



Faculty of Sciences

Department of Analytical Chemistry

Tryptic cleavage of proteinaceous paint: a high-performance protein binder analytical technique

Doctoral dissertation to meet the requirements to take the doctoral exam
Doctor of Science: Chemistry

Wim Fremout

Academic year 2013-2014

Supervisor: Prof Dr Luc Moens
Co-supervisor: Prof Dr Peter Vandenabeele
Co-supervisor: Dr Steven Saverwyns
Co-supervisor: Dr Jana Sanyova

Cover: chicken ovalbumin on canvas.

It represents the fusion of the (often considered incompatible) artistic and (bio)analytical worlds; the boundary that was explored in this doctoral dissertation. Ovalbumin is the main protein in egg white, an ingredient of many historical paint recipes and as such a frequently encountered analyte. The photograph of the canvas and the 3D structure of ovalbumin are public domain.

Members of the jury

Prof Dr Frank Vanhaecke

*Ghent University, faculty of Sciences, department of Analytical Chemistry
Chairman*

Prof Dr Luc Moens

*Ghent University, faculty of Sciences, department of Analytical Chemistry
Supervisor*

Prof Dr Peter Vandenabeele

*Ghent University, faculty of Arts and Philosophy, department of archaeology
Co-supervisor*

Dr Steven Saverwyns

*Royal Institute for Cultural Heritage, Laboratory department
Co-supervisor*

Dr Jana Sanyova

*Royal Institute for Cultural Heritage, Laboratory department
Co-supervisor*

Dr Eleni Kouloumpi

National Gallery & Alexandros Soutzos Museum (Athens, Greece), laboratory of Physicochemical Research

Dr Stepanka Kuckova

Institute of Chemical Technology (Prague, Czech Republic), department of Biochemistry and Microbiology

Dr Hilde De Clercq

Royal Institute for Cultural Heritage, Laboratory department

Prof Dr Dieter Deforce

Ghent University, faculty of Pharmaceutical Sciences, department of Pharmaceutics

Prof Dr Frederic Lynen

Ghent University, faculty of Sciences, department of Organic Chemistry

Prof Dr Laszlo Vincze

Ghent University, faculty of Sciences, department of Analytical Chemistry

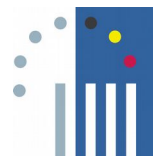
This is a joint research project of the following partners



Ghent University
Faculty of Sciences
Department of Analytical
Chemistry



Royal Institute for Cultural
Heritage
Laboratory department



Belgian Science Policy

TABLE OF CONTENTS

Abbreviations.....	9
I Introduction.....	13
II Proteinaceous binders in painted works of art.....	19
II.1 Artists' paints.....	19
II.1.1 Pigments.....	20
II.1.2 Binding medium.....	20
II.1.3 Additives.....	22
II.2 Introduction to protein theory.....	23
II.2.1 Amino acids.....	23
II.2.2 Protein synthesis.....	27
II.2.3 Important protein properties.....	29
II.2.4 Homologous protein in different species.....	30
II.3 Protein binders in painting.....	30
II.3.1 Animal glue.....	31
II.3.2 Egg based binders.....	33
II.3.3 Milk-based binders.....	36
II.3.4 Other protein sources in paint.....	37
II.4 Chapter epilogue.....	38
III Protein binder analysis.....	39
III.1 General overview of the analytical techniques.....	39
III.1.1 Amino acid analysis (AAA).....	42
III.1.2 Pyrolysis GC-MS.....	45
III.1.3 Immunostaining.....	46
III.1.4 Peptide analysis.....	49
III.2 Sample pretreatment in peptide analysis.....	50
III.2.1 Dissolving and denaturation.....	50
III.2.2 Proteolysis.....	55
III.2.3 Clean-up and preparation for detection method.....	57

III.3 Peptide chromatogram fingerprinting using HPLC-DAD.....	59
III.3.1 Principles.....	59
III.3.2 Experimental set-up.....	60
III.4 Peptide mass fingerprinting using MALDI-TOF-MS.....	60
III.4.1 Principle of MALDI-TOF-MS.....	60
III.4.2 Experimental set-up.....	63
III.5 Peptide identification using HPLC-ESI-QTOF-MS/MS.....	63
III.5.1 Measurement.....	63
III.5.2 Data treatment.....	69
III.6 Chapter epilogue.....	74
 IV Identification of protein binders in works of art by high-performance liquid chromatography diode array detector analysis of their tryptic digests.....	77
IV.1 Introduction.....	78
IV.2 Experimental.....	80
IV.2.1 Reagents.....	80
IV.2.2 Samples.....	80
IV.2.3 Apparatus and chromatographic conditions.....	82
IV.2.4 Analytical procedures.....	83
IV.3 Results and discussion.....	85
IV.3.1 Pure protein samples.....	85
IV.3.2 Paint models and historical paint samples.....	87
IV.4 Conclusions.....	91
 V Classification of protein binders in artists' paints by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry: an evaluation of principal component analysis and soft independent modelling of class analogy.....	93
V.1 Introduction.....	94
V.2 Experimental.....	96
V.2.1 Reagents.....	96
V.2.2 Samples.....	96
V.2.3 Sample pretreatment.....	97
V.2.4 Instrumentation.....	98
V.2.5 PCA and SIMCA.....	98
V.3 Results and discussion.....	98
V.3.1 Determination of the main protein binder class.....	98
V.3.2 Species determination of animal glues.....	107
V.3.3 <i>St Margaret of Antioch</i>	108
V.4 Conclusions.....	110

VI Tryptic peptide analysis of protein binders in works of art by liquid chromatography tandem mass spectrometry.....	113
VI.1 Introduction.....	114
VI.2 Experimental.....	116
VI.2.1 Reagents.....	116
VI.2.2 Samples.....	116
VI.2.3 Instrumentation and chromatographic conditions.....	118
VI.2.4 Analytical procedures.....	119
VI.3 Results and discussion.....	121
VI.3.1 Development of the sample preparation.....	121
VI.3.2 Optimisation of the clean-up and instrumental set-up.....	126
VI.3.3 Determination of the required sample size.....	127
VI.3.4 Proteins in historical samples.....	128
VI.4 Conclusions.....	131
VII A dedicated peptide tandem mass spectral library for conservation science..	133
VII.1 Introduction.....	134
VII.2 Experimental.....	137
VII.2.1 Reagents.....	137
VII.2.2 Samples.....	137
VII.2.3 Analytical procedures.....	139
VII.2.4 Instrumentation and chromatographic conditions.....	139
VII.2.5 Data treatment.....	140
VII.3 Results and discussion.....	141
VII.3.1 Building and test-running the spectral library.....	141
VII.3.2 Application of the libraries on historical samples.....	153
VII.4 Conclusions.....	158
VIII Case study: The double ground layer in Rembrandt's <i>Portrait of Nicolaes van Bambeeck</i> analysed by proteomics techniques.....	161
VIII.1 Introduction.....	161
VIII.2 Overview of the analytical study.....	163
VIII.3 Experimental.....	166
VIII.4 Results and discussion.....	167
VIII.4.1 Proteins in the lower red layer.....	167
VIII.4.2 Proteins in the upper grey layer.....	171
VIII.5 Conclusion.....	178
IX Conclusions and future perspectives.....	179

Bibliography.....	187
Dutch summary (<i>samenvatting</i>).....	203
Acknowledgements (<i>dankwoord</i>).....	211

ABBREVIATIONS

The one- and three-letter codes for the standard amino acids are listed in table II.1, page 23

AAA	Amino acid analysis
ABC	Ammonium bicarbonate
ACN	Acetonitrile
ASL	Annotated spectrum libraries
BLAST	Basic local alignment search tool
CHCA	α -cyano-4-hydroxycinnamic acid
CID	Collision-induced dissociation
COL1A1	Collagen α -1(I) chain
COL1A2	Collagen α -2(I) chain
COL2A1	Collagen α -1(II) chain
COL3A1	Collagen α -1(III) chain
COL4A1	Collagen α -1(IV) chain
DAD	Diode array detector
DHB	2,5-dihydroxybenzoic acid
DNA	Deoxyribonucleic acid
DTT	1,4-dithiotreitol
EDTA	Ethylenediaminetetraacetic acid
EDX	Energy-dispersive X-ray detector
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionisation
FA	Formic acid
FAME	Fatty acid methyl ester
FD	Fluorescence detector
FTICR	Fourier-transform ion cyclotron resonance mass spectrometry
FTIR	Fourier-transform infrared spectroscopy
GC	Gas chromatography
GPM	Global Protein Machine (www.thegpm.org)
HDL	High-density lipoproteins

HILIC	Hydrophilic interaction liquid chromatography
HLB	Hydrophilic-lipophilic balance
HPLC	High-performance liquid chromatography
Hyl	5-hydroxylysine
Hyp	4-hydroxyproline
IAA	2-iodoacetamide
IFM	Immunofluorescence microscopy
ISB	Institute for Systems Biology (Seattle, USA; www.systemsbiology.org)
KIK/IRPA	Royal Institute for Cultural Heritage (Koninklijk Instituut voor het Kunstpatrimonium / Institut royal du Patrimoine artistique, Brussels, Belgium; www.kikirpa.be)
LDL	Low-density lipoproteins
LIF	Laser-induced fluorescence spectroscopy
MALDI	Matrix-assisted laser desorption/ionisation
MaSC	Users' Group for Mass Spectrometry and Chromatography (www.mascgroup.org)
MOWSE	Molecular weight search
mRNA	Messenger ribonucleic acid
MRS	Micro-Raman spectroscopy
MS	Mass spectroscopy
MS/MS	Tandem mass spectrometry
MS ²	Tandem mass spectrometry
MTBSTFA	<i>N-tert</i> -butylmethylsilyl- <i>N</i> -methyltrifluoroacetamide
NCBI	National Center for Biotechnology Information (Bethesda, USA; www.ncbi.nlm.nih.gov)
NIRS	Near-infrared spectroscopy
NIST	National Institute for Standards and Technology (Gaithersburg, USA; www.nist.gov)
PCA	Principal component analysis
PDA	Photodiode array detector
PLM	Polarised light microscopy
PMF	Peptide mass fingerprinting
<i>p</i> NPP	<i>p</i> -nitrophenyl phosphate
PS	Polystyrene
py	Pyrolysis
QTOF	Quadrupole time-of-flight
RNA	Ribonucleic acid
RT	Room temperature
SCX	Strong cation exchange

SDS	Sodium docecyl sulphate
SEM	Scanning electron microscopy
SERS	Surface-enhanced Raman spectroscopy
SIM	Selected ion monitoring
SIMCA	Soft independent modelling by class analogy
SIMS	Secondary ion mass spectrometry
SPE	Solid phase extraction
SR	Synchrotron radiation
TBDMCS	<i>t</i> -butyldimethylchlorosilane
TCEP	Tris(2-carboxyethyl)phospine
TFA	Trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
THM	Thermally assisted hydrolysis and methylation
TMAH	Tetramethylammonium hydroxide
TOF	Time-of-flight
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer ribonucleic acid

I

INTRODUCTION

In 2003 the Royal Institute for Cultural Heritage launched an interdisciplinary research project entitled “Pre-Eyckian panel paintings in the Low Countries”. Questions concerning the binding media arose during this study: did the innovative techniques introduced by Jan van Eyck and the Flemish Primitives come out of the blue? Or were they inspired by an experimenting preceding generation of artists? Answers did not always follow. Museum conservators were not keen on allowing the necessary microsampling on their rare and precious works of art. Sampling, albeit tiny amounts, would inevitably leave visible marks on the already small artworks. Moreover, analytical results were often inconclusive because of several causes: samples being too small, not “layer-pure” (originating from a single layer) or for some other reason sometimes yielded no or unlikely results. In short, current analytical techniques sometimes fell short.

There exists no magical solution. As of today, no non-invasive analytical techniques are reported that are able to discriminate protein binders in paint; sampling thus remains the only possibility. Minimising the sample quantity by developing improved analytical techniques does not solve the difficulty of taking layer pure samples. On the contrary, sampling still happens by hand and often in absence of stereomicroscopic equipment. Larger multilayer samples converted into cross sections, on the other hand, potentially contain a wealth of information on which

(proteinaceous) binder is present in which layer. However, the mainly spectroscopic methods used on cross sections are relatively limited in sensitivity and specificity compared to the mainly chromatographic methods used on layer pure samples. Besides layer purity of the samples, protein mixtures may have other causes: mixtures of different protein binders by the artist, proteins in other ingredients of the paint, contaminations or even fungal or bacterial activity.

Both approaches, cross-sectional and layer-pure sample analysis, are the subject of current research. And since both have considerable advantages and disadvantages, a combination of both will be necessary to answer the complex questions.

This doctoral dissertation does not aim to resolve all the problems related to protein binder analysis, but focuses on one specific goal instead: to develop novel, better performing protein binder analysis techniques inspired by contemporary proteomics¹ whilst decreasing sample consumption. The analysis of peptides, a popular subject in proteomics, offers better specificity and a far improved ability to analyse blends of multiple proteins in comparison the established techniques for protein binders. Proteins are broken down into peptides by enzymatic processes. These peptides are small enough to be accurately measured, but large enough to hold the necessary specificity to unambiguously characterise their parent proteins. Potentially even small mutations between analogous proteins in different species can be distinguished this way. As a result of this high specificity, individual proteins in even the most complex mixtures would be easy to identify next to each other.

This dissertation consists of two introductory and theoretical chapters, followed by four chapters devoted to the development and application of increasingly powerful analysis protocols. Finally, the last chapter treats a case study.

To understand the disadvantages of the current techniques and the potential of proteomics techniques applied on proteinaceous paints, an introduction to the structure and properties of proteins is crucial. Chapter II treats the general structure

¹ The designation *proteome* is a blend of protein and genome, and is used to coin the entire complement of proteins, including the modifications made to a particular set of proteins, produced by an organism or system. Within this context, *proteomics* is the large-scale study of proteins, in particular their structure and function. Often a stricter sense of the term is used for protein purification and mass spectrometry techniques.

of proteins, but also focuses on the most prominent proteins encountered in artists' paints.

An overview of the most common analytical techniques used for protein binder analysis is given in chapter III. Much attention is given to the proteomics-inspired peptide analysis methods developed in this dissertation, compared to those published elsewhere.

The principle of protein digestion by the enzyme trypsin is explored in chapter IV. The resulting complex tryptic peptide mixture is subsequently separated by liquid chromatography and the eluting peptides detected by their UV absorption. This no-frills technique is straight-forward to use and the instrumentation is available in most conservation labs. The peptides, however, cannot be identified by the uncharacteristic UV absorption spectra.

With mass spectrometry, tryptic peptides can be identified by their mass. In chapter V direct measurement of the peptides (without chromatographic separation) is used. This fast and relatively straightforward technique allows us to distinguish the different protein binders more easily. In (single) mass spectrometry, however, a peptide is only characterised by its mass, which hampers the interpretation of unexpected protein sources, complex mixtures and contaminated samples. Only protein sources with a known mass spectrum (in the library) can thus be recognised. This chapter also focuses on interpretation assisted by chemometrics.

Tandem mass spectrometry goes a step further: each peptide is not only characterised by its mass (and charge), but also by its highly specific fragmentation pattern. The inherent slowness of this detection method implies the use of liquid chromatography to deliver the peptides one by one. Chapter VI discusses the set-up of liquid chromatography coupled to mass spectrometry to measure complex paint samples. Identification of the proteins is based on the identification of their tryptic peptides. These are on their turn identified by comparison by a search algorithm that compares the observed tandem mass spectra with the expected fragmentation patterns of known peptide sequences. This extremely powerful approach is aided by the fast-growing libraries of protein sequences available. This

allows the conservation scientist to unambiguously identify proteins in complex mixtures, even from unexpected sources – as long as these proteins are sequenced.

Determination of the species on which a protein binder is based, remains a challenge. While small mutations in the protein sequence between two animals are reflected in the tandem mass spectra of its tryptic peptides, sequence library search algorithms often struggle to determine the species. Firstly, although rapidly improving, sequence libraries do not cover the sequences of all variants of the same protein in different species. The animals and protein interests of conservation scientists apparently differ from those of biochemists. Secondly, the theoretical fragmentation does not completely fit with the experimentally obtained data. It does not help if the target proteins are evolutionarily well conserved and have a repetitive nature, as is the case for collagen (animal glue). In chapter VII an alternative approach is proposed based on a library of tandem mass spectra. This way mutated peptides can be attributed to a species, even if the exact sequence of this peptide and its parent protein are unknown. Both data interpretation approaches combined deliver vastly superior analysis results.

Each of the developed techniques in chapters IV to VII are corroborated with the analysis of samples from real works of art. The focus of these chapters was on the analytical technique. To solve art-historical questions or conservation issues a broader multi-instrumental study is often required. This was the case in Rembrandt's *Portrait of Nicolaes van Bambeeck*, in which remarkable ingredients were found in the double ground layers (chapter VIII). Only the combination of a whole series of current and high-profile techniques enabled to solve this complex jigsaw puzzle. Proteomics-based peptide analysis played an important role in the determination of proteins and flour in the ground layers.

This dissertation discusses the development and application of proteomics inspired peptide analysis methods, starting from straightforward, fast and cheap protocols and building up to extremely powerful state-of-the-art analytical techniques. This dissertation is evidently not the end point in peptide analysis research. Proteins are present in many other domains in cultural heritage: textile fibres, leather, parchment and archaeological remains. The spectral library as is today is only a

proof-of-concept and should be expanded considerably. Peptides with unknown sequence could also be sequenced based on their tandem mass spectra. A large opportunity exists in the understanding of protein degradation phenomena in age-old art objects: even though the proteins seem relatively unaffected by ageing – at least we did not observe age-based degradation, very few systematic studies have been made. Finally, as stated before, to successfully answer the complex issues in conservation science, a strategy combining different analytical techniques is of crucial importance.

II

PROTEINACEOUS BINDERS IN PAINTED WORKS OF ART

This introductory chapter focuses on the theory of both paints and the proteins that can be found in artists' paints. Firstly, the definition and composition of artists' paint will be clarified, as well as its classification based on the binders that are present. Secondly, the category of the protein containing binders, the so called proteinaceous binders or protein binders, will be the further focus of this chapter. Before going deeper into the different proteins that can be encountered in artists' paints at the end of this chapter, some relevant aspects of protein theory will be summarised.

II.1 Artists' paints

Painting as a form of human expression is one of the oldest human inventions, as witnessed by prehistoric cave paintings. During many ages the designation and composition of paint evolved. Modern acrylic paint shares few properties with the mainly mineral based paints found in cave paintings. And although a vast number of components has been – and is – used in paint, the latter can be defined as a blend of pigments, binding media, a solvent as appropriate and a whole range of possible additives.

Both art historians and conservators need to understand the nature of the materials, and as such the exact composition of paint layers. The former for example when researching and describing the painting techniques and stylistic properties of an artist or in a period of time. For restorers, chemical analysis of paints is indispensable for helping to distinguish original layers and later additions, for selecting suitable materials for conservation and restoration (e.g. selective removing overpaintings) and identifying the reasons for degradation of specific paints. The knowledge of the composition of the paints also plays a significant role in authenticating paintings. During the previous decades many analytical techniques have been developed to (try to) answer the diverse questions regarding the different aspects of paint, including pigments, binding media and all kinds of additives(1,2).

