers, D.R. Goodlett ,G.L. Corthals, *Proteome analysis of tissues by mass spectrometry.* Mass Spectrometry Reviews, 2019. **38**(4-5): p. 403-441.
127. Hood, B.L., M.M. Darfler, T.G. Guiel, B. Furusato, D.A. Lucas, B.R. Ringeisen, I.A. Sesterhenn, T.P. Conrads, T.D. Veenstra ,D.B. Krizman, *Proteomic analysis of formalin-fixed prostate cancer tissue.* Molecular & cellular proteomics, 2005. **4**(11): p. 1741-1753.
128. Blackler, A.R., *10: Proteomic Analysis of Archival Samples.* Pathological Specimens And Genomic Medicine: Emerging Issues, 2019: p. 201.
129. Metz, B., T. Michiels, J. Uittenbogaard, M. Danial, W. Tilstra, H.D. Meiring, W.E. Hennink, D.J. Crommelin, G.F. Kersten ,W. Jiskoot, *Identification of formaldehyde-induced modifications in diphtheria toxin.* Journal of Pharmaceutical Sciences, 2020. **109**(1): p. 543-557.
130. Kim, T.H. ,J. Dekker, *Formaldehyde Cross-Linking.* Cold Spring Harb Protoc, 2018. **2018**(4).
131. Wang, Z.-J., J.-B. Yang, G.-P. Li, N.-N. Sun, W.-C. Sun, Q.-S. Peng ,N. Liu, *Chemical Modifications of Peptides and Proteins with Low Concentration Formaldehyde Studied by Mass Spectrometry.* Chinese Journal of Analytical Chemistry, 2016. **44**(8): p. 1193-1199.
132. Depuydt, L., *Coptic and Coptic literature.* A companion to Ancient Egypt, 2010: p. 732-754.

133. Thompson, S.E., *Ancient Egypt: Facts and Fictions*. 2019: ABC-CLIO.
134. Depuydt, L. ,D.A. Loggie, *Catalogue of Coptic Manuscripts in the Pierpont Morgan Library*. Vol. 2. 1993: Peeters Louvain.
135. Sayed, G.G.A., G. Gabra ,T. Vivian, *Coptic Monasteries: Egypt's Monastic Art and Architecture*. 2002: Oxford University Press.
136. Library, P.M. ,H. HYVERNAT, *A Check List of Coptie Manuscripts in the Pierpont Morgan Library.[By Henri Hyvernat. With Facsimiles.]*. 1919: Privately printed.
137. SOUTHWORTH, G. ,F. TRUJILLO, *The Coptic Bindings Collection at the Morgan Library & Museum: History, Conservation, and Access.*
138. Needham, P., *Twelve centuries of bookbindings, 400-1600*. 1979.
139. Han, X., L. He, L. Xin, B. Shan ,B. Ma, *PeaksPTM: mass spectrometry-based identification of peptides with unspecified modifications.* Journal of proteome research, 2011. **10**(7): p. 2930-2936.
140. Ramsøe, A., V. van Heekeren, P. Ponce, R. Fischer, I. Barnes, C. Speller ,M.J. Collins, *DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes.* Journal of Archaeological Science, 2020. **115**: p. 105080.
141. Bettinger, J.Q., K.A. Welle, J.R. Hryhorenko ,S. Ghaemmaghami, *Quantitative analysis of in vivo methionine oxidation of the human proteome.* Journal of proteome research, 2019. **19**(2): p. 624-633.
142. Schilling, B., R.H. Row, B.W. Gibson, X. Guo ,M.M. Young, *MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides.* Journal of the American Society for Mass Spectrometry, 2003. **14**(8): p. 834-850.

Conclusion and perspective

This doctoral thesis aimed to reach a more in-depth comprehension of proteinaceous compounds in artworks by implementing various strategies based on mass spectrometry proteomics, some of them not yet adopted in cultural heritage studies. The different methodologies proposed were intended to identify proteins from samples at trace levels and disclose new molecular information not yet been fully explored in cultural heritage studies.