II.1.1 Pigments

Pigments constitute the colouring matter of a paint. They are granular solids that are insoluble in the medium. Paints are thus dispersions of one or more different kinds of pigment powders in a medium. Usually pigments are classified by their origin (natural or synthetic) and by their chemical form (organic or inorganic, and further divisions by their chemical structure). Dyes can be used in paints as well; generally they are adsorbed onto a suitable substrate such as alumina or calcite, to form a lake pigment(3). Many present-day paints contain also extenders, which are cheap pigments with low hiding power to provide bulk to the paint; common examples are silica, calcium carbonate and barium sulphate. (1)

Many different analytical techniques have been developed especially for pigment characterisation, allowing to answer a broad range of questions. Nowadays vibrational spectroscopy (infrared, Raman...) and X-ray related techniques (X-ray fluorescence spectroscopy, electron microscopy...) are very commonly used(4–11).

II.1.2 Binding medium

The binding medium (sometimes called the “vehicle”) consists of a filmogen matter and, if necessary, a solvent or thinner. Its task is to accomplish adhesion, both between the pigment particles and the ground to which the paint is applied, as within the paint layer itself. After drying (evaporation of the solvents and thinners)

and/or curing (coalescence and polymerisation), which is dependant on the type of binder, the binding medium forms a solid film around the pigment particles(1,2). The choice for a specific binder has an enormous effect on the aspect and properties of a paint and as such determines the paint class and in many cases also the analytical methods used needed for their identification:

- Oil paints, probably the most used as artists' paints during the entire history, are made using drying oils such as linseed oil, poppy seed oil, walnut oil or in more recent oil paints, safflower, soy bean and tung oil. These oil paints are habitually analysed using gas chromatography mass spectrometry (GC-MS) to study the relative proportions of the different fatty acids and dicarboxylic acids. For GC-MS to analyse these products in paints, they have to be derivatised to render them volatile (e.g. methylation(12,13) into so called fatty acid methyl ester (FAME) or silylation(14)). This reaction is done in advance or, in case of pyrolysis(15–17), on-line.
- Modern binders are mainly based on synthetic polymers. Acrylic and alkyd paints are, for example, as their name indicates made up of acrylics and polyester binding media. Acrylic paints can be both diluted in mineral spirit as well as emulsified in water (often called “latex paint” in case of household paints). Alkyd paints are based on fast curing polyesters, made of polyols and dicarboxylic acids, and most often modified by the addition of oils or fatty acids. FTIR, direct temperature-resolved mass spectrometry (DTMS) and pyrolysis-GC-MS are frequently used to characterise synthetic binding media(18).
- Watercolour and gouache are both made up of aqueous gum media, such as gum arabic (*Acacia spp.*), gum tragacanth (*Astragalus spp.*) or cherry gum (*Prunus spp.*). The analysis of natural gums, principally polysaccharides, most frequently involves high-performance liquid chromatography (HPLC) (19) or GC-MS(20–22). In some cases gums can be identified by the methods normally applied on protein binders because of the low concentration of glycoproteins present.

- Another group of paints is based on proteinaceous binders. In Western European arts, they are mainly found in pre-Renaissance painted objects, while in Eastern Europe they remained extensively used until the late 18th century(23,24). Nowadays protein binders are only occasionally used (e.g. in casein mural paints). A few products that are rich in proteins are known to be used as paint binders, including animal glues, egg fractions and milk products. Amino acid analysis (AAA)(25–32) and pyrolysis(33–36) methods have been applied for the identification of proteins in paint. For AAA, chromatographic separation and detection is done with either GC (using flame ionisation detection (FID)(37) or more commonly MS) or HPLC of the amino acids after hydrolysis of the protein content. In recent years focus shifted more towards the analysis of peptides after proteolysis of the proteins and immunostaining techniques. Both classical and newer approaches will be discussed further-on.

II.1.3 Additives

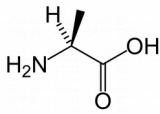
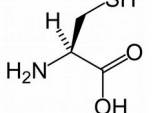
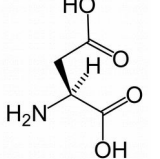
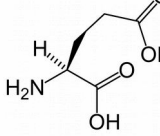
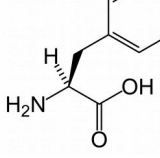
To explain the superiority of the great old masters many hypotheses circulate about some secret or miracle ingredients the artists added to their paints, giving them their superb hue, brilliance or aspect. While some are without any foundation, additives were surely used in some cases to alter and optimise several paint properties. Nowadays all commercially available artists' paints contain multiple additives (typically defined as minor ingredients of maximum 3% of the fresh paint formulation). Thickeners adapt the viscosity of the paints so the paint may be applied properly, have an impact on the film thickness and the flow of the applied paint. Surfactants are added to the paint by maintaining the pigments dispersed, by lowering the surface tension (better spreading of the paint) and by acting as wetting agent. Other additives include driers, defoamers, co-solvents, biocides and UV-stabilisers. (1,2)

II.2 Introduction to protein theory

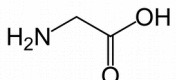
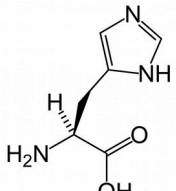
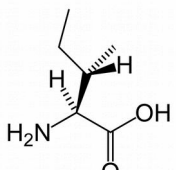
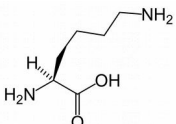
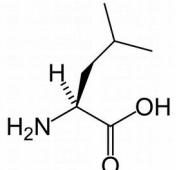
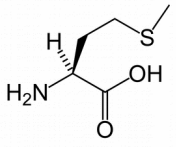
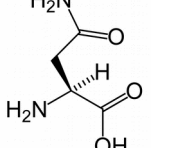
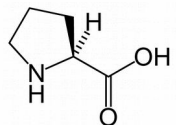
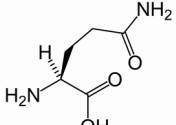
II.2.1 Amino acids

Amino acids are a class of molecules, all containing an amine and a carboxylic acid group. In eukaryotes twenty amino acids (often called standard amino acids, table II.1) make up the basic building blocks for proteins and are directly encoded by the genetic code. All of them are α -amino acids, which means both the carboxylic acid and the amine groups are bound to the same α -carbon. The twenty amino acids are differentiated by a varying side chain, which determines the properties of the amino acid (polarity, charge, size...) and as such largely influences the protein structure and function. (38)

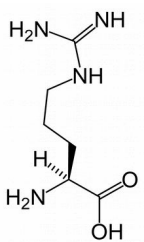
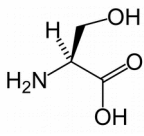
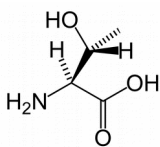
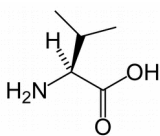
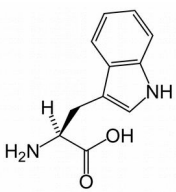
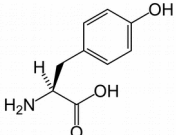
Table II.1. Standard amino acids.

Amino acid	3-letter code	1-letter code	Molecular structure	Average mass (Da)
L-Alanine	Ala	A		89.09404
L-Cysteine	Cys	C		121.15404
L-Aspartic acid	Asp	D		133.10384
L-Glutamic acid	Glu	E		147.13074
L-Phenylalanine	Phe	F		165.19184

(table continued on next page)

Amino acid	3-letter code	1-letter code	Molecular structure	Average mass (Da)
Glycine	Gly	G		75.06714
L-Histidine	His	H		155.15634
L-Isoleucine	Ile	I		131.17464
L-Lysine	Lys	K		146.18934
L-Leucine	Leu	L		131.17464
L-Methionine	Met	M		149.20784
L-Asparagine	Asn	N		132.11904
L-Proline	Pro	P		115.13194
L-Glutamine	Gln	Q		146.14594

(table continued on next page)

Amino acid	3-letter code	1-letter code	Molecular structure	Average mass (Da)
L-Arginine	Arg	R		174.20274
L-Serine	Ser	S		105.09344
L-Threonine	Thr	T		119.12034
L-Valine	Val	V		117.14784
L-Tryptophan	Trp	W		204.22844
L-Tyrosine	Tyr	Y		181.19124

With the exception of glycine, all standard amino acids have their α -carbon bonded to four different groups in a tetrahedral arrangement. The α -carbon is thus a chiral centre and these amino acids have two possible enantiomers. Instead of the systematic *R/S*-nomenclature, as commonly used in organic chemistry, biochemists tend to use the *D/L*-nomenclature, based on the absolute configuration, for describing the enantiomers of amino acids (and sugars), because all amino acids (except glycine) in living organism proteins are L-amino acids. In what follows, all amino acids are considered to occur in their L-state unless specifically stated otherwise. (38)

A few other amino acids exist in proteins occurring in living organisms that are formed after the actual translation by enzymatic *in vivo* post-translational modifications. The most important examples in the context of this dissertation are 4-hydroxyproline (sometimes abbreviated with the three-letter code Hyp, figure II.1) and 5-hydroxylysine (abbreviated as Hyl). The first is found in plant cell wall glycoproteins, and both are components in collagens, the main proteins found in animal glues. In collagen they are produced by hydroxylation of proline and lysine by the enzymes prolyl hydroxylase, respectively lysyl hydroxylase. In the canonical collagen Gly-X-Y triad (where X and Y can be any amino acid), a proline or lysine occupying the Y position is hydroxylated to give a Gly-X-Hyp or Gly-X-Hyl sequence(39,40). Other than in collagen and elastin, hydroxyproline occurs in no mammalian proteins; hydroxyproline is therefore considered a marker for animal glues when found in paint samples using AAA.

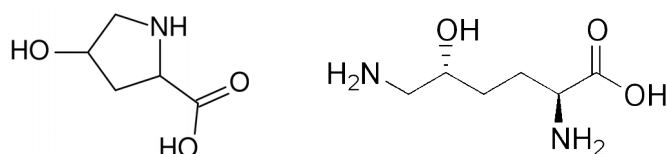


Figure II.1. Left: 4-Hydroxyproline ; right: 5-hydroxylysine.

The α -amine group of an amino acid can react with the α -carboxylic acid of another (condensation reaction), forming an amide group, as shown in figure II.2. This covalent bond is called a peptide bond and the resulting molecule a dipeptide. Further polymerisation can happen, resulting in linear oligo- or polypeptides. The outer end of a peptide with a free α -amino group is called the amino- or N-terminal residu, while the other end, which has a free α -carboxyl group, is the carboxyl- or C-terminal residue. Long amino acid sequences are the primary structure of a protein. (38)

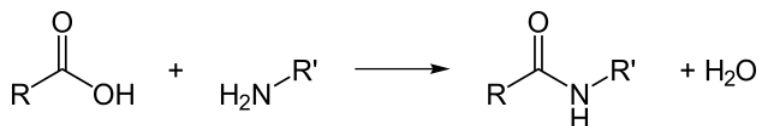


Figure II.2. Peptide bond synthesis

II.2.2 Protein synthesis

Proteins are organic compounds made of amino acids arranged in one or several (so called multi-subunit proteins) linear chains and folded into a globular form. In living organisms, proteins are assembled from amino acids using information encoded in the deoxyribonucleic acid (DNA). A gene is a stretch of DNA that codes for a protein; its unique amino acid sequence is specified by the nucleotide sequence of the gene. DNA contains four different nucleotides, which are composed of a nucleobase (adenine, thymine, guanine and cytosine), deoxyribose and a phosphate group. Each set of three nucleotides – a codon – is translated in one of the standard amino acids. The total number of possible codons is 64, hence some amino acids are specified by more than one codon. (38)

Genes encoded in DNA are first *transcribed* into messenger ribonucleic acid (mRNA) by proteins such as RNA polymerase. This mRNA is then used as a template for protein synthesis by the ribosome; this process is known as *translation* (figure II.3). The mRNA is loaded onto the ribosome and is read three nucleotides at a time by matching each codon to its base pairing anticodon located on a transfer RNA (tRNA) molecule, which also carries the amino acid corresponding to the codon it recognises. In the ribosome, proteins are step by step biosynthesized from N-terminus to C-terminus. (38)

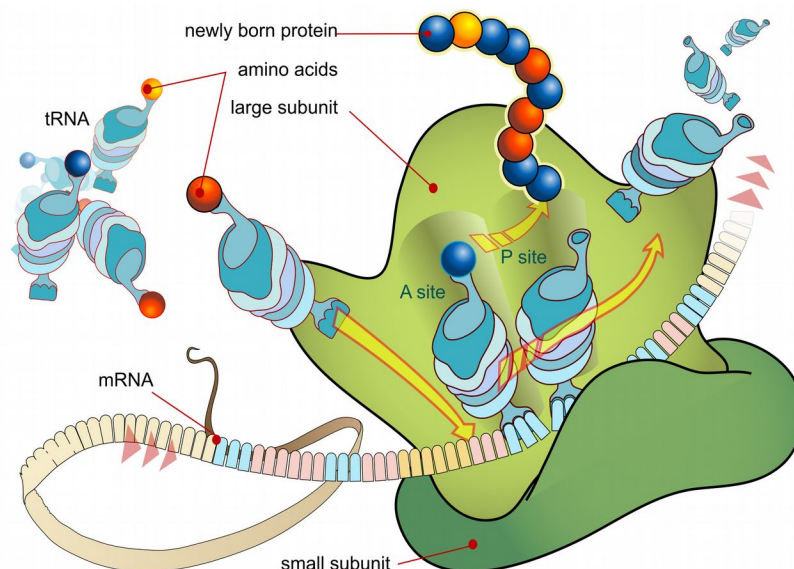


Figure II.3. Translation: the synthesis of proteins in ribosomes.
(source: Ruiz Villarreal(41))

Amino acid chains fold into three-dimensional structures. Many proteins can fold unassisted, simply through the chemical properties of their amino acids, while others require the aid of molecular chaperones to fold into their native states. Conventionally, four levels in a protein's structure are distinguished (figure II.4): (38)

- The *primary structure* is the linear sequence of amino acids.
- The *secondary structure* consists of frequently recurring local structures that are stabilised by hydrogen bonds. The most common examples are the α -helix, β -sheet and turns. Because secondary structures are local, many regions of different secondary structure can be present in the same protein molecule.
- The *tertiary structure* is the overall shape of a single protein molecule or the way the secondary structures fold together. It is stabilised by several non-local interactions, most commonly the formation of a hydrophobic core, but also through salt bridges, hydrogen bonds, disulphide bonds, and even post-translational modifications.
- In case of multi-subunit proteins, the *quaternary structure* is the final structure formed by several protein subunits, which function as a single protein complex.

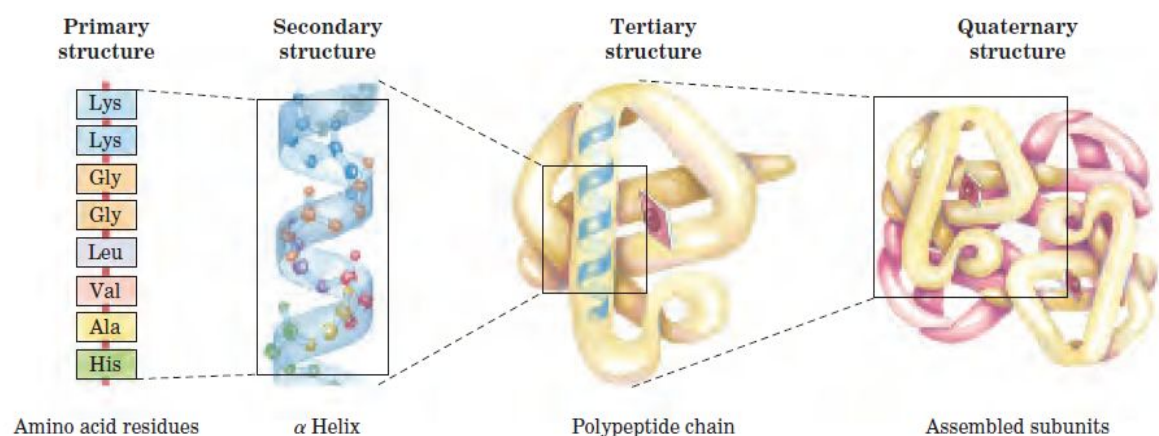


Figure II.4. Four levels of protein structure. (source: Nelson and Cox(38))

Proteins can fold in many different ways; these tertiary or quaternary structures are usually referred to as conformations. The native conformation is the form into which a protein naturally folds. Its form is a critical requisite for the functioning of the protein.

Proteins can be informally divided into three main classes, which correlate with typical tertiary structures: globular proteins, fibrous proteins and membrane proteins. Most all globular proteins are soluble. Fibrous proteins are often structural, such as collagen, the major component of connective tissue, or keratin, the protein component of hair and nails. Membrane proteins often serve as receptors or provide channels for polar or charged molecules to pass through the cell membrane. (38)

II.2.3 Important protein properties

The size of a synthesized protein can be measured by the number of amino acids it contains and by its total molecular mass, which is normally reported in units of Dalton (Da, synonymous with atomic mass units).

In living organisms, with the exception of the non-chiral glycine, only L-isomers of amino acids are used to assemble proteins. After the tissue died, slow racemisation takes place, slowly tending to an L/D ratio of 1. The speed of this process is dependent on temperature, making that dating proteinaceous products is not absolute. (42)

As a result of ageing and dehydration, quaternary, tertiary and even secondary structures are subject to changes over time, forcing the proteins in non-native conformations. As a consequence, the properties of these denaturated proteins change, even though the actual amino acid sequence remains in many cases unaltered. The most notable change is the insolubility of many proteins in aged paint layers. In buried samples the microbial degradation of proteins occurs generally much faster than in the paint layers of paintings. (42)

II.2.4 Homologous protein in different species

The main driving force behind evolution is mutation of the genes. These are generally small changes (point mutations, insertions and deletions) in the nucleotide sequence of the DNA. Differences between species are related to differences in their DNA. If two organisms are closely related, the DNA sequences of their genes are similar; the sequences increasingly diverge as the evolutionary distance between two organisms increases. There are three common types of mutations: point mutations (a nucleotide is substituted for another), deletions (one or more nucleotides are missing) and insertions (one or more nucleotides are inserted).

If the nucleotide sequence in a gene is changed, the translation into amino acids may be influenced in different ways. At best the altered codon encodes the same amino acid (silent mutation) or another amino acid with similar properties, which does not hamper the folding of the resulting protein. If the amino acid that is encoded by the codon has totally different properties, or if the codon encodes for a stop, the protein may be prevented to fold into its active conformation, and thus be non-functional. In the case of insertions or deletions the reading frame of the gene can be shifted, causing all following codons to be different. Given that the loss of functionality of a protein is likely to decrease drastically the chance of survival of the organism, only subtle differences (if any) in amino acid sequences between closely related species are noted.

Consequently, proteins can enable conservation scientists to determine the species origin of the protein-containing component of paint. In case of AAA the differences in terms of relative amino acid composition are rarely significant. Modern techniques in use in proteomics, however, are capable of identifying these mutations.

II.3 Protein binders in painting

The most evident sources of proteins in art objects are the proteinaceous binders and adhesives used both in painting and for other purposes in art, such as animal glues, products based on egg (egg white and / or yolk) and on milk, of which a

large proportion consists of protein matter. But proteins play a capital role in every living organism and as such, many natural products made from animal or plant sources contain to some extent proteins. Proteins can be expected as a major or minor component in many other natural products used in paint, in the support (canvas, parchment...), in conservation treatment products or they can be introduced by contamination. Most of the publications on protein binders are limited to animal glues, eggs and milk based binders, and so these will be the main focus in this dissertation, but the methods that are developed herein are expected to be able to handle and identify the proteins from unexpected or minor sources.