Firstly, the research intended to furtherly optimises the already widely implemented bottom-up approach to enhance the range and coverage of detected proteins notwithstanding the heterogeneity and degradation of the artwork and the low amount of sample collected. The developed method was successfully applied to investigate two different artistic materials (secco wall paintings and protein-sized drawings from Thomas Gainsborough). In both studies, a clear identification of proteins was achieved together with the definition of their biological origin and the precise localisation of their PTMs (principally induced by ageing and degradation processes). The strategy was demonstrated particularly useful in the investigation of drawings in which samples were collected with minimally invasive techniques. The achievement of MS and MS/MS spectra with a good signal to noise ratio allowed the detection of additional minor proteins with adequate protein coverage. Hence, the findings have proven that a wide range of proteins of interest can be confidently determined even from samples at trace levels (almost non-invasive sampling).

Secondly, the doctoral research pursued the development of the top-down strategy to investigate proteins collected at trace levels from the cultural heritage objects. In a sample collected with a minimally invasive technique from Gainsborough's drawing, the successful detection of various casein proteoforms, both hydrolysed and highly modified, opened up new horizons in the proteomics studies applied to cultural heritage. So far, this alternative approach was rarely considered for artistic researches due to its need for a high sample amount and its numerous technical difficulties. These findings have proven the potentialities of top-down investigation capable of disclosing information missed during the peptide cleavage (including data on *in situ* proteolysis and the loss of both sequence fragments and PTMs).

The complementarity of the two proteomic approaches was highlighted: top-down still requires further improvements to guarantee an exhaustive investigation of proteins at trace levels when used alone; however, it can bring added values to data formerly investigated with bottom-up methodology.

Furthermore, the research was intended to transcend the classical protein investigation by pursuing a better insight into the structural alterations of proteins in an artwork. Innovative mass spectrometry approaches (never applied before in cultural heritage studies) were conducted, such as cross-linking analysis (on paintings and a manuscript) and Hydrogen/Deuterium exchange (HDX) studies in painting models.

The differences in deuterium uptakes observed for the unpigmented protein versus the fresh painting-models contributed to gain new information on the action of some of the most common inorganic pigments (lead and zinc white, cinnabar and minium) on the molecular and conformational changes of proteins. Various levels of oxidations were observed at the protein level, depending on the pigment used. The decrease of deuterium exchange observed in the pigmented models suggested the modification of the protein conformation and/or indicate a pigment positioning between the amide hydrogen and the solvent. The analyses have proven that pigments can play a critical role in the molecular and structural alteration of proteinaceous binders in the first stages of painting formulation, long before the ageing.

Complementary information on structural modifications was inquired through investigating protein networking and cross-linking via mass spectrometry analysis (bottom-up approach). A data elaboration strategy capable of detecting and localising cross-links in an aged protein-based artwork was developed to improve the results' representability and provide supplementary data on an object's state of conservation. Hence, the detection of reticulated peptides can enhance the findings of standard linear analyses by unveiling the protein fraction potentially undetected due to the reticulation, and it can give a more detailed vision of the object’s environmental conditions (i.e. basic pH, elevated exposure to photo-oxidation or presence of reactive compounds).

The strategy was initially tested on the study of cross-links induced by oxidative stress in painting mock-ups. The method was successfully applied to the study of historic paintings showing its potentialities in the examinations of natural induced reticulations.

Subsequently, the same strategy was shown to effectively disclose chemical-induced reticulations caused by the intentional introduction of a chemical compound (particularly formaldehyde) during an old restoration treatment. The investigation of reticulations, combined with an unbiased modification search, led to a more accurate understanding of the biological composition and conservation history of a Coptic manuscript made with parchment and subjected to an old and invasive restoration. In particular, the observation of fragmentation patterns characteristic of formaldehyde-based cross-links provided the first analytical evidence of a parchment consolidation treatment based on gelatin-formol.

Overall, both HDX experiments and cross-linking studies resulted in more comprehensive identification of the proteinaceous compounds and were remarkably efficient to identify undetected peptides with a classical bottom-up approach due to their chemical and structural modifications.

The development of innovative strategies and the subsequent results have inspired new research ideas intended to further improve the comprehension of the alterations in heterogeneous proteinaceous compounds in historical artworks.