II.3.1 Animal glue

Animal glue is made from diverse connective tissues (bone, cartilage, skin, fish air bladder...). After degreasing, washing, and the removal of hair and keratins, the collagen-rich tissue is subjected to prolonged boiling. As a result, the collagens hydrolyse and form a strong adhesive that found many applications throughout the history of art. In painting it was very frequently used as sizing medium to seal canvas, paper, wood or stone before painting, but also as a paint binder itself: so called *distemper* (ambiguously). Traditionally rabbit skin glue was used for the latter, due to its higher elasticity compared to other hide glue. Because of the hygroscopic nature of collagen, paints based upon animal glue have a relatively high risk of cracking(25). Animal glues also found application in e.g. woodworking (typically hide glue), book binding, as a varnish coating, for consolidation of flaking paint and relining canvases(25). Mild treatment of collagenous material results in a dark coloured animal glue that becomes very hard after drying, while very long boiling treatment results in gelatin, a lighter coloured and softer material. The difference in colour and properties is caused by dark humins and impurities in the glues. (12,38)

A collagen molecule (tropocollagen, figure II.5) is made up of three polypeptide strands (α -chains), all in a left-handed helix conformation, which differs from the more common α -helix. Together, the three α -chains form, a quaternary structure in the form of a right-handed coiled coil, which is stabilised by numerous hydrogen bonds between the chains. In turn, these tropocollagens aggregate further into fibrous right-handed coils and larger fibrillar structures. (38)

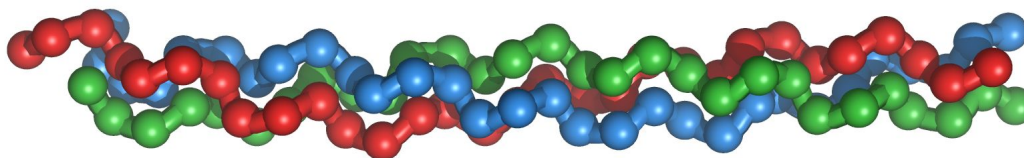


Figure II.5. Tropocollagen: triple helix. (source: adapted from Vossman(43))

There are many types of collagen in living organisms, all with their own specialised functions, but types I to IV account for about 90% of all collagens, and of these, only types I to III are found in the connective tissues used for making animal glues. Collagen type I is the most abundant collagen and is present in connective tissues, including skin, bone and cartilage. It consists of two α -1(I) collagen chains and one α -2(I) chain. Collagen type II is the main component of cartilage and consists of three equal α -1(II) collagen chains. Collagen type III is found alongside type I in most connective tissues (skin, bone and cartilage) and made of three equal α -1(III) chains. (38)

The collagen α -chains are characterised by repeating patterns in their amino acid sequences: it is mainly build-up of Gly-X-Y triads where X and Y are any amino acid. X is often proline and Y hydroxyproline (post-translational modification, see section II.2.1, page 23). Hydroxyproline is only present in very few other animal proteins and is therefore considered a marker for animal glues when found in paint samples using AAA. However, hydroxyproline also occurs in high quantities in some glycoproteins found in plant cell walls, which might be causing confusion when analysing gum containing paint samples. Proline and hydroxyproline each constitute about 1/7 of the total sequence (table II.2), which thus is characterised by regularly repeating Gly-Pro-Y and Gly-X-Hyp triads. These three amino acids play an important role in the structure of collagen: the large side chains of proline and hydroxyproline force the chain in sharp twists, while glycine, without side chain, is required at every third position (ca. 1/3 of the total sequence). The assembly of the triple helix puts this residue at the interior of the helix, which allows for a very close and tight wrapping of the three α -chains. Animal glue is further characterised by its low amounts of essential amino acids (low nutritional value), such as valine, phenylalanine and leucine. (38)

Throughout the animal kingdom, collagen has changed little during evolution, and therefore it is nearly impossible to identify the animal source of a glue based on its amino acid composition(25). Fish glues contain a slightly lower amount of proline and hydroxyproline, which is supposed to affect the wrapping and thus the stability of the collagen molecules(12,26).

Table II.2. Amino acid composition of proteinaceous binding media (in mole percent); Cys, Tyr, Trp, His, Arg, Asn, Gln and Hyl did not yield a reproducible result using the applied AAA method and are therefore not included. (source: Schilling et al.(26))

Amino acid	Animal glue	Egg glair	Egg yolk	Milk casein
Alanine	12.1 ±0.7	10.5 ±0.7	9.4 ±0.6	5.1 ±0.7
Aspartic acid	4.9 ±0.4	11.0 ±1.6	12.2 ±2.6	7.2 ±1.9
Glutamic acid	7.4 ±0.5	12.7 ±1.2	11.8 ±1.9	19.6 ±1.5
Phenylalanine	1.4 ±0.2	5.1 ±0.5	3.6 ±0.6	4.1 ±0.5
Glycine	36.0 ±0.8	7.1 ±0.4	6.3 ±0.1	4.0 ±0.5
Isoleucine	1.4 ±0.1	6.1 ±0.9	6.0 ±0.7	6.0 ±0.9
Lysine	2.8 ±0.3	6.3 ±1.3	6.1 ±1.1	7.5 ±1.2
Leucine	2.9 ±0.3	10.3 ±0.9	10.4 ±0.4	10.3 ±1.5
Methionine	0.6 ±0.4	2.9 ±1.5	1.8 ±0.6	2.5 ±0.7
Proline	12.9 ±0.8	5.0 ±0.5	5.3 ±0.4	13.5 ±2.6
Serine	3.8 ±1.0	9.1 ±1.2	11.5 ±0.8	7.3 ±1.1
Threonine	2.0 ±0.6	5.3 ±0.9	8.3 ±3.0	4.9 ±0.8
Valine	2.3 ±0.2	8.7 ±1.5	7.4 ±0.5	7.9 ±1.2
Hydroxyproline	9.6 ±1.6	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0

II.3.2 Egg based binders

Chicken eggs have been widely used as paint binding media, as a whole or as yolk or even egg white (glair, albumen) solely. Paints that are based on yolk and water are very often denominated *egg temperas*, but its composition varies, sometimes including egg white, linseed oil, etc.

Glair

For glair binders, the egg white is stirred into a stable foam – this denatures the proteins – and left to stand overnight; the reliquidised fraction is used as medium.

Proteins are, next to water, the largest group of ingredients of egg white, about 15% of the total mass. The most abundant protein in the white is ovalbumin (ca. 54%), a 45 kDa phosphorylated glycoprotein. Other major proteins are ovotransferrin (formerly conalbumin, ca. 12%, 76 kDa), ovomucoid (ca. 11%, 28 kDa), ovoglobulins G2 and G3 (each ca. 4%, 49 kDa), ovomucin (ca. 3.5%, 5.5-8.3 kDa) and hen egg white lysozyme (ca. 3.4%, 14.4 kDa)(44).

In terms of amino acid composition (table II.2), egg white is characterised by relatively high amounts of aspartic and glutamic acid, in more or less the same amount.

Yolk

To prepare yolk binder, the yolk is separated from the white, dried on a towel without breaking the membrane and held above a recipient. The yolk membrane is subsequently punctured carefully to avoid that fractions of the membrane end up in the drained yolk, which is ready to be used as a binder, eventually slightly diluted with water. A small amount of vinegar is sometimes added as a preservative. (45)

Egg yolk's three main components are water, lipids (ca. 27%) and proteins (ca. 16%). Yolk is a notable source of cholesterol, lipid soluble vitamins and also of lecithin, a non-proteinaceous emulsifier. It are, however, the proteins in the form of complex lipoproteins that are responsible for emulsifying the complete lipid fraction, that is either present in triglycerides and phospholipids. The protein fraction of these lipoproteins are in this context called apoproteins. On the basis of its dry matter, yolk has five major constituents: 68% low-density lipoproteins (LDL, lipovitellenins), 16 % high-density lipoproteins (HDL, lipovitellins), 10% globular proteins (livetins), 4% phosphoprotein (phosvitin), and 2% minor proteins. (46)

The LDL or lipovitellenins are micelles-like spherical particles (17-60 nm diameter) with a lipid core and surrounded by a monolayer of phospholipids and apoproteins, which contributes only to 25% of the total LDL mass (figure II.6). There are six major LDL apoproteins (apovitellenins), all of which are glycosylated. There is a lack of knowledge concerning the exact identification of the apoproteins of

LDL. Very-low density lipoproteins (VLDL) of hen blood seem to be the precursors of egg yolk LDL. VLDLs contain mainly two apoproteins: apo-VLDL II and apo-B. During its transfer into the yolk, hen apo-B is enzymatically cleaved, resulting in the production of apo-B fragments. The only apoprotein from blood lipoproteins to be transferred to yolk in large amount without any modification is apo-VLDL II, called apovitellenin I in the yolk. (46)

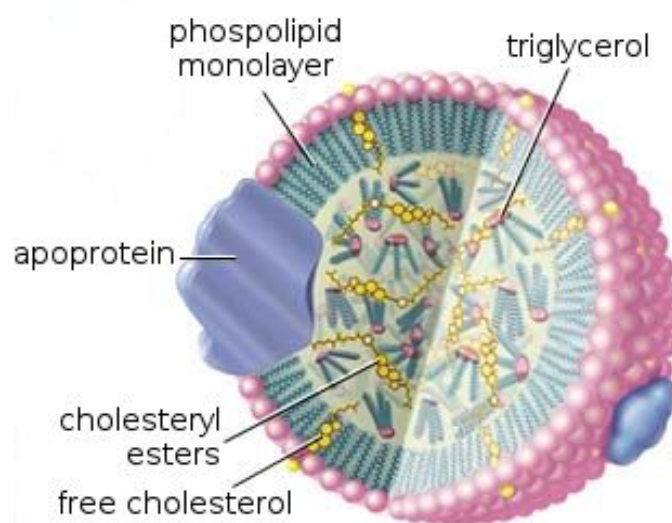


Figure 11.6. Low density lipoprotein structure. (source: adapted from Encyclopædia Britannica(47))

The HDL or lipovitellins are smaller structures (7-20 nm diameter) that have a more molecular form and much lower emulsifying properties: 75-80% proteins versus only 20-25% lipid fraction. HDL are dimers, each having a cavity formed by two β -sheets of mainly hydrophobic amino acids. In this cavity about 35 phospholipids are bound, which form the hydrophobic environment needed to contain the triglycerides and cholesterol. Each monomer of HDL is composed of about five glycosylated apoproteins (vitellins), which have their origins in the precursor vitellogenin chains. Vitellogenins are dimeric proteins synthesized in the liver. During the transfer into the oocyte, they undergo a proteolytic cleavage generating, amongst others, the vitellins. (46)

Phosvitin is a phosphoglycoprotein with almost 50% of its amino acids are serine, out of which 90% are phosphorylated. There are two different phosvitins: an α -form (160 kDa) and a β -form (190 kDa). Both forms are aggregates of several subunits

that are synthesized from the same vitellogenin precursors that include the HDL vitellins. (46)

The livetins are a small group of different water-soluble proteins that correspond to hen blood serum proteins: α -livetin equals serum albumin (70 kDa), β -livetin is α_2 -glycoprotein (45 kDa) and γ -livetin is immunoglobulin Y (170 kDa), all present in blood serum. (46)

Although composed of entirely different proteins, yolk and glair have almost the same amino acid composition (table II.2, page 33). Due to these subtle differences, analytical techniques based on AAA are seldom able of distinguishing between the two (or mixtures of them)(12,25). Moreover, their amino acid profiles are also similar to that of other proteins, such as serum albumin that is present in blood(25).

II.3.3 Milk-based binders

Milk is an emulsified colloid of fat in a water-based solution. Most milk-based binders are based on the curd of the milk. This is the solid fraction that precipitates when adding an acid (such as vinegar or lemon juice) to skimmed milk. The whey (milk plasma) remains after curdling and straining of the milk. Most recipes use fresh fat-free curd cheese (quark), which are essentially the curds, as the base for casein paint (sometimes termed *milk tempera*). With water they form sludges, but with alkaline solutions they form a colloidal suspension. Traditional recipes generally use slaked lime, whereas more recent recipes use ammonia or ammonium carbonate. Also, borax (disodium tetraborate) is sometimes used to hydrolyse and tackify the caseins (so called borax casein paints).

The curd proteins are called caseins: phosphoproteins that are involved in emulsifying the lipid fraction, and account for 80% of the total proteins in milk. Four different caseins can be found: α_{s1} -casein, α_{s2} -casein, β -casein and κ -casein. The exact structure of the casein micelle is a point of discussion amongst dairy scientists. Several micellar models have been proposed, but none is accepted universally(48). In all models κ -casein is regarded as a key component in forming the outer edge of the micelles, while the other caseins and the many phosphate groups (so called

colloidal calcium phosphate) in the interior of the micelles are responsible for many of the important characteristics of casein, especially its ability to bind relatively large amounts of calcium(49).

The whey contains the remaining 20% of the proteins. As a consequence, in casein binding media, the concentration of whey proteins should be minimal. The main whey proteins are β -lactoglobulin (50% of the whey proteins), lactalbumin (20%), blood serum albumin (10%) and several immunoglobulins G₁, G₂, M and A (20% in total, respectively IgG₁, IgG₂, IgM, IgA). (49)

The amino acid composition of casein based binders is shown in table II.2 (page 33).

II.3.4 Other protein sources in paint

Few extensive studies on analytical methods for proteinaceous components in paints exist, taking into account a broad set of protein sources, not only from the binders, but also minor protein fractions in other natural ingredients or introduced by conservation treatment or contamination. In this respect Schilling and co-workers compiled a database of amino acid composition from diverse protein sources in paints(26).

Paint ingredients that contain proteins either as a principal or a minor ingredient include blood, flour, garlic and gums. About 60% of the proteins present in blood are serum albumins, while several immunoglobulins account for 18%. Flour, made from wheat (*Triticum spp.*) or other plant sources, was occasionally used in paints (chapter VIII) and contain about 10% (w/w) protein matter. Gliadin and glutenin comprise about 80% of the protein contained in wheat seed. Garlic (*Allium sativum*) was as an adhesive in gildings and contains 15–17% (w/w of dry matter) proteins(42). Gums, made from plant exudates, are mainly composed of polysaccharides, but contain a minor fraction of glycoproteins. High amounts up to 29% of hydroxyproline are present in gum arabic(26), the most frequently and widespread gum made from *Acacia spp.* and was used as an adhesive or as a paint or ink medium.

Other sources of proteins in paints are introduced by contamination, such as saliva, skin or hair fragments and bacterial or fungal activity. Saliva was (and regrettably still is) often used as a ready-to-use cleaning agent or used to wet the paint brush. It contains a small amount of digestive and antimicrobial enzymes. Keratins are major proteins in both the epidermis and hair; as such avoiding contamination of paints with keratin is nearly impossible, as is contamination of the paint samples for analysis during the sample preparation. Finally, in non-ideal storage circumstances, a high humidity for example, microbe or fungal activity might introduce proteins into the paint layers.

II.4 Chapter epilogue

The knowledge of the analytes, the proteins found in works of art, is the first requirement in order to understand the limitations of the current analytical techniques and to develop new, better performing alternatives. In the following chapter both current and new approaches will be discussed.

III

PROTEIN BINDER ANALYSIS

The quest for novel, high-performing methods of analysis for protein binders is nowadays focused on two recent developments in proteomics techniques: immunostaining on the one hand and peptide analysis on the other. In this thesis the focus is on the latter. Herein, enzymatic cleavage of the protein matter results in smaller, yet very characteristic peptides. To justify the need for and the advantages of this approach, the *modus operandi*, features and drawbacks of other frequently applied analytical techniques for protein binders will be shortly reviewed in this chapter. This is followed by an in-depth discussion on the technical aspects and the scientific reasoning of the proteomic procedures adapted for art investigation.

III.1 General overview of the analytical techniques

The analysis and identification of paint and its components are of the utmost importance for conservators, conservation scientists and art historians. During conservation, the parties involved are often faced with questions regarding the composition of paint, its originality or how to selectively remove an overpainting or oxidised varnish layer. Art historians studying a specific painter or an artistic style want to know which products are characteristic for that painter, style or epoch.

The enormous diversity in chemical composition of binders thwarts a single analytical method that is suitable to identify all different products in detail. Albeit that some methods are capable of distinguishing the main binder classes from each other, for instance proteinaceous binders from oils, they often fail in determining the exact protein source. These analytical methods, sometimes non-destructive or even non-invasive, are in this context called non-specific methods and can be used to help selecting the most suitable detailed analytical approach afterwards. In many cases the results obtained with these methods are sufficient for the conservator or art historian. A whole range of non-specific analytical techniques have been developed that can distinguish between the major binding media groups such as oils, proteins, gums, waxes, resins and synthetic binders. The most important in this respect are the historically first but still practised microchemical tests(50) and the modern spectroscopic techniques, including Fourier-transform infrared spectroscopy (FTIR)(6,7,51–55), near-infrared spectroscopy (NIRS)(56), micro-Raman spectroscopy (MRS)(57–59) and laser-induced fluorescence (LIF)(60,61).

Sometimes a more detailed result is required, such as the source of oil or the type of protein binder. Traditionally, (micro-)destructive, mostly chromatographic techniques have been used for this task. These specialised methods are able to discern different types of one, or at the utmost few, binding media groups. For protein binders, by far the most widely used techniques are chromatography based amino acid analysis (AAA)(25–27,29,30,28,31,32) and pyrolysis(33–36), both having considerable drawbacks mainly due to the small and often non-characteristic markers that are monitored. One of the consequences of this is the fast-growing complexity up to non-interpretable data in case of mixtures of different protein sources, the presence of protein contaminants or unexpected sources of proteins. Both AAA and pyrolysis based techniques require a small amount ($\geq 100\text{ }\mu\text{g}$) of paint sample. While efforts to minimise the quantity were successful, sampling remains perceived with reluctance by conservators and restorers or simply cannot be tolerated due to specific characteristics of the work of art. Minimising the sample quantity also hampers the possibility to obtain a representative and layer-pure sample of a specific paint layer, which is crucial for these types of analytical techniques.

To tackle these downsides, attempts were made to leverage non-specific methods, such as MRS, with mixed success(58). Over the recent years, statistics came into play to process spectroscopic data. The interest to use spectroscopy for protein analysis lays not only in the quest for a universal analytical technique, but would also enable both cross-sectional and non-invasive, even *in situ* measurements. Working with cross sections permits to obtain valuable information on the distribution of the binding media over the whole of the paint layers, something that simply cannot be done with AAA or pyrolysis. While generally good results are obtained on fresh laboratory samples with MRS, very few positive results are published on aged laboratory and historical paint samples.

Another approach focuses on immunostaining, which is based on the highly specific antibody-antigen interaction. Immunostaining is a common name for a range of techniques used in the field of proteomics. The first application on proteinaceous binders, already in 1971, was based on immunofluorescence microscopy (IFM)(62). This technique uses antibodies that are tagged with a fluorophore and can be carried out on cross sections. However, application is limited to dark paints that do not exhibit autofluorescence and the response is generally weak and often leads to false results. Due to these drawbacks, scientific papers dealing with this topic appeared very sparsely over the next decades(50,63–65). More recently, novel immunostaining techniques that are developed for proteomics were applied in conservation science, in which tags are described that offer improved reliability, but these still suffer from the high number of false positives and negatives(66–69). Another immunostaining technique, enzyme-linked immunosorbent assay (ELISA) has a higher sensitivity but requires (and consumes) a layer-pure sample.(65,70,71)

There is an obvious need for better performing protein binder analytical techniques, requiring even smaller microsamples and more importantly, giving accurate and precise results, also in the cases where all other methods fail. The state of the art techniques that are nowadays applied in proteomics(72–74) potentially offer several advantages over the techniques that are commonly used for proteinaceous binding media analysis. Breaking down the protein matter by enzymatic cleavage into peptides that remain highly specific for the protein, allows for unambiguous identification of the proteins, even in complex mixtures. Typically,

matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS)(75–84) or high-performance liquid chromatography electrospray ionisation mass spectrometry (HPLC-ESI-MS)(85–90) are used to measure these peptides. Both techniques offer a very high sensitivity and therefore they generally require only low sample amounts. Their application on protein binders requires adapted sample preparation and measurement procedures in order to cope with the specifics of these samples: the presence of pigments and other interfering paint components, the aged and degraded state of old paints and the difficulty to dissolve these paints.

In the next sections, the two classic techniques for protein binders (amino acid analysis and pyrolysis) and the two emerging techniques (immunostaining and peptide analysis) will be discussed in depth.

III.1.1 Amino acid analysis (AAA)

In this technique, protein characterisation is based on amino acid profiles. Amino acids are rarely characteristic for a protein as most proteins contain all standard amino acids. A notable exception on this rule is hydroxyproline, a non-standard amino acid, which only occurs in collagen (animal glue) and in some glycoproteins found in plant cell walls (gums). The relative ratios of the amino acids, however, are fairly representative. Several approaches for AAA have been proposed for use on proteinaceous paint microsamples, but in general they are based on the same principles and follow a similar procedure with successively the hydrolysis, derivatisation, separation and detection steps. The separation can either be in liquid(25,27,29,30) or gas phase(26,28,31,32,30), implying different derivatisation and detection methods.

In the first step, the peptide bonds of the proteins in the paint sample are hydrolysed using 6 M HCl at elevated temperatures (105°C to 110°C). Proteins are thus broken down into their individual free amino acids. Some procedures prescribe hydrolysis under a nitrogen atmosphere(28) and/or in gaseous phase(27). This procedure improves the preservation of certain amino acids, such as tryptophane and cysteine, that would otherwise oxidise and it also avoids

contamination. Microwave-assisted acid hydrolysis is proposed as an alternative that drastically reduces the long reaction time of up to 24 h till 40 min (91,92).

To achieve high sensitivity and low detection limits, most labs that use HPLC-based AAA, derivatise the amino acids before detection. A popular method is pre-column derivatisation with a fluorescing agent, such as 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AccQ fluorescence reagent) in conjunction with fluorescence detection(27,30). To analyse them with gas chromatography coupled to mass spectrometry (GC-MS), the amino acids need to be derivatised to increase their volatility. Many derivatisation agents have been proposed based on silylation (e.g. using *N*-*tert*-butylmethylsilyl-*N*-methyltrifluoroacetamide, MTBSTFA(28), figure III.1) or alkylation (e.g. ethyl chloroformate, ECF(26)).

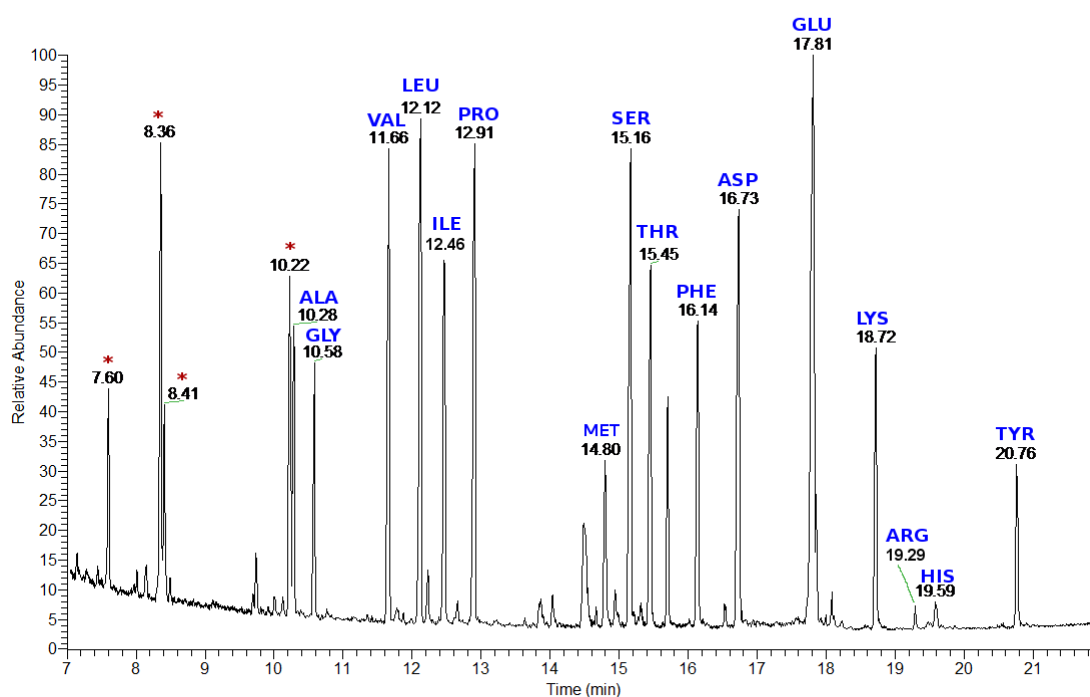


Figure III.1. GC-MS chromatogram obtained on model milk casein based paint sample (see page 80) after derivatisation with MTBSTFA. The *tert*-butyltrimethylsilyl derivatives of the amino acids are annotated in blue; MTBSTFA system peaks are marked in red.

Independent of the choice between HPLC or GC-MS, different methods for data treatment can be used. Evidently, the chosen data treatment procedure needs to be adapted to exclude amino acids that are unstable in the followed

measurement procedure. Schilling considers only seven “stable” amino acids(93). The response of the peak heights or integrations are corrected using a standard with known amino acid concentrations. Relative amino acid quantities are commonly expressed in percentages of the summed amount. Basic interpretation can be done based on these values. The easiest example of this is a moderate amount of hydroxyproline that is characteristic for animal glue. Pearson's r correlation coefficient of the seven stable amino acids is frequently used to compare the set of relative amounts of amino acids with those of reference products(94). Finally, principal component analysis (PCA) is often used to classify unknown amino acid profiles(95).

It is important to understand that protein binders are mixtures of different proteins. Egg white's main protein is ovalbumin, but it only accounts for 54% of the total protein fraction. Other egg white proteins (ovotransferrin, ovomucoid, ovomucin, ovoglobulins, lysozyme) have a significant impact on the amino acid profile of an egg white binder. The same is equally true for the other binders. The characterisation of these binders with AAA should thus be based on the amino acid profiles of a set of reference samples of complete binders, not on these of pure proteins in published data.

The availability of HPLC and GC-MS, the high sensitivity of the analytical methods and the relatively fast sample preparation protocols made AAA the method of choice in most conservation science labs. However, the concept of amino acid profiles has some serious drawbacks. A distortion of any kind in a single amino acid peak has a tremendous impact on all relative amino acid quantities, rendering the sample unidentifiable or falsely identified. Such a distortion can have various causes, ranging from co-eluting components or selectively degraded amino acids. Some paint components, copper based pigments for instance, are known to react with amino acids(96). Also mixed protein binders, impregnated proteinaceous products (original, e.g. isolation layers, or applied during a conservation treatment), samples comprising multiple layers, unexpected proteinaceous additives or contaminants entangle a correct identification. Although amino acid profiles obtained by AAA can be compared with theoretical mixtures of reference samples using Pearson's r correlation coefficients, the results are rarely reliable: high

correlation coefficients for the same sample can frequently be obtained using totally different theoretical mixtures.

III.1.2 Pyrolysis GC-MS

Pyrolysis, most frequently coupled to a gas chromatograph and a mass spectrometer (py-GC-MS), is a routine technique for the analysis of paint materials. It is commonly applied for the identification of natural(97–100) and synthetic resins(101–104), gums(21,105), waxes(106,107) and even synthetic organic pigments(18,108). Successful classification of protein binders as egg, milk casein or animal glue has been described by several authors(33–36). Pyrolysis thus qualifies easily as an allround technique that allows for the identification of a broad range of organic paint components.

Thermochemolysis, or thermally assisted chemolysis, is a designation for pyrolysis with online derivatisation. This is often a methylation reaction (thermally assisted hydrolysis and methylation, THM) using derivatisation agents such as tetramethylammonium hydroxide (TMAH). As an alternative to methylation, silylation is also used. Thermochemolysis adds support for the identification of drying oils and improves the results on other classes(16,109,110).

In its basic set-up minimal amounts of paint scrapings are applied without any sample pretreatment on a suitable carrier (wire or sample cup), depending on the type of pyrolysis instrument. To aid sample transfer and to obtain better contact with the heated surface, finely ground paint sample is often dispersed in a fast evaporating liquid such as methanol and subsequently pipetted onto the carrier.

Even at temperatures as low as 375°C, the proteins are completely broken down into small pyrolysis products, that are often poorly specific and low in intensity. In general, the marker component peaks are only visible when using selected ion monitoring (SIM) on the masses of these marker ions. Chiavari(34) reports pyrrole (m/z 67) and diketopyrrole (m/z 186) as thermal degradation products of hydroxyproline; these are thus markers for animal glues. Toluene (m/z 91), indole (m/z 117) and skatole (6-methylindole, m/z 130) are thermal degradation products of phenylalanine and thryptophane. These are not specific for one of the

common protein binders, but are significantly more abundant in egg-based binders. Milk casein-based paints, on the other hand, can be characterised by markers of carbohydrates (glycoproteins) instead of markers from protein pyrolysis products: furanmethanol (m/z 81+97) and maltol (m/z 126) are found in the pyrograms in a higher abundance in paints samples based on milk casein, than in those based on egg(34). Interpretation in case of unexpected proteins, binary mixtures or keratin-contaminated samples is very hard.

In presence of TMAH a range of methylated side products can be observed in addition to the above mentioned markers, making the pyrograms even harder to interpret. Markers that needed SIM monitoring without methylation, often end up barely higher than the noise levels(34). Although not ideal for protein identification, this methodology is a compromise that often is used as a universal analytical approach for binding media. As a side effect, the derivatisation agent, dissolved in water or methanol, more firmly attaches the dispersed analyte to the pyrolysis wire through crystallisation upon evaporation of the solvent. This way sensitivity and repeatability of the measurements improves.

III.1.3 Immunostaining

The principle of immunostaining is based on the highly specific antigen-antibody interaction. Antibodies (immunoglobulins) play an important role in the immune system: the Y-shaped proteins recognise and bind to specific regions – the antigen – of foreign molecules (bacterial cell walls, viruses, foreign proteins or macromolecules). When these pathogens enter the body, the organism produces adapted antibodies that can bind to the antigen site of the foreign molecule. This way the intruding molecule is “marked” for destruction and in some cases also deactivates it. (64,65)

Immunostaining is a common term for a range of techniques that use this principle to visualise the presence of specific proteins. Highly specific antibodies can be cultivated that bind only to a specific target protein (the antigen) present in proteinaceous binders. Once extracted and purified, the antibodies can be fitted with a functional group that enables their detection. A solution with these labelled antibodies is subsequently dropped onto the (paint) sample. If the corresponding

protein is present in the latter, the labelled antibodies bind; if not, the antibodies are washed away (figure III.2 A). Nowadays a double antibody approach is preferred: in a first step an unlabelled primary antibody binds to the antigen, while in the second step a labelled secondary antibody binds to a conserved region of the primary antibody (figure III.2 B). The primary antibody thus acts as the antigen for the secondary antibody. This operation procedure allows a greater versatility: the same labelled secondary antibody can be used for a range of different primary antibodies (cultivated in the same species). As an additional advantage, in some cases primary antibodies can be fitted with multiple secondary antibodies (not shown in the figure), which leads to a multiplication of the signal and sensitivity. (64,65)

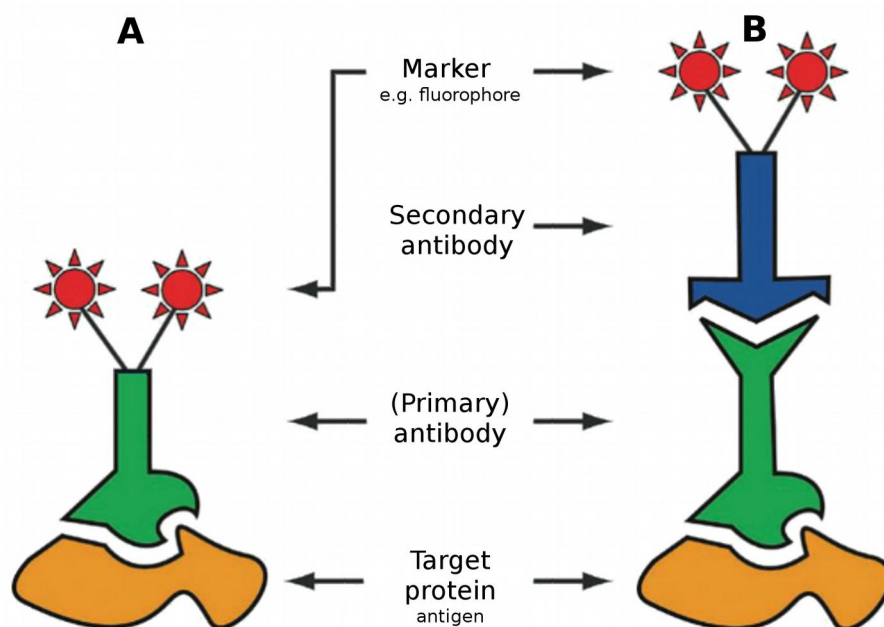


Figure III.2. Principle of immunostaining with (A) single and (B) double antibody approach. (source: adapted from Heginbotham et al. (65))

The earliest applications of immunostaining in conservation science involved a fluorescent dye as a label and ovalbumin (egg white) as antigen(62–65). A solution with antibodies was directly applied on a cross section; the stratigraphic distribution of the fluorescing antibodies, and thus egg white binder, was visualised with a standard fluorescence microscope. Only recently, Cartechini reported the use of IFM for the identification of casein and collagen (animal glue) in cross sections(68).

As is the case with traditional staining techniques on cross sections, IFM often suffers from false positive and false negative results. Firstly, the fluorescence response is known to be weak (false negatives). Secondly, in some cases non-specific binding of the antibodies is observed (false positives). Finally, light colours and autofluorescence of pigments and binders may seriously hamper the interpretation (false positives), although this can in many cases be tackled by the sequential use of fluorophores that have a response at different wavelengths. This in turn, significantly increases the workload and cost of IFM analysis, and in between each application a thin layer of the cross section needs to be removed by polishing. Another disadvantage of IFM is its long reaction time of the antibody solution on the surface of the cross section, which is reported to cause dissolution of some of the paint layers. (69)

Over the past few years, alternatives for fluorescent dye labels have been proposed to resolve some of the reasons for false results. Cartechini(68) discussed so called quantum dots (nanoparticles of semiconductors composed of groups II-VI and III-V elements) as fluorescent markers for antibodies, giving a very sharp and more specific fluorescence signal. Chemiluminescent labels solve the problems related to autofluorescence of pigments and binders(66,67). Both alternatives, however, require a specialised epifluorescence microscope and nitrogen cooled ultra-sensitive CCD detector, which are not available in most conservation science labs. Another technique is based on surface-enhanced Raman spectroscopy (SERS): antibodies equipped with SERS-active labels allow for very specific analysis on a cross section using a micro-Ramanspectrometer(69).

ELISA, another immunostaining technique that has been used for the identification of proteinaceous binders, offers a much higher response, but requires a layer-pure paint sample to be taken. The proteins are extracted from this sample and applied on an ELISA plate (polystyrene, PS) on which the proteins bind. Primary (unlabelled) antibodies are then added and left to bind with the proteins (if present). After washing the unbound antibodies away, enzyme-conjugated secondary antibodies are left to bind the primary antibodies (if present). Usually alkaline phosphatase is used as enzyme(65). After washing the unbound secondary antibodies away, a colourless substrate (e.g. *p*-nitrophenyl phosphate, *p*NPP) is added that can be enzymatically catalysed into a coloured product (e.g. nitrophenol) if the bound

enzyme is present, or thus if the sought-after protein was present in the paint sample (figure III.3). An absorbency spectrophotometer is used to detect these colour changes. The enzymatic reaction instigates an amplification of the response in comparison with IFM, which explains the superior sensitivity for this technique. (65,70,71)

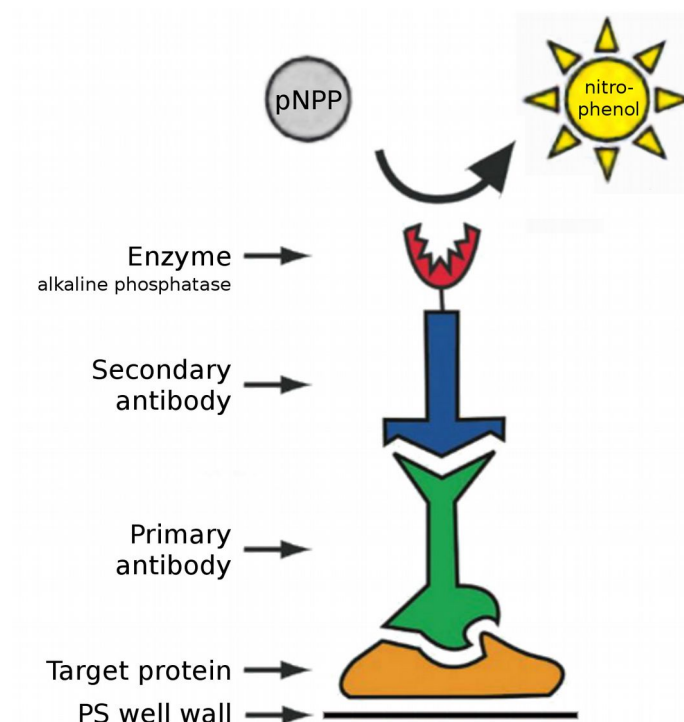


Figure III.3. Principle of enzyme-linked immunosorbent assay. (source:adapted from Heginbotham et al.(65))

III.1.4 Peptide analysis

While the study of proteins through the analysis of peptides has long become an established method in proteomics, its application in conservation science is picking up fast. In peptide analysis, the proteins are first enzymatically cleaved into small, yet very characteristic oligopeptides. The 5 to 15 amino acids in these peptides allow for an exponential number of combinations and are thus characteristic for the parent protein. Due to mutations between the corresponding proteins in different animals, many of these peptides even allow for identifying the species of the protein.

Over the past years different methodologies emerged for the identification of proteins through peptide analysis in artists' paints, historical mortars and archaeological remains. Essentially, these methods differ mainly in the way the proteins are extracted from the century old dry samples and the detection method: HPLC with diode array detector (HPLC-DAD, chapter IV), MALDI-TOF-MS (chapter V, (75–84)) or HPLC-ESI-MS (chapters VI and VII, (85–90)). The latter is in most cases a quadrupole time-of-flight tandem mass spectrometer (HPLC-ESI-QTOF-MS/MS). In the next sections these peptide analysis approaches will be discussed profoundly, with the focus on proteinaceous paints.

III.2 Sample pretreatment in peptide analysis

Different research groups active in the development of methods for peptide analysis for paint samples independently implemented their own sample pretreatment protocols. Evidently, these are adapted to and optimised for the chosen method of detection. The principles and properties of these are covered in the next sections. However, a large diversity also exists in how the proteins are extracted from the sample and how they are dissolved. Age-old paint samples are expected to be significantly harder to dissolve and are possibly not susceptible to the solvents that are used to dissolve fresh proteinaceous material. A common pathway is observed in the published methods: the first series of steps involves dissolving and denaturing, followed by the proteolysis step, clean-up and instrument-specific sample preparation. An overview of published methods that are used on paint samples, is shown in table III.1.