A more in-depth investigation of proteins' structural alterations within a paint system is currently conducted by expanding the cross-linking formation research in relation to inorganic pigments and other organic molecules. In particular, enrichment strategies of the cross-linked fractions are presently investigated. Complementary information regarding the stable formation of metal complexes from different pigments at intact protein level are studied through native protein analysis (already observed using other analytical methods)

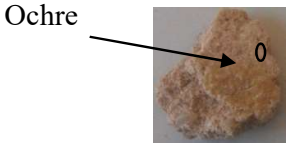
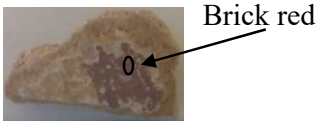
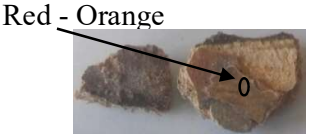
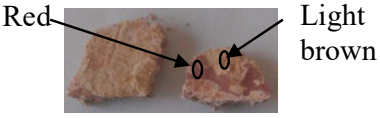
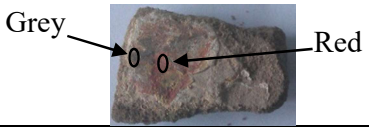
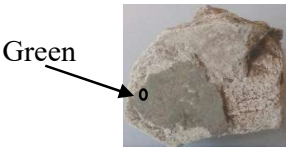
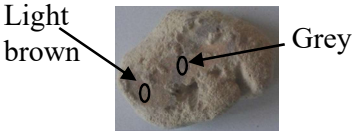
Our next studies also concern the observation of the evolution of the protein structure from the moment of the formulation of the pictorial models and drying until prolonged ageing in various environments. The data would help understand whether the structures established following the pigment introduction tend to remain stable or change over time.

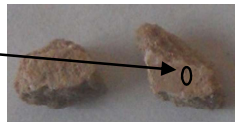
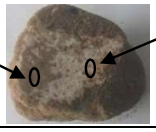
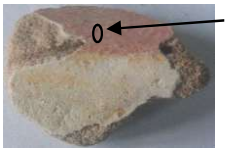
Additional information on the painting system's protein structure can be obtained using complementary techniques such as nuclear magnetic resonance (NMR). The method is hardly performed on precious objects due to its need for high concentrations of analytes; nonetheless, applied to the study of model mock-ups, it could be crucial to define the 3D structures of proteins.

Annexe: samples investigated

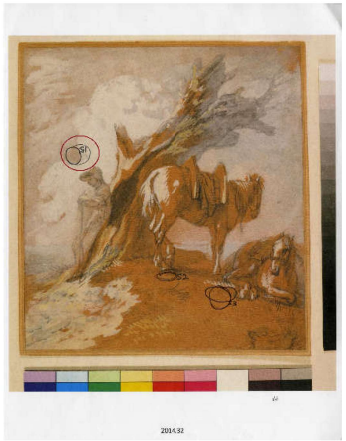
Chapter II Bottom-up proteomic analysis of historical and artistical samples Nubian wall paintings

Annexe i Investigated fragments of Nubian wall painting (Middle Nile Valley) dated between the 6th and 14th century AD from the kingdom of Makuria (Northern Sudan and Southern Egypt). Only one sample is from the Kingdom of Alwa (actual central and southern Sudan). The area of the sampling for the MS analyses and the colours are indicated on each fragment. The fragments were collected from different museum collections (including the Sudan National Museum in Khartoum, National Museum in Warsaw, Humboldt University in Berlin, Polish Centre of Mediterranean Archaeology, University of Warsaw) and supplied by Dr Dobrochna Zielinska (*Institute of Archaeology of the University of Warsaw*).

Sample	Origin	Sampling
Kingdom of Makuria		
Zuma_1	El-Zuma, Tumulus 23 (ca. AD 450-550)	
Se_5_2	Selib church (end of 6th -7th century)	
D_CB.III.12	Old Dongola, Cruciform Building (7th-8th century?)	
Gh_01	i and ii Ghazali Monastery, North Church (second half of the 7th century)	
D_H_A_18	Old Dongola, House A (8th century)	
D_BV_6	i and ii Old Dongola, Church of Archangel Raphael (9th century)	
D_BV_5b	Old Dongola, Church of Archangel Raphael (9th century)	
D_D_1	i and ii Old Dongola, Church on Kom D (9th-10th century)	

D_Th_8	Old Dongola, Throne Hall (9^{th} century)	Light brown 
US. 22_8c	i and ii Us Island, Church SR.22A (10^{th}-11^{th} century)	Brown White 
D_H_SW_13	Old Dongola, Monastery on Kom H, Southwest Annex (second half of the 11^{th} century up to the 13^{th} century)	Light Brown 
B_U_8	Banganarti, Upper Church, Chapel 3 (late 11^{th} to 14^{th} century)	Red 
Kingdom of Alwa		
So_90M14/61-12	Soba, Building 4 (Church) (9^{th}-11^{th} century)	Light grey 