III.2.1 Dissolving and denaturation

To attain the maximal proteolysis yield, proteins are typically dissolved and denatured before proteolysis. Leo *et al.* (87) use alkaline extraction with ammonia to separate protein content from the rest of the paint matrix. This reaction step, as part of an extensive procedure, is only done in cases where direct digestion (without any protein extraction, dissolving and denaturation steps) fails. Dissolving centuries old proteinaceous samples such as paints can be a daunting task. To aid this process, grinding and sonification are often used. Chaotropic agents, such as sodium dodecylsulphate (SDS), urea, thiourea, 2,2,2-trifluoroethanol (TFE) and

guanidine hydrochloride, disrupt the molecular structure of the proteins and aid in denaturation by breaking non-covalent bonds (hydrogen bonds, Van der Waals forces) in the protein structure. The addition of high quantities of chaotropes is often problematic for proteolysis (section III.2.2, page 55) and detection (section III.2.3, page 57), and therefore requires extra reaction steps to remove or dilute them.

Disulphide bonds, responsible for much of the tertiary structure of proteins, are commonly reduced at elevated temperatures by 1,4-dithiothreitol (DTT) or *tris*(2-carboxyethyl)phosphine (TCEP). The reaction mechanism of disulphide bond reduction using DTT is shown in figure III.4 A. To avoid refolding of the tertiary structure, the thiol groups need to be stabilised. The cysteine free thiol groups are irreversibly alkylated into *S*-carboxyamidomethylcysteine by 2-iodoacetamide (IAA) at room temperature in the dark (figure III.4 B). When using this reagent in combination with mass spectrometry, a mass difference of 57.02 Da for each cysteine in the peptides has to be taken into account; in most software packages one needs to include the corresponding post-translational modification.

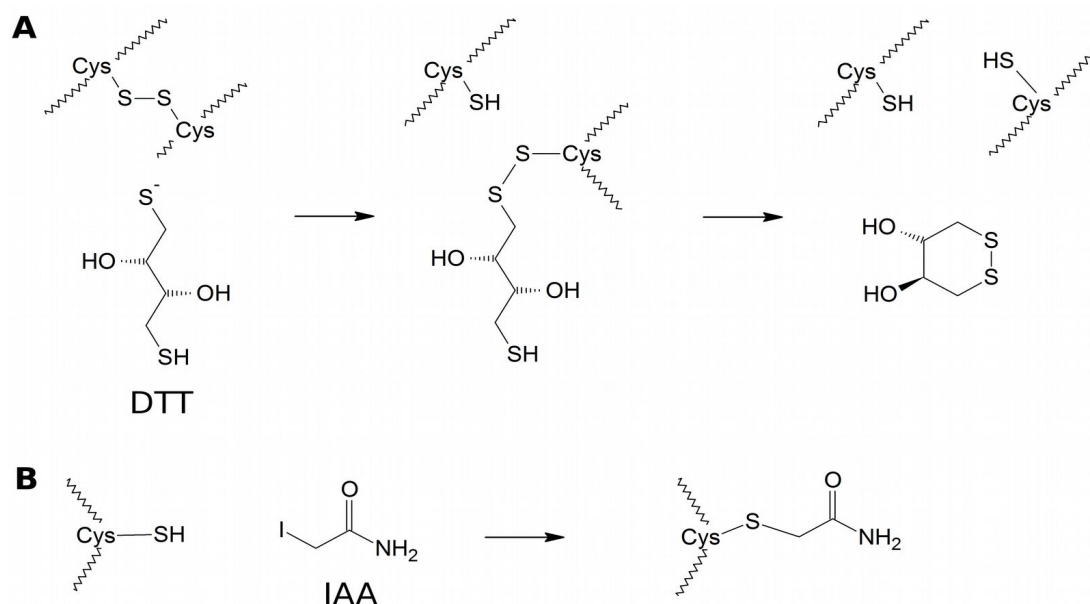


Figure III.4. Reaction mechanism of (A) cysteine disulphide bond reduction by DTT and (B) free thiol alkylation by IAA.

Table III.1. Published peptide analysis methods on paint samples.

	Kuckova (77), chapter V	Kirby (82)	Tokarski (85)
Dissolution and denaturation		30 μ L TFE +30 μ L 50 mM ABC 60°C, 30 min Sonificate 60°C, 10 min Vacuum centrifuge +30 μ L 50mM ABC 60°C, 30 min +3 μ L 20mM TCEP 37°C, 20 min +3 μ L 40mM IAA RT, dark, 30 min	50 μ L 1% TFA Crushing resin Extraction in H ₂ O Sonification 80 μ L 6 M guanidine, 100 mM Tris, pH 7.25 50°C, 30 min + DTT (to 12 mM) 50°C, 4h + IAA (to 300 mM) RT, dark, 30 min
Proteolysis	15 μ L 400 ng μ L ⁻¹ trypsin in 50 mM ABC RT, 2h	8 μ L 20 ng μ L ⁻¹ trypsin in 50 mM ABC 37°C, >2h	1 ng trypsin in 50 μ L ABC (/10 μ g sample) 37°C, overnight
Clean-up	μ -SPE: ZipTip C18 4 μ L 0.1 M DHB, 0.1% TFA in 7:3 water/ACN	2 μ L sample solution + 20 μ L saturated CHCA solution in 40% ACN, 0.1% TFA	<i>Poorly documented</i> DHB solution for MALDI-TOF- MS
Instrument	MALDI-TOF-MS	MALDI-TOF-MS	HPLC-ESI-QTOF-MS/MS MALDI-TOF-MS

Abbreviations: **ABC** ammonium bicarbonate, **ACN** acetonitrile, **CHCA** α -cyano-4-hydroxycinnamic acid, **HILIC** hydrophilic interaction liquid chromatography, **IAA** iodoacetamide, **RT** room temperature, **Tris** tris(hydroxymethyl)aminomethane, **μ -SPE** miniaturised solid phase extraction

Assuming that proteins in historical and laboratory paint samples are already denatured by natural processes, Kuckova(77) and (for most samples) Leo(87) skip the dissolution and denaturation steps completely and perform proteolysis directly on untreated microsamples. While this unavoidingly impacts the proteolysis yield, as not every protein or protein fragment is reachable for proteolysis, it also has several advantages. Firstly, analysis becomes more economical and faster: not only these steps are avoided, also eventual extra steps to neutralise or remove the

Fremout 2009 chapter IV, (88)	Leo (87)	Fremout 2010² chapters VI and VII	Dissolution and denaturation
Grinding with spatula + 25.5 µL 50 mM ABC Sonification	None OR 3x (500 µL 2.5 M NH ₃ Sonification, 45 min Collect liquid Evaporation + 6M guanidine, 0.3 M Tris, 10 mM EDTA, pH8	50 µL 0.7% SDS, 5 M urea, 2 M thiourea Sonification	
+ 1.5 µL 100 mM DTT 95°C, 5 min	+ DTT (to 40 mM) 37°C, 2h	+5 µL 110 mM DTT 60°C, 60 min	
+ 3 µL 100 mM IAA RT, dark, 15 min	+ IAA (to 845 mM) RT, dark, 30 min Desaltion on column in 10 mM ABC	+5 µL 120 mM IAA RT, dark, 15 min +440 µL 50mM Tris, pH7.6	
2 µL 100 ng µL ⁻¹ trypsin in 50 mM ABC 37°C, overnight	16 µM trypsin in 10mM ABC 37°C, 16h OR microwave 700 W, 15 min	5 µL 10 ng µL ⁻¹ trypsin in 50 µL ABC 37°C, overnight	Proteolysis
Vacuum evaporation 25 µL 0.01% FA	<i>Poorly documented</i> Probably none	SPE: TopTip HILIC Vacuum centrifugation 23.5 µL 0.1% FA	Clean-up
HPLC-DAD HPLC-ESI-QTOF-MS/MS	HPLC-ESI-QTOF-MS/MS	HPLC-ESI-QTOF-MS/MS	Instrument

DHB, 2,5-dihydroxybenzoic acid, **DTT** dithiothreitol, **EDTA** ethylenediaminetetraacetic acid, **FA** formic acid, **SDS** sodium dodecylsulphate, **TCEP** tris(2-carboxyethyl)phosphine, **TFA** trifluoroacetic acid, **TFE** tetrafluoroethanol,

added reagents can be skipped. Secondly, data treatment and interpretation is significantly simplified. The latter is of particular interest for MALDI-TOF-MS analysis: all peptides are measured simultaneously, causing complex mass spectra. In this case, the loss of sterically unreachable peptides and (incompletely) modified

2 In the currently applied analytical method, the amount of 120 mM IAA solution is increased to 10 µL in order to improve the alkylation yield of both the reduced disulphide bridges in the proteins and the DTT excess.

peptides is an advantage. At the same time there are fewer reagent peaks that potentially hinder the mass spectrum. Finally, risks of contaminations are reduced.

A comparison in terms of efficiency of the pretreatment methods used by the different research teams has yet to be made. Several authors showed interest to cooperate in a round robin test in the near future. It is, however, important to realise that the development of a sample pretreatment protocol is commonly based on calculated yields of a *limited* set of home made reference samples or – *subjectively* – on the results obtained on real life samples from works of art: the large variation of sample types, matrices and degradation symptoms may very well imply that there is no single method that fits all cases or outshines all others.

During this doctoral research, the three main detection methods (HPLC-DAD, MALDI-TOF-MS and HPLC-ESI-MS) are covered. Each of these requires an adapted sample pretreatment:

- For HPLC-DAD analysis (chapter IV) we have applied a relatively straightforward method, avoiding extraction steps and the potential sample loss involved. Microsamples are manually ground with a spatula in the sample tube and rigorously sonicated in a 50 mM ammonium bicarbonate buffer. The protein disulphide bonds are reduced using a 1.5 μ L of buffered 100 mM DTT solution. A short reaction time of only 5 min proves to be effective due to an elevated temperature of 95°C. The free thiol groups are subsequently alkylated using 3 μ L of freshly prepared buffered 100 mM IAA solution, which is let to react for 15 min. Due to the instability of the reagent this is done in the dark and at room temperature. This procedure proves to be successful for HPLC-DAD analysis, and was later applied by Chambery(88) using HPLC-ESI-QTOF-MS/MS. In many cases, however, an insoluble debris is observed, possibly indicating an incomplete dissolution.
- For MALDI-TOF-MS analyses (chapter V) we applied the established method as developed by Kuckova and colleagues(77). The dissolution and denaturation are skipped completely and the enzyme solution is added directly to the sample.

- The chromatographic separation, high specificity and high sensitivity of a HPLC-ESI-QTOF-MS/MS set-up (chapters VI and VII) enabled us to maximise the yield of the pretreatment. Historically, the different research teams independently developed different approaches. We used 50 μ L of a strong chaotrope environment of 5 M urea, 2 M thiourea and 0.7% SDS detergent, a common combination in proteomics, combined with rigorous sonification. While able of dissolving almost all proteinaceous samples, both models and real life age-old paint samples, several implications have to be taken into account. Firstly, this solution must be prepared freshly due to the instability of urea in aqueous solution. This avoids unwanted carbamylation of lysine, arginine and cysteine as post-translational modifications(111). For the same reason, the reduction step using DTT (10 mM final concentration) needs to be done at a reduced temperature and as a consequence during a longer time span (60 min at 60°C). Secondly, final concentrations of SDS above 0.1% deactivate trypsin(112). Therefore, an extra dilution step with Tris buffered solution is inserted before the proteolysis step. Finally, the high concentrations of urea, thiourea and SDS seriously disturb the HPLC system and need to be removed before injection using solid phase extraction (section III.2.3, page 57).

III.2.2 Proteolysis

Soft and selective cleavage methods are used to cut protein chains into peptides. In proteomics, a number of enzymes is used for this task, each exhibiting its own characteristics. The cleavage process itself is called proteolysis, the enzymes are called proteases, or more specifically endopeptidases (cleavage of peptide bonds of non-terminal amino acids). A selection of proteases is shown in table III.2. Trypsin is by far the most used protease in the field of proteomics. Up to present, this is also the only protease used for the analysis of proteins in conservation science. Buckley *et al.*(113) have experimented with collagenase for the identification of archaeological animal samples, but they are now also using trypsin.

Table III.2. Common proteases for peptide analysis.

Protease	Biological source	Cleavage point
Trypsin	Bovine pancreas	Carboxyl side of Lys and Arg
Submaxillarius protease	Mouse submaxillary gland	Carboxyl side of Arg
Chymotrypsin	Bovine pancreas	Carboxyl side of Phe, Trp and Tyr
V8 protease	<i>Staphylococcus aureus</i>	Carboxyl side of Asp and Glu
Asp-N-protease	<i>Pseudomonas fragi</i>	Amino side of Asp and Glu
Pepsin	Porcine stomach	Amino side of Phe, Trp and Tyr
Endoproteinase Lys C	<i>Lysobacter enzymogenes</i>	Carboxyl side of Lys
Collagenase	Different sources	Y-Gly in collagen Gly-X-Y triads

Trypsin is produced in and obtained from the pancreas. It cleaves the protein chain following a positively charged amino acid, thus the carboxyl side of lysine and arginine, except if either of these are followed by proline. It belongs together with chymotrypsin and others to the group of the serine proteases, in which a catalytic triad of histidine (His 57), serine (Ser 195) and aspartic acid (Asp 102) are responsible for the proteolysis(38). The catalysis procedure is summarised in figure III.5. Trypsin proteolysis efficiency is heavily subject to experimental conditions such as heat and acidity; ideal conditions are 37°C and pH 7 - 8. Thanks to the relatively high abundance in many proteins (including the vast majority of those expected in art objects) of lysine and arginine tryptic peptides of 5 to 15 amino acids are created. This is an ideal length in terms of uniqueness and thus in characterising their parent proteins, while their mass allow for precise detection by standard mass spectrometers.

The differences in implementation of the proteolysis step between the different research groups are minimal, as can be seen in table III.1 (page 52).

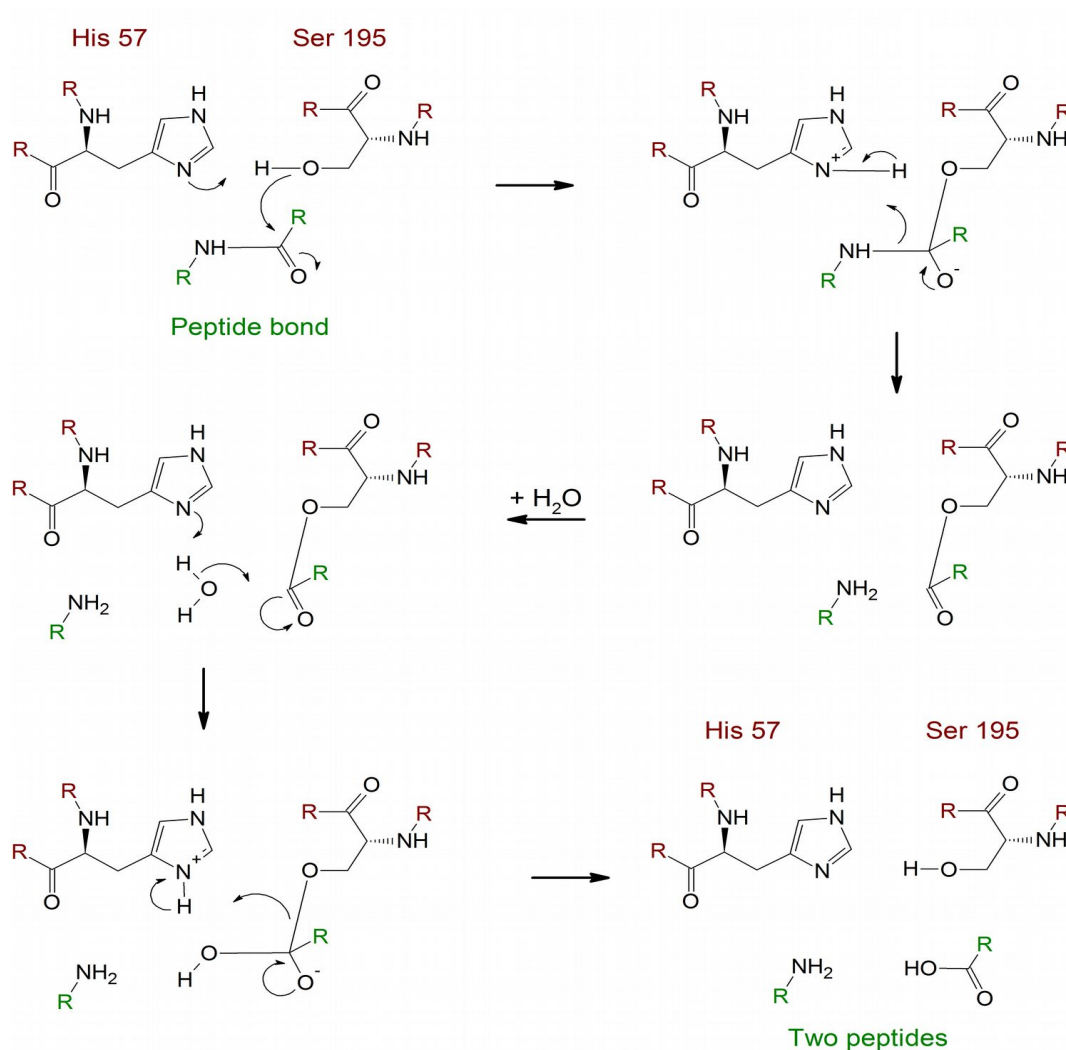


Figure III.5. Reaction mechanism of proteolysis with a serine protease (e.g. trypsin). Asp 102 forms a low-barrier hydrogen bond with His 57 upon binding of the target protein (not shown).

III.2.3 Clean-up and preparation for detection method

Before the measurement with HPLC-DAD, MALDI-TOF-MS or HPLC-ESI-MS/MS, the samples have to be prepared accordingly. Some of the involved research groups apply clean-up steps in the form of miniaturised solid phase extraction (SPE) to avoid soluble paint components or thwarting reagents from previous pretreatment steps from entering the instrument.

In our own experiments the following clean-up steps are applied:

- No SPE is done before HPLC-DAD experiments (chapter IV). Instead salts and soluble paint components are removed online using a trap column (section III.3, page 59).
- In our MALDI-TOF-MS experiments (chapter V), soluble components in the paint samples that potentially can overwhelm the mass spectra, are removed by miniaturised SPE reversed-phase C18 pipette tips (ZipTip, Millipore). A minimal quantity of C18 stationary phase is immobilised on the interior of the tips. The tips are used in accordance with the suppliers instructions. The other two sample preparation protocols used for MALDI-TOF-MS analysis by Kirby(82) and Tokarski(85) do not specify any kind of sample clean-up.
- In addition to the soluble paint components, the large concentrations of urea, thiourea and SDS in our HPLC-ESI-QTOF-MS/MS method (chapters VI and VII) would seriously hamper the detection. Particularly SDS has been observed to irreversibly bind and even jam the trap column. Common reversed-phase solid phase extraction (SPE), e.g. C18, did not remove the SDS. Strong cation exchange (SCX) SPE did remove the SDS, but inserted KCl salts instead. An alternative is the less common hydrophilic interaction liquid chromatography (HILIC) SPE. In HILIC, the mobile phase is mainly an aprotic solvent, most often acetonitrile, mixed with a small amount of water. As a stationary phase almost all polar phases can be used. It is believed that two separate mechanisms amplify retention(114,115): the polar analyte is adsorbed in the water-rich layer covering the polar stationary phase (liquid-liquid extraction), while the polar analyte undergoes cation exchange with the stationary phase. This way, peptides are retained, while salts and SDS are washed away. Moreover, the ammonium acetate used to control the pH is volatile and was removed upon evaporation to dryness. None of the other published(85,87) HPLC-ESI-QTOF-MS/MS methods uses non-ionic detergents such as SDS that significantly improve the digestion yield on hydrophobic proteins (e.g. egg yolk and certain milk proteins). Therefore, these methods do not require SPE to remove those, at the cost of a lower

yield. Tokarski tries to overcome this by using a grinding resin in the preparatory steps, which however requires to be removed afterwards by extraction.

Finally, although sometimes poorly documented in literature, the sample needs to be adapted to the chosen analytical instrument upon introduction. For MALDI-TOF-MS a suitable matrix, such as α -cyano-4-hydroxycinnamic acid (CHCA)(82) or, as in our experiments (chapter V) 2,5-dihydroxybenzoic acid (DHB), is added, the resulting solution is dropped onto a target plate and let to dry. For the chromatographic methods, HPLC-DAD and HPLC-ESI-QTOF-MS/MS, the sample solution is evaporated upon dryness and reconstituted in a suitable solvent.

III.3 Peptide chromatogram fingerprinting using HPLC-DAD

III.3.1 Principles

Trypsin proteolysis essentially creates a complex, water-soluble mixture of tryptic peptides from all proteins in the original paint sample. Liquid chromatography, e.g. based on a reversed-phase stationary phase, is a proficient tool to separate the tryptic peptides according to their polarity. A suitable detector is the DAD (also known as photodiode array detector, PDA), which detects the UV and visible light absorption of the eluent throughout the entire run.

Every peak in a chromatogram corresponds to either an eluting tryptic peptide or a contaminant. The UV-Vis absorption spectrum of a typical peptide is characterised by two bands: one at 214 nm that can be attributed to the peptide bond and one at 280 nm, attributed to aromatic side chains. Although the relative intensity of these two bands gives some insight in the number of aromatic amino acids in a peptide, determination of the amino acid sequence of the peptides using this technique is not possible. The resulting chromatograms act as a fingerprint for a certain proteinaceous paint binder, which can be identified by comparison with the chromatogram library of reference samples.

Retention time accuracy is essential in performing chromatogram comparison. Therefore, any retention shift obstructs identification; an internal standard should therefore be considered. The advantages of using HPLC-DAD for peptide analysis are the straightforward procedure, easy interpretation and the low cost of these instruments. Most conservation science labs have access to suitable instruments.

III.3.2 Experimental set-up

In the experiments described in chapter IV we used the SpectraSystem HPLC system (Thermo Fisher) at the Royal Institute for Cultural Heritage. This instrument consists of a P1000XR quaternary pump, an AS3000 autosampler with a 20 μL sample loop and a UV6000LP UV/Vis DAD detector equipped with a 50 mm detector cell. We used the lowest flow rate the quaternary pump could handle ($200 \mu\text{L min}^{-1}$) to minimise sample consumption. The HPLC eluents were (A) 0.01% formic acid in milli-Q water and (B) 0.01% formic acid in acetonitrile; the HPLC was programmed to gradually increase solvent B from 10% to 50% in 60 min.

To prevent salts and soluble paint components from entering the system a trap column (Phenomenex Security Guard Max-RP C12, 4.0x2 mm) was placed between the autosampler and the analytical column. This trap column was replaced regularly. The analytical column was a Phenomenex Jupiter Proteo C12, 250x2.0 mm, with a pore size of 9.0 nm and a particle size of 4 μm . Both columns were thermostatted at 25°C in a column heater to improve retention time accuracy. The slit opening of the DAD detector was set at 11 nm for discrete wavelength scans at 214 nm and 280 nm and at 5 nm for complete scans (200 - 400 nm).

III.4 Peptide mass fingerprinting using MALDI-TOF-MS

III.4.1 Principle of MALDI-TOF-MS

MALDI is an ionisation technique of which the exact working mechanism (figure III.6) is not yet fully understood. The peptide solution is mixed with a large excess of matrix that strongly absorbs the laser energy. After co-crystallisation of this solution on a MALDI target plate, a pulsed UV laser beam induces desorption of

the matrix containing the peptides. The peptides are thought to be protonated (soft ionisation, $(M+H)^+$) inside the plume of neutral and ionised matrix molecules and peptides. These protonated peptides are subsequently accelerated by a high-voltage electromagnetic field towards the TOF-MS. MALDI has a tendency to create single-charged ions. (74)

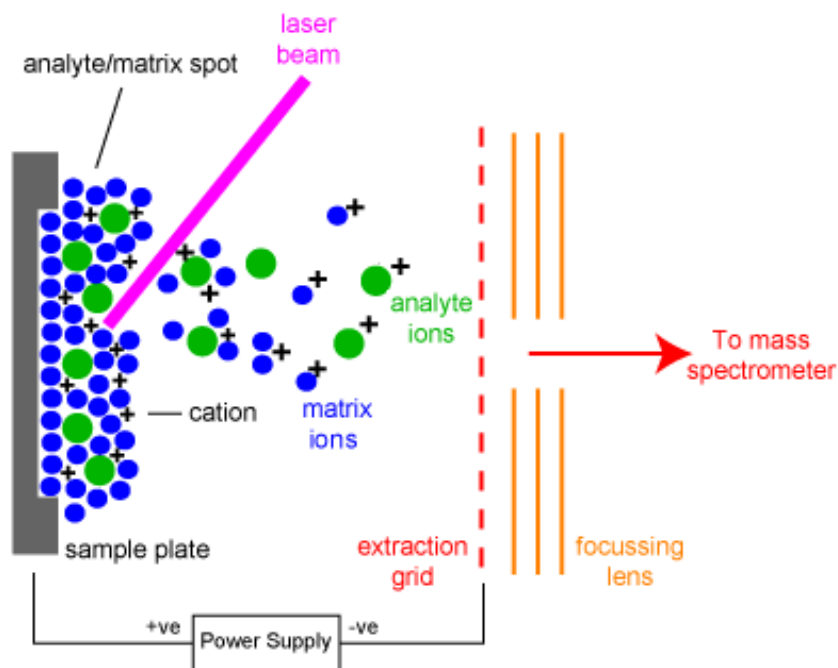


Figure III.6. A schematic of matrix-assisted laser desorption/ionisation. (source: adapted from Gates(116))

This short period of acceleration induces a separation of the ions based on their moment of inertia, or thus their mass. Once passed the lens, inside the TOF-MS (figure III.7), the ions enter a field-free drift range in which all ions continue their path with the speed they had reached before: lighter ions travel faster than heavier molecular ions. The length of the drift range defines the mass resolution of the instrument. In reflectron mode, all ions are accelerated in the reversed direction towards the dynode detector. (74)

When no chromatographic separation is used, all components (including contaminants) are measured simultaneously and typically give rise to complex and noisy mass spectra. Each component, such as a peptide, is represented by a series of peaks representing its isotopic distribution. The only information that is obtained

from this peptide is the measured mass-to-charge ratio (m/z). Due to the tendency of MALDI to create single-charged ions ($z = 1$), this is likely equal to $(M+H)^+$.

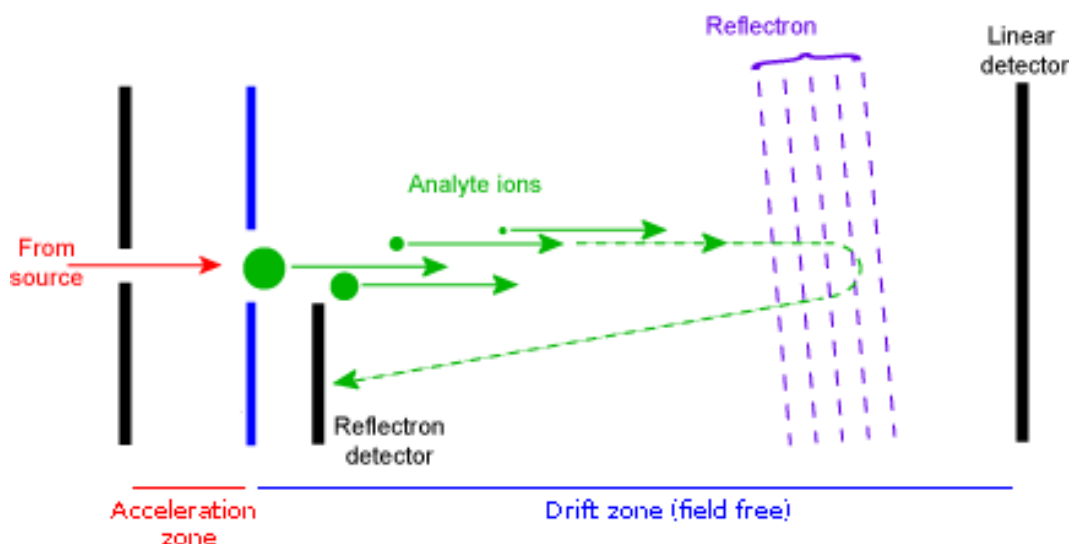


Figure III.7. A schematic of a time-of-flight mass spectrometer operating in reflectron mode. (source: adapted from Gates(116))

The determination of the mono-isotopic mass of a peptide is by itself not a solid characterisation, as different components might share the same mass. However, the whole of peptides that originates from a certain protein, or a mixture of proteins as is the case in paint binders, is characteristic. Peptide mass fingerprinting (PMF) is the technique where all experimental masses are compared with known masses. In proteomics, these known masses are often calculated based on the masses of predicted tryptic peptides, which in their turn are based on known protein sequences. This can be automated by software packages, such as Mascot PMF search (Matrix Science). Unfortunately, this approach is only applicable for pure and single proteins – not for protein binders that are mixtures of different proteins. A variant of PMF is based on the comparison of mass spectra with those obtained from reference samples. The marker peaks that are obtained this way can then be identified with their sequence by, amongst others, the online tool FindMod (ExPASy) (117). Evidently, this approach is limited to known and studied substances and becomes hard in the case of complex mixtures, contaminations or degraded samples.

III.4.2 Experimental set-up

In the experiments in chapter V we used the Biflex IV MALDI-TOF-MS (Bruker Daltonics) at the Institute of Chemical Technology (Prague). This system is equipped with a pulsed nitrogen laser (337 nm) and was used in positive reflectron mode to acquire the mass spectra. To enhance reproducibility, at least 200 laser shots on different locations of the sample were collected and accumulated.

The data files were converted into readable ASCII format using mMass(118–120), a freely available open source tool. The mass spectra were uniformised in a batch operation using a Python script: all measurements are limited to the same mass range, the number of datapoints is reduced and finally the spectra are normalised. The Unscrambler X.1 (Camo Software) was used for principal component analysis (PCA) and soft independent modelling by class analogy (SIMCA). The calculation of the principal components is based on the pre-processed paint model data and the series of different animal glues; afterwards the coordinates in the reduced dimensions are calculated for the unknowns, which are thus not taken into account for the factor axis calculations.

III.5 Peptide identification using HPLC-ESI-QTOF-MS/MS

III.5.1 Measurement

Nano-HPLC

The typical set-up of HPLC in combination with mass spectrometry usually differs from the one with other types of detectors. The (nano-)ESI interface cannot handle the large flow rates of traditional HPLC. To overcome this, the high flow is sometimes reduced using a splitter between the column (or a non-destructive traditional detector) and the ESI interface. Depending on the flow rate and the ESI interface, typically 5 to 10% of the total flow is directed towards the ESI interface. This implies, however, a loss of 90 to 95% of each of the tryptic peptides: large paint samples are required and low sensitivity is obtained.

Nano-HPLC is an improved set-up (figure III.8): first the tryptic peptides are trapped using a normal flow rate on small trap column and subsequently eluted with a

nano-flow onto a capillary column, where they are separated. Almost all peptides that were injected in the HPLC also reach the ESI-MS. This results in a significantly higher sensitivity, while minimizing the paint sample quantity.

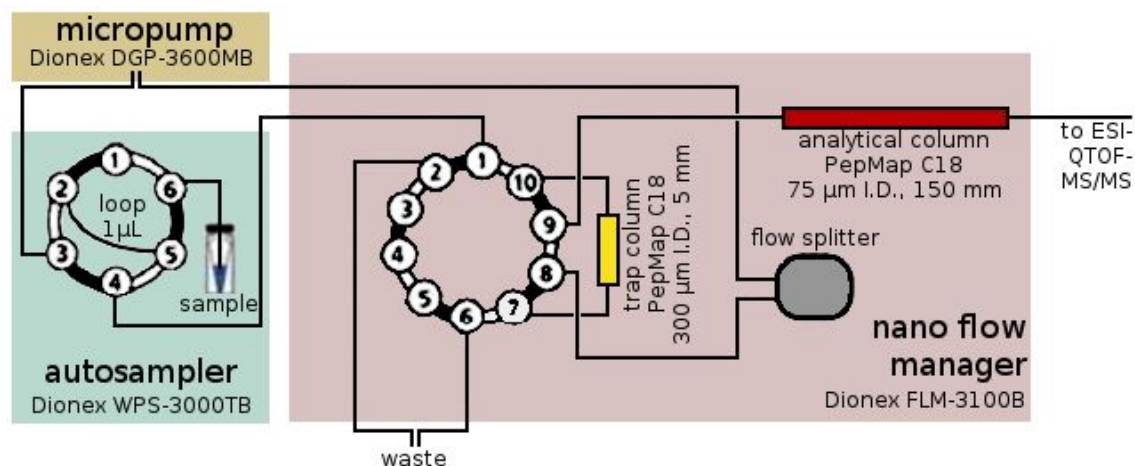


Figure III.8. Diagram of the nano-HPLC set-up.

The system used in chapters VI and VII is the Ultimate 3000 Dual nano-HPLC system (Dionex) at the Laboratory of Pharmaceutical Biotechnology of Ghent University. This system consists of a DGP-3600MB dual ternary low-pressure proportioning micro pump, a WPS-3000TB nano pulled-loop thermostatted autosampler and a FLM-3100B Nano flow manager. The system is controlled via the Chromeleon software package. The solvents used are (A) 0.1% aqueous formic acid and (B) 0.1% formic acid in 80:20 acetonitrile/water. All solvents are LC/MS grade supplied by Biosolve. 20 µL of the sample solution was injected, purified and concentrated in a PepMap C18 trap column (5 µm, 100 Å, 300 µm I.D. x 5 mm from LC Packings) using a flow of solvent A. Separation of the peptides took place in a PepMap C18 capillary column (3 µm, 100 Å, 75 µm I.D. x 150 mm from LC Packings) using a linear 45 min gradient at 300 nL min⁻¹ in which the solvent B increased from 4% till 100%. (121)

ESI-QTOF-MS/MS

Like MALDI, ESI is a soft ionisation technique in which the peptides are protonated. The latter, however, has a tendency to create multiple positively charged peptides. Another key difference between the two ionisation techniques is the physical state in which the analyte enters the system: whereas for MALDI a crystallised sample is

required, ESI ionises the analytes in solution (figure III.9 A). This renders this technique ideal for online coupling with nano-HPLC systems, allowing the tryptic peptide mixtures to be separated. MALDI is therefore preferred for relatively simple peptide mixtures (e.g. tryptic digest of a single protein), whereas HPLC-ESI-MS systems are used for complex samples.

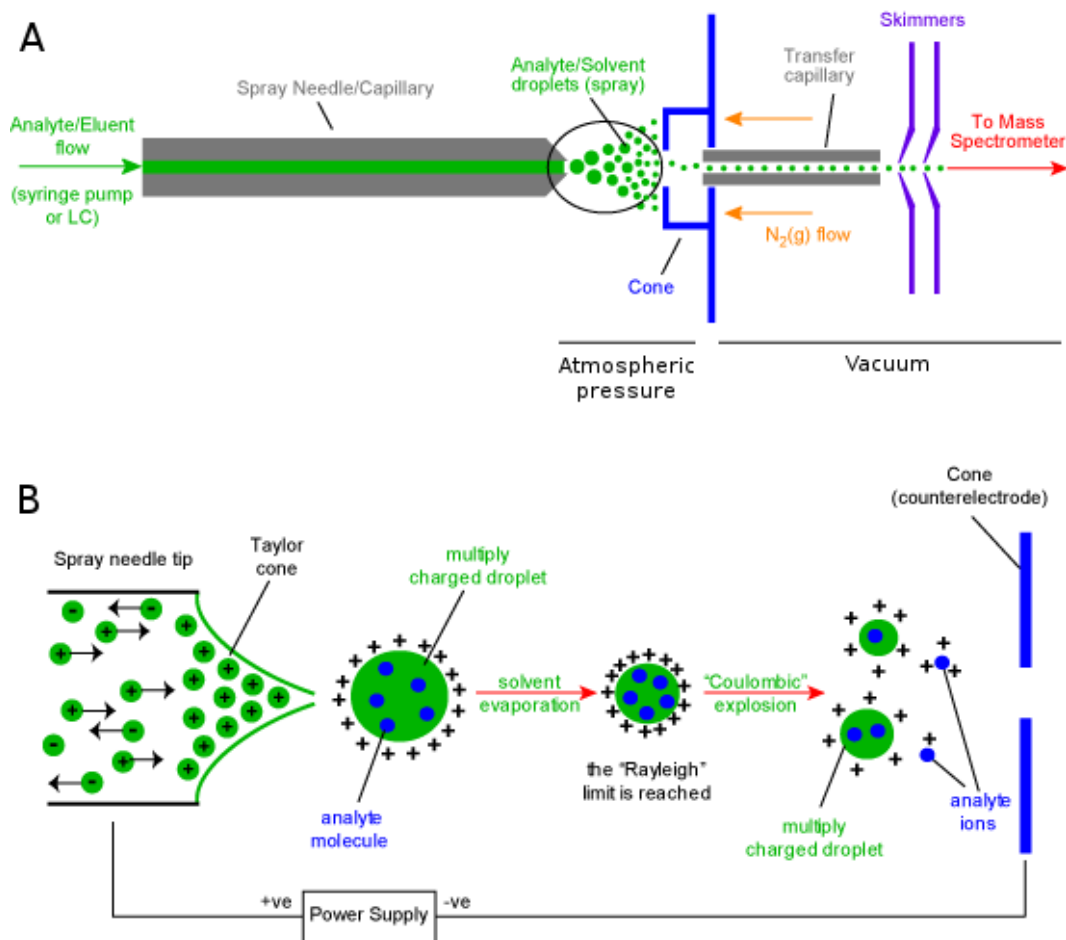


Figure III.9. A schematic of electrospray ionisation: (A) general overview of the interface, (B) mechanism of ion formation. (source: adapted from Gates(116))

ESI is based on the evaporation and fission of highly charged droplets in a nebulised sample liquid (figure III.9 B). Arriving from the HPLC, the solution containing the analytes is passed through a capillar to which a high voltage is applied. The highly charged solution nebulises upon leaving the capillar following the Taylor cone phenomenon(122). During the subsequent pressure and potential gradient, aided by a nitrogen flow, the highly charged droplets reduce in size by evaporation of the solvent. During this phenomenon, the charge density increases

until it reaches the Rayleigh limit and droplet fission occurs (Coulomb explosion), further reducing the droplet size. This evaporation and fission cycle is repeated until ultimately only fully desolvated ions are left. (123)

The ESI interface can be used in conjunction with almost all types of single and tandem mass spectrometers. In the experiments in chapters VI and VII tandem MS was used: whereas single mass spectrometry only determines the peptide mass, tandem MS gives much more information on the peptide structure. Four types of tandem MS instruments are nowadays frequently used in proteomics: (linear) ion trap, orbitrap, TOF-TOF (the combination of two TOF sections), QTOF (the combination of a quadrupole and TOF) and Fourier-transform ion cyclotron resonance (FTICR). QTOF, which is used in most applications in conservation science, offers a high sensitivity, mass accuracy and resolution. (74)

During this doctoral research project two similar QTOF instruments were used (chapters VI and VII): the reference samples and most historical samples were measured with an Ultima instrument, while a few other historical samples were run on a next-generation Premier system (both by Waters Corporation). The ground plan of the Premier instrument is shown in figure III.10; the main difference with the older generation Ultima set-up is the replacement of two quadrupoles by T-wave parts as ion guide and collision cell. This enhances sensitivity, selectivity and analytical speed.