Annexe ii Investigated Gainsborough drawings (1750-1780) with the sampling localisation for milk identification via nano-LC-MS/MS. The minimally invasive technique implemented is defined as well. The samples were provided by Dr Julie Arslanoglu and Dr Federica Pozzi (*Department of Scientific Research, The Metropolitan Museum of Art, New York*) and Reba Snyder (*Thaw Conservation Center, The Morgan Library&Museum, New York).*

Accession number	Title	Date	Sampling location	Sampling label	Microsampling	Picture
2005.82	*Study of Trees*	1750	White wash, background	B	6µm grit	
				S1C2	small eraser	
2014.32	*Open Landscape with Drover and Packhorses*	1775-80	Background (possible white wash)	S1B2	6µm grit	
				S1C2	small eraser	

III, 63	*Landscape with Horse and Cart Descending a Hill*	ca. 1780	Upper right corner; bare spot where G. could have held the paper while dipping in fixative.	S1A	15µm grit	
			Background: middle sky, left of tree (no paint apparent)	S2C2	small eraser	
			Background: sky, right side (no paint apparent)	S3B2	6µm grit	
2017.89	*Coastal Scene with Shipping, Figures, and Cows*	ca. 1780	Background (no apparent paint)	S2B1	6µm grit	
				S2C2	small eraser	

III, 61	*Wooded Landscape with Horseman, Figures, and Bridge*	ca. 1780	Center of pond (no paint evident)	S2B2	6µm grit	
				S2C2	small eraser	
III, 62	*Hilly Landscape with Cows on the Road*	ca.1780	Background (no paint apparent)	S1C2	small eraser	
			Foreground, shaded road	S3B2	6µm grit	
III, 63a	*Wooded Landscape with Cows in a Pool*	1780-85	Front shoulder of centrallight coloured cow	S1C	small eraser	
			Verso (back), lower right corner	S4B	6µm grit	

Chapter IV Protein Structural Analysis in artworks by Mass Spectrometry: Hydrogen/Deuterium Exchange and cross-linking investigation

Annexe iii Egg-based historical artworks investigated to detect and localise oxidising-induced cross-links in proteins. The samples were provided by Dr Julie Arslanoglu, *The Metropolitan Museum of Art* (Department of Scientific Research) in New York. The sample D_BV_6 was supplied by Dr Dobrochna Zielinska (Institute of Archaeology of the University of Warsaw).

Accession number	Title	Painting method	Date and origin	Picture
09.197	"Madonna Adoring the Child with the Infant Saint John the Baptist and an Angel":	Tempera on wood	Early 1490s by Lorenzo di Credi (Lorenzo d'Andrea d'Oderigo). Italy, Florence	
13.175	"Saint Mary Magdalen Holding a Crucifix"; (reverse) "The Flagellation"	Tempera on canvas	Ca. 1395-1400 by Spinello Aretino (Spinello di Luca Spinelli) Italy, Arezzo	
2012.332.2	Gilt-leather wall hanging (part of a set)	Adhesive on "Gilt-leather wall hanging"	1650-1670 by De Gecroonde Son or De Vergulde Roemer Amsterdam (Dutch)	
D_BV_6		mural painting made by secco technique	i and ii Old Dongola, Church of Archangel Raphael (9th century)	

Annexe iv Investigated samples collected from the third page of the Coptic manuscript *"Martyrdom of St. Pteleme"*, which belongs to the Hamuli Collection. The samples were provided by Dr Julie Arslanoglu and Dr Federica Pozzi (*Department of Scientific Research, The Metropolitan Museum of Art, New York*) and conservators Maria Fredericks and Frank Trujillo (*Thaw Conservation Center, The Morgan Library&Museum, New York*).

Martyrdom of St. Pteleme, Egypt, ninth century to early tenth century. MS M.581, folio 3 recto. Morgan Library & Museum; Purchased by Pierpont Morgan in 1911 and 1912. Photograph by Federica Pozzi, Associate Research Scientist, Network Initiative for Conservation Science (NICS), Department of Scientific Research, The Metropolitan Museum of Art.