In MS mode, the main (middle) quadrupole acts merely as an ion guide to the TOF analyser, where the mass analysis takes place. In the tandem MS mode, the precursor ions are selected in the quadrupole and undergo fragmentation through collision-induced dissociation (CID) in the collision cell. During CID the ions collide with neutral particles such as argon. Herein, some of the kinetic energy is converted into internal energy which results in bond breaking and the fragmentation of the peptide ion into smaller fragments. These product ions are analysed in the TOF device. (121)

The measurements were recorded in data-directed acquisition mode. Herein a single MS survey is run during the whole measurement run; for each significant mass detected, ideally an eluting peptide, the instrument automatically switches to

MS/MS mode: the peptide ion is isolated, fragmented and detected. The MS survey mass range was set from 500 Da to 1150 Da. A maximum of six simultaneous MS/MS acquisitions were run with a mass range from 50 Da to 1998 Da. The precursor charge state selection for MS/MS analysis was set at $2^+ - 3^+$ with a minimum precursor total ion count of 50. Data acquisition was done using MassLynx (Waters).

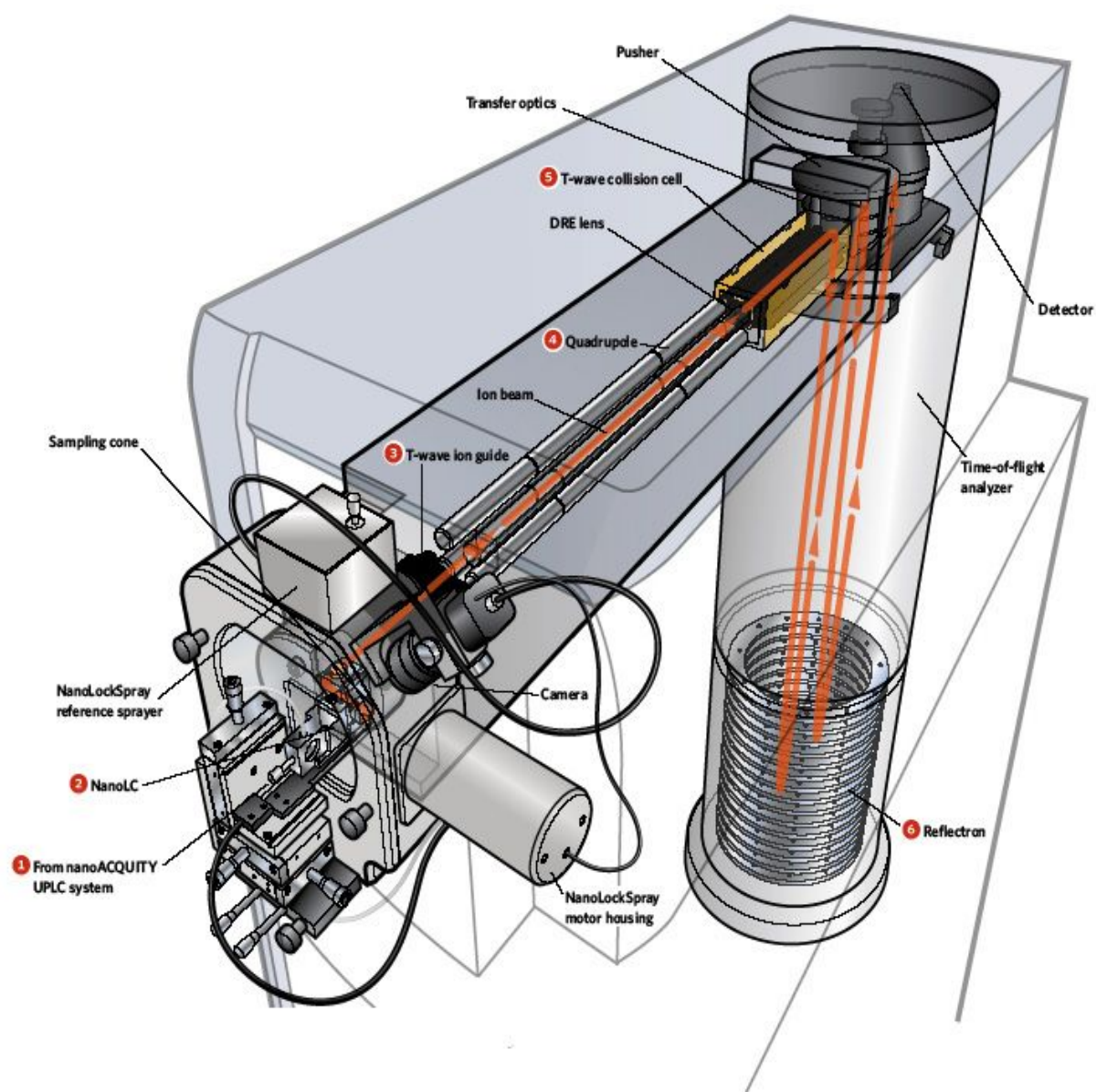


Figure III.10. Waters QTOF Premier. (source: Constans(124))

Each component that elutes from the HPLC is detected in single MS mode and subsequently automatically isolated (in the quadrupole), fragmented (in the

collision cell) and detected (in the TOF). These fragments are characteristic for the precursor peptide. The fragmentation processes are yet not fully understood, but seem to follow certain rules, depending on, amongst others, the type of instrument: for QTOF instruments cleavage happens mostly along the peptide bond, creating so called y and b ions (figure III.11)(125). The result is a tandem mass spectrum in which the sequence can be “read” (figure III.12): the b_1 ion peak corresponds to the mass of the first (protonated) amino acid of the peptide sequence, the mass difference between the b_1 and b_2 ion peaks equals the mass of the second amino acid, etc. This mass spectrum is a unique fingerprint for the tryptic peptide, substantially enlarging the “certainty” of its parent protein (binder) identification.

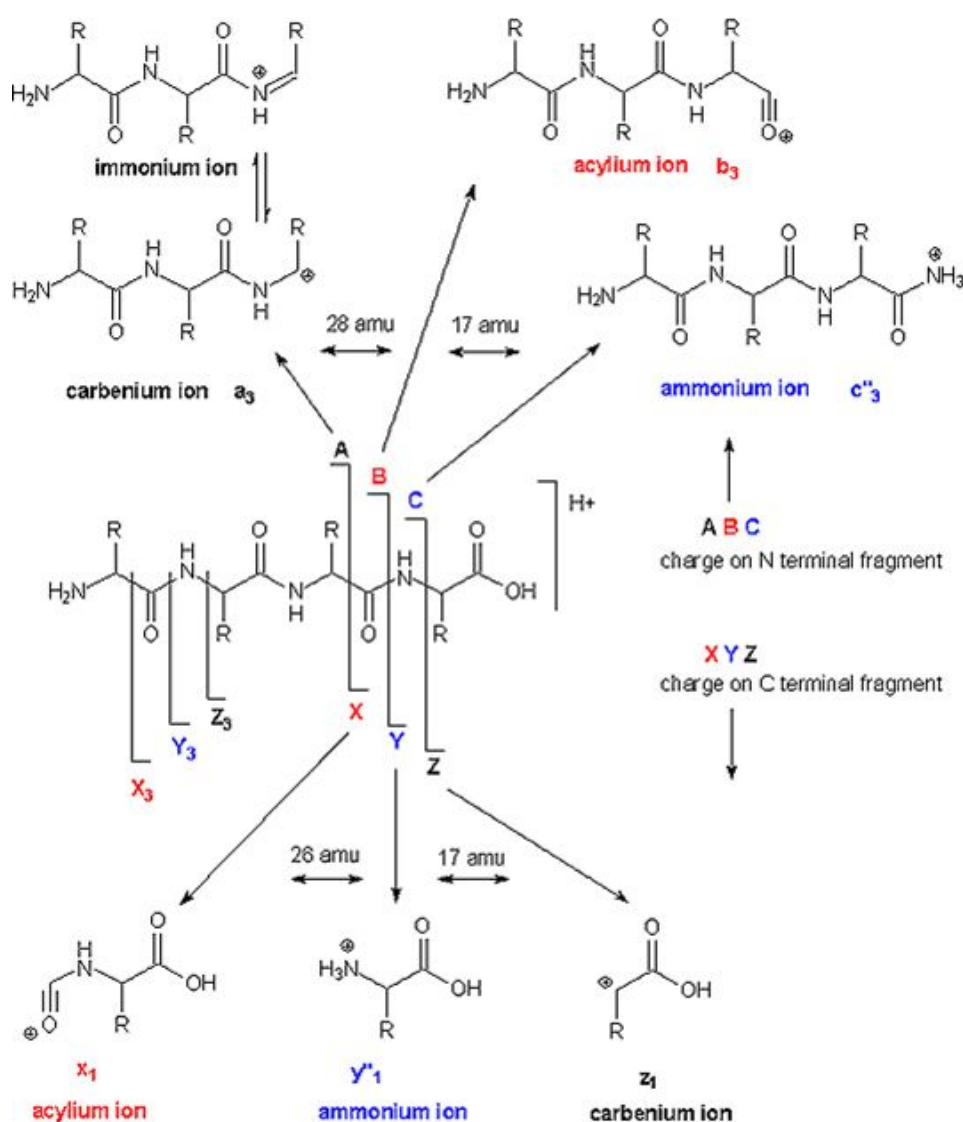


Figure III.11. Peptide fragmentation in tandem mass spectrometry.

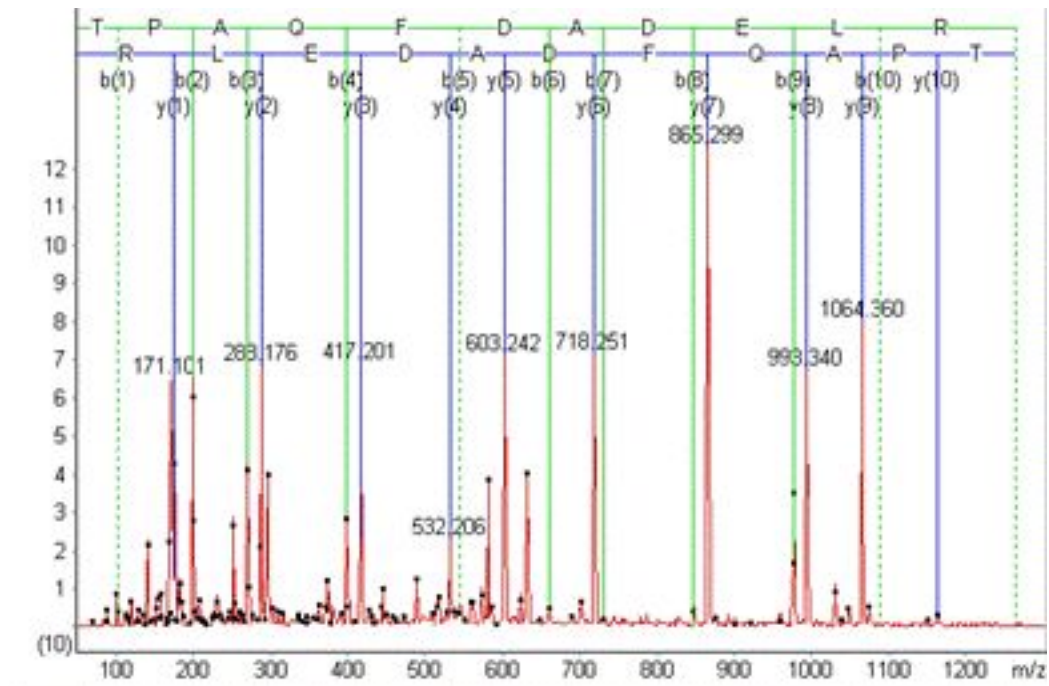


Figure III.12. Tandem mass spectrum of the peptide ion TPAQFDAELR. The green lines show the b ions, the blue lines the y ions (and the sequence in reverse order).

III.5.2 Data treatment

Datasets of a single measurement containing several hundreds of tandem mass spectra are common. Powerful software packages are needed to work through these vast amounts of data. Two different pathways for the identification of the tryptic peptides (and thus the proteins) have been developed (figure III.13): sequence library versus spectral library search approaches. Beside these, two other techniques aiding in the interpretation of the data are discussed as well: *de novo* sequencing and BLAST. In the following paragraphs we will discuss these approaches in detail.

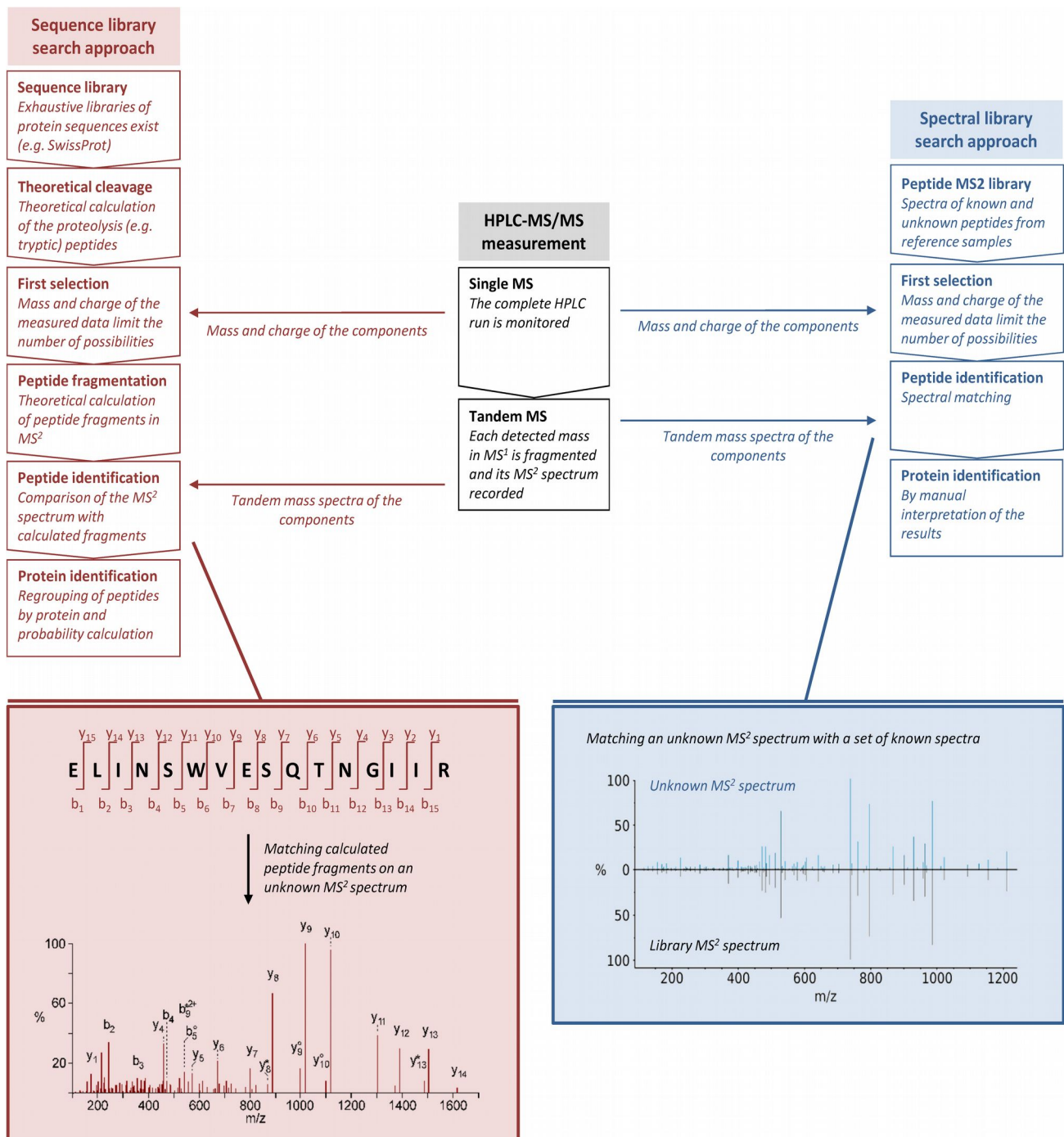


Figure III. 13. Comparison of the sequence library search and the spectral library search approaches.

Sequence library search

A first approach tries to link tandem mass spectra with the protein sequences that are available in sequence libraries. Many libraries exist, including a few very large

ones that are maintained by large biochemistry institutions and consortia and that can be freely consulted on the internet. Examples are SwissProt, Trembl (both maintained by UniProt(126)) and the library of the National Center for Biotechnology Information (NCBI)(127). These libraries, containing hundreds of thousands sequences, or proteins, from many different animals, form the starting point for specialised search programs such as Mascot MS/MS Ions Search (Matrix Science), SEQUEST (Thermo Fisher) and X!Tandem (The Global Proteome Machine Organization, GPM). In this work a locally installed version of Mascot was used.

- Mascot requires input of experimental and search parameters that are necessary for its calculations. It needs to know which protease has been used and which post-translational modifications have to be taken into account. These can be fixed (the modification is expected on all possible instances) or variable (may or may not be present). Another option is the “error-tolerant search”, in which all modifications are considered. The instrument type that has been used has to be inserted, due to its influence on fragmentation. A sequence library should be chosen, with eventual restrictions on e.g. taxonomy. Also mass tolerances, both for peptides and for fragments, peptide charges to be considered and the maximum number of missed cleavages should be inserted in the software.
- Based on the cleavage rules of the selected protease, the algorithm calculates all possible peptides for all protein sequences available in the sequence library. The calculated masses, taking the selected post-translational modifications and peptide charges into account, are compared with the masses observed in single MS mode. For each of the observed components in the measurement, a list of possible peptide annotations is thus composed, based solely on the precursor mass and charge.
- For each of the detected components, a similarity score (MOlecular Weight SEarch score or MOWSE score(128)) is attributed to each of the possible peptide annotations based on the tandem MS spectrum. It uses a statistical model that calculates the probability of a number of matching peaks by random chance. It does so by assigning a probability value to each

matched m/z peak using a training set of sequences and multiplying all probability values to compute the composite probability P , which for convenience is expressed as $-10 \log P$. When this MOWSE score is above a limit value, the corresponding peptide annotation is considered as a "hit". Note that MOWSE and Mascot's derived ion scores are not real probabilities, since there is no requirement that a higher score for one sequence reduces the score for other sequences(129).

- Mascot regroups all hits per protein. Based on the individual MOWSE scores of the individual peptides, a protein score is calculated. Considering that the same peptides may occur in different proteins (for example corresponding proteins from different animals) in the sequence library, the Mascot search report shows these different proteins.

Since different proteins are determined independently from each other, sequence library algorithms are able to process measurements of complex mixtures of protein paint binders, samples contaminated with keratin or saliva, or samples consisting of multiple protein containing layers. The approach also benefits from the vast and fast-growing number of sequenced proteins from an increasing number of organisms. Determination of unexpected protein sources therefore becomes feasible. Library search algorithms can only assign a sequence to a spectrum if that sequence is present in the database and will fail to identify a spectrum derived from a sequence that is not. Unfortunately, many more proteins, mainly of relatively low interest to biochemistry research, are still lacking, some of which are of particular interest to conservation science. In these cases, Mascot allows at best to identify those peptides that are common (evolutionarily conserved) with a protein of a closely related species. Non-conserved peptides, on the other hand, simply remain ignored by Mascot. This leads to a false identification. Difficulties also exist with evolutionarily well conserved proteins, such as collagen and with samples containing mixtures of corresponding proteins from different species.

Spectral library search

More recently a different, more straightforward method for peptide identification is implemented, based on spectral matching of a tandem mass spectrum of an

unknown peptide with a library of spectra(130,131). Several proteomics research groups are developing these spectral libraries and specialised search software. Well-known examples in this area are the National Institute for Standards and Technology (NIST)(132), the Global Protein Machine (GPM)(133) and the Institute for Systems Biology (ISB)(134) providing both libraries and software (MS search and MSepSearch by NIST, X!Hunter by GPM, SpectraST by ISB).

Compared to sequence library searching, the concept of spectral library searching is easier and promises a more discriminative precision through direct comparison of spectra. The use of real library spectra for matching is the major advantage of spectral library searching, since it takes into account real peak intensities and non-canonical fragments (peptide fragmentation in tandem spectroscopy is not fully understood), both under-utilised information in sequence database searching(135). Spectral library searching is also disproportionately more successful in identifying low-quality and complex spectra(135). Finally, spectrum library searching calculations are up to 1000 times faster. The spectral library search for peptides was only developed recently due to the lack of libraries and the difficulties to build them, while sequence libraries were already available. Nowadays, spectral library building can rely on a fast-growing number of confidently identified MS/MS spectra. It is therefore not (yet) a replacement for the sequence library search approach, but rather a complementary technique. (131)

A fast-growing number of spectral libraries is available on the internet, but they typically focus on those proteins that are important in current day proteomics. The proteins encountered in conservation science are largely lacking. Regardless of the available libraries, most software packages including NIST Search MS that was used in chapter VII allow to make dedicated spectral libraries.

De novo sequencing

If a protein from a certain species is missing in the sequence libraries, its tryptic peptides that are unique for that species cannot be annotated by sequence library search. Without prior knowledge of the amino acid sequence, the mass difference between the peaks of a spectrum is studied to identify the different series of fragments (for QTOF the main series are the b and y fragments). The mass

differences between to subsequent peaks of the same series allows us to determine the amino acid and its place in the peptide sequence. This time-consuming process is called *de novo* sequencing and is usually calculated by software packages (e.g. Mascot Distiller by Matrix Science). It is known to be error-prone: *de novo* sequencing does not allow to distinguish between amino acids with the same mass (e.g. leucine and isoleucine). Also, post-translational modifications seriously hamper the analysis.

De novo sequencing was not used in this work, but could eventually be a useful tool to annotate tandem mass spectra in the spectral library that could not be annotated by sequence library search. Due to its error-proneness, *de novo* sequenced peptides should be checked for similarity with corresponding evolutionarily non-conserved peptides in other species.

BLAST

Even though strictly not an approach to annotate unknown peptide spectra, the basic local alignment search tool (BLAST) deserves attention as a useful data treatment tool in peptide analysis. BLAST is a software algorithm for comparing a sequence with a sequence library. When supplied with a peptide sequence, BLAST outputs a list of proteins in which this peptide or similar peptides are encountered; a similarity value is given and differences between the two sequences are shown. BLAST is therefore vital in the determination of the species or to verify *de novo* sequenced peptide annotations. Online BLAST interfaces exist for all major protein sequence libraries (UniProt, NCBI...).

III.6 Chapter epilogue

This chapter illustrated the technical background, the advantages and disadvantages of the most frequently used methods for proteins in works of art. Three proteomics methods are discussed in depth: peptide fingerprinting using HPLC-DAD (chapter IV), peptide mass fingerprinting with MALDI-TOF-MS (chapter V) and peptide identification (chapter VI). For each of these, a detailed review of the sample preparation and the specific choices for the three detection methods was given. Finally the expected results of these methods and the tools used to interpret

these data were discussed. In the following chapters, these methods will be thoroughly tested and applied on paint models and micro-scrapings of real works of art.

IV

IDENTIFICATION OF PROTEIN BINDERS IN WORKS OF ART BY HIGH- PERFORMANCE LIQUID CHROMATOGRAPHY DIODE ARRAY DETECTOR ANALYSIS OF THEIR TRYPTIC DIGESTS

*Based on the scientific paper:
W. Fremout, J. Sanyova, S. Saverwyns, P. Vandenabeele, L. Moens,
Anal. Bioanal. Chem. 393 (2009) 1991-1999.*

The principles and potential advantages of peptide analysis were discussed in depth in the previous chapter. Enzymatic cleavage of the protein matter in paint samples results in peptides that are small enough for precise and sensitive detection, yet big enough to be characteristic for the proteins from which they originate. In this chapter the first, most basic iteration of a protocol for peptide analysis is elaborated. Here the tryptic peptides are measured with high-performance liquid chromatography (HPLC) coupled to UV diode array detection (DAD), instrumentation which is available in most conservation science labs. All reaction steps are performed in the same vial; no extraction method nor sample transfer is needed, reducing the risk of sample losses. A collection of pure binders, paint models and historical paint samples have been investigated with this method. Chromatograms of unknowns recorded at 214 nm and 280 nm are compared with

those of the reference samples as a fingerprint. The results are comparable with the results of other techniques used for binder identification on the same samples, with the differentiation between egg yolk and glair as an added value.

IV.1 Introduction

Understanding the materials, such as binding media, used by artists is of the utmost importance in conservation science and art history. Diverse proteinaceous materials have been used as binding media in paint, such as animal glues prepared from animal skin or bones (contains several sorts of collagen), egg white (contains ovalbumin, ovotransferrin, lysozyme...), egg yolk (contains vitellogenin...) and milk-based binder (contains several types of casein).

Most actual methods(136) for protein binder identification are based on complete hydrolysis of the protein into its constitutive amino acids (amino acid analysis, AAA) followed by derivatisation and separation/detection with gas chromatography coupled to mass spectrometry (GC-MS)(26,28,30,92,93,95,96,137–139) or with HPLC coupled to a fluorescence detector (HPLC-FD)(27,29,30). Identification of the protein is based on the relative amounts of amino acids present(27,28,93) or using a chemometrical approach, such as principal component analysis (PCA) (30,95,96,140). However, much information about the precise nature of the protein binder, its degradation and the pigment-binder interactions is lost in this highly destructive hydrolysis step. In the case of mixtures of binding media, identification becomes cumbersome.

Other techniques that are used for the identification of binders, such as staining techniques(141), Fourier-transform infrared spectroscopy (FTIR)(142), near-infrared spectroscopy (NIRS)(56) and micro-Raman spectroscopy (MRS)(57–59) are able to discriminate between oils, proteins and other binder classes, but are unable to discriminate between the different proteins. Recently there is a renewed interest in immunofluorescence techniques(63,64) in which labelled antibodies are left to interact with the specific antigens of the binding media proteins. Finally, pyrolysis is used to split the macromolecules into very small fragments that are separated and detected by GC-MS. This approach gives rise to complex pyrograms that are

difficult to interpret. Identification of the binders is based on the presence of marker fragments(33,36,143–145).

In proteomics, however, the approach is usually different: proteins are digested enzymatically into peptides with trypsin. This way, amino acid sequences can be studied, which are much more specific for a protein than single amino acids are: identifying a peptide unambiguously identifies the protein to which it belongs. Peptides also retain more structural information about the protein than amino acids, making a study of degradation effects in proteinaceous paint feasible. Two methods can be used in proteomics. Firstly, an approach based on peptide mass fingerprint, in which peptides are analysed using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS). Secondly, a more complete approach based on tandem mass spectrometry including structural information on peptides and proteins (exact sequence, modifications of the peptides and proteins). Generally, peptides are analysed and fragmented using a nano-HPLC, coupled to a tandem mass spectrometer equipped with nano-ESI. Although well-established routine techniques in proteomics exist, only limited results in conservation science have recently been published: egg temperas have been studied to some extent with the tandem MS approach(85) and the peptide mass fingerprint approach, using MALDI-TOF-MS, has been successfully developed for protein binder identification(75–77,146).

In this study, a DAD (or photodiode array, PDA), which is quite often available in the laboratories of art conservation institutes, is proposed as a cheap and straightforward alternative for the detection of tryptic peptides of protein binders. The novel method is developed using pure protein standards. Different sample preparation methods are tested. In a second phase, the method is tested with both reference paint model samples and real samples from historical works of art. The results are compared with those obtained by amino acid analysis using either GC-MS or HPLC-FD.

IV.2 Experimental

IV.2.1 Reagents

For sample preparation, ammonium bicarbonate, dithiotreitol (DTT) and iodoacetamide (IAA) were supplied by Pierce (Perbio Science, Belgium). Sequencing-grade modified trypsin (Promega, The Netherlands) was used for the digestion. The HPLC eluents, formic acid 99% LC-MS grade and acetonitrile HPLC grade, were acquired from Biosolve (The Netherlands). The deionised water, used for sample preparation and HPLC, was produced with a milli-Q system equipped with a 0.22 µm filter (Millipore, Belgium).

For GC-MS analysis, hydrochloric acid (HCl) solution for amino acid analysis was supplied by Fluka (Sigma-Aldrich, Belgium), *N-tert*-butylmethylsilyl-*N*-methyl-trifluoroacetamide with 1% *t*-butyldimethylchlorosilane (MTBSTFA/TBDMCS) by Grace Davison (Belgium), pyridine hydrochloride 98% by Aldrich (Sigma-Aldrich) and pyridine by Pierce.

IV.2.2 Samples

Standards

Collagen from calf skin (ref. C3511), albumin from chicken egg white (ref. A5503), egg yolk from chicken (ref. E0625) and casein from bovine milk (ref. C7906) were all acquired from Sigma-Aldrich (Belgium).

Paint models

Rabbit skin glue (8% w/w) was allowed to swell in water for 24 h and subsequently heated to 50°C to dissolve completely the glue. Fresh eggs and soft curd cheese (quark cheese) were bought in a local supermarket. The egg glair and yolk were separated. The egg glair was stirred into a stable foam and allowed to stand for several hours. The reliquidised fraction was separated from the foam and was used as egg white binder. The egg yolk was suspended above a recipient and the membrane was cut open. 25 g of soft curd cheese was washed with water on a Büchner funnel to eliminate the soluble gums and sugars. After adding 10 ml water

and 1 g of borax (UCB) to the curds, the mixture was stirred for 40 min until a syrupy solution was formed, which was used as borax casein binder after dilution (1:4). The unaged paint samples, applied on glass slides and stored in a climatised environment in the dark before analysis, were made using the above binders and the same or comparable pigments as found in the historical samples.

Historical paint samples

Samples from three so called pre-Eyckian painted objects were collected. Pre-Eyckian art denominates the Southern Netherlandish works of art preceding Jan van Eyck (early 15th century)(147,148). From the reliquary of the Virgin's Veil (Church of Our Lady, Tongeren; figure IV.1) a black and an orange paint sample were taken from the interior of the chest. From the Crucifixion with St Catherine and St Barbara (Cathedral of the Holy Saviour, Bruges; figure IV.2) a red/brownish sample was taken from the upper edge of the panel. Finally, a sample of the overpainting of the red background from the small St Ursula shrine (St John's Hospital, Bruges, figure IV.4) was collected during its restoration.



Figure IV.1. Reliquary of the Virgin's Veil. (source: KIK/IRPA photo library)



Figure IV.2. Crucifixion with St Catherine and St Barbara. (source: KIK/IRPA photo library)



Figure IV.3. St Ursula Shrine. (source: KIK/IRPA photo library)

IV.2.3 Apparatus and chromatographic conditions

HPLC-DAD for peptide analysis

A SpectraSystem HPLC system (Thermo Scientific, Belgium) was used. It consists of a P1000XR quaternary pump, an AS3000 autosampler with a 20 μ L loop and a

UV6000LP UV/Vis DAD detector equipped with a 50 mm detector cell. The eluents were (A) 0.01% formic acid in milli-Q water and (B) 0.01% formic acid in acetonitrile. The elution program starts from 10% to 50% B in A in 60 min at a flow rate of 200 μ L/min. The analytical column was a narrowbore Phenomenex Jupiter Proteo C12, 250x2.0 mm, with a pore size of 90 Å and a particle size of 4 μ m (Bester, The Netherlands) and the guard column a Phenomenex Security Guard Max-RP C12, 4.0x2 mm (Bester). Both columns were kept at a constant temperature of 25°C in a column heater to improve retention time accuracy. The slit of the DAD detector was set at 11 nm for discrete wavelength scans (at 214 nm and 280 nm) and at 5 nm for complete scans (200 – 400 nm). The wavelength 214 nm was chosen as this is the absorption band of the peptide bonds, while 280 nm is the absorption wavelength of the aromatic side chains.

HPLC-FD for amino acid analysis

The system consists of a HP series 1050 pump and a Waters (Belgium) 470 fluorescence detector (set at 250 nm excitation and 395 nm emission). The analytical column is an AccQ.Tag 3.9x150 mm (Waters), which was kept at 35°C during the analysis.

GC-MS for amino acid analysis

Amino acid analysis using GC-MS is performed on a Finnigan TraceGC Ultra coupled to a Finnigan PolarisQ ion trap MS detector (Interscience, Belgium) and an AT-5ms column (30 m + 5 m retention gap, i.d. 0.25 mm, film thickness 0.25 μ m; Grace Davison, Belgium).

IV.2.4 Analytical procedures

Peptide analysis by HPLC-DAD

The protein standards (100 to 200 μ g) were dissolved in 25.5 μ L 50 mM ammonium bicarbonate buffer in milli-Q water (pH 7.8). 1.5 μ L 100 mM DTT in milli-Q water was added as a reducing agent to irreversibly break the disulphide bonds, unfolding the tertiary structure of the proteins. The samples were incubated at 95°C in a heating block for 5 min. After cooling down, the *SH*-groups of cysteine that were

formed during the reduction step, were irreversibly alkylated into the stable S-carboxyamidomethylcysteine by adding 3 μL freshly prepared 100 mM IAA in milli-Q water. In this alkylation step, the samples were incubated for 15 min at room temperature in the dark. 2 μL of 100 ng μL^{-1} trypsin solution was added and the mixture incubated at 37°C overnight. After evaporation under vacuum, redissolution in 25 μL 0.01% formic acid in milli-Q water and centrifugation (14000xg, 15 min), the samples were ready for HPLC analysis.

For paint samples the procedure is extended: the samples (ca. 200 μg) were first manually ground with a spatula in the sample tubes. After addition of 25.5 μL 50 mM ammonium bicarbonate buffer in milli-Q water (pH 7.8), the samples were rigorously sonicated to enhance dispersion and dissolution of the aged and degraded proteins.

Amino acid analysis by HPLC-FD

For amino acid analysis with HPLC, acid hydrolysis (6 M HCl constant boiling, 1% v/v phenol) is performed in the gas phase on ca. 200 μg of sample. After freezing in liquid air (-170°C) and five evacuation (oil pump) / flushing (N_2) cycles, the sample vessel is put in an oven for hydrolysis (20 h at 110°C). Afterwards the samples are dried in a desiccator and redissolved in 0.1 M HCl. Amino acid analyses are carried out with the AccQ.Tag amino acid analysis method (Waters). A more detailed description of the procedure is found in two publications by Wouters *et al.*(27).

Amino acid analysis by GC-MS

The procedure used here is described in detail by Scott *et al.*(28) and is summarised here. The samples (ca. 200 μg) are first hydrolysed with 6 M HCl (105°C, 24 h) to obtain free amino and fatty acids and glycerol. After some purification steps to remove all acids and water from the samples and a desiccation step (30 min under vacuum), the compounds are silylated with MTBSTFA/TBDMCS in a silylation solvent (30:70; pyridine hydrochloride saturated pyridine) for 30 min at 60°C and 5 h at 105°C. After centrifugation, 1 μL of the mixture is injected into the instrument.

IV.3 Results and discussion

Aged and degraded paint samples are hard to dissolve. While some protocols are described using grinding resins or adding aggressive solvents that need to be removed before the analysis – reaction steps with a potential risk of sample loss, we obtained good results with a simple single-tube method. With the help of a spatula, the sample is ground in a 0.5 mL sample tube. Further sonification of the samples in the digestion buffer disperses and dissolves the sample adequately. Different method variants were examined: improvements were seen when a reduction and an alkylation step were performed before the digestion step. DTT reduces the disulphide bonds, helping to unfold the protein's secondary structure, while IAA prevents refolding by alkylation of the cysteine residues.

IV.3.1 Pure protein samples

The chromatograms of four pure proteins are shown at 214 nm (figure IV.4) and 280 nm (figure IV.5). Collagen (from calf skin, contains different kinds of collagen) only has a weak signal in both the 214 nm and the 280 nm absorption wavelengths, probably caused by its trypsin-resistant triple helix structure(149). In the 280 nm-chromatogram, the peptide peaks are very sparse due to the very low content of aromatic amino acids in collagen (only 2.3% phenylalanine and 1.0% tyrosine(12)). The chromatograms of ovalbumin (from egg white) and yolk differ on several places, proving a significant difference in protein content. This eases the discrimination between the two compared to the amino acid analysis techniques in which the composition of amino acids exhibit none or only few subtle differences(12). The casein sample gives a strong signal and many peptide peaks at both wavelengths.

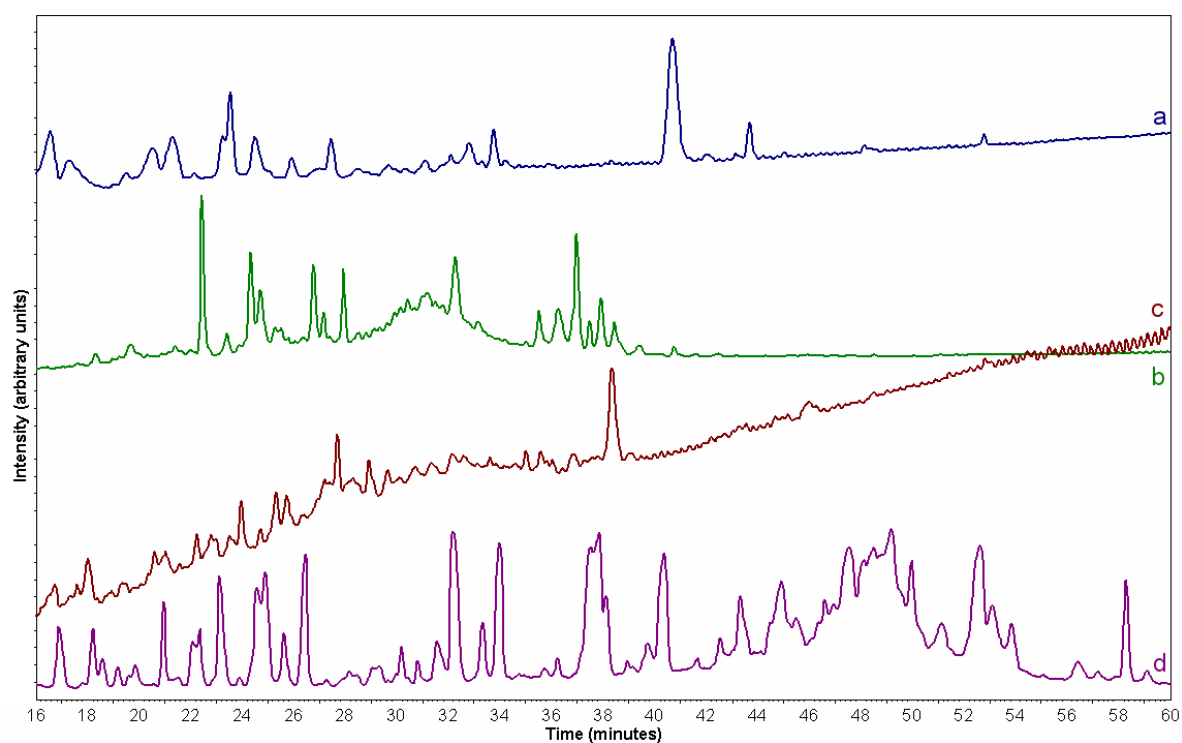


Figure IV.4. Chromatograms at 214 nm of pure protein samples: (a) collagen, (b) ovalbumin, (c) egg yolk and (d) casein.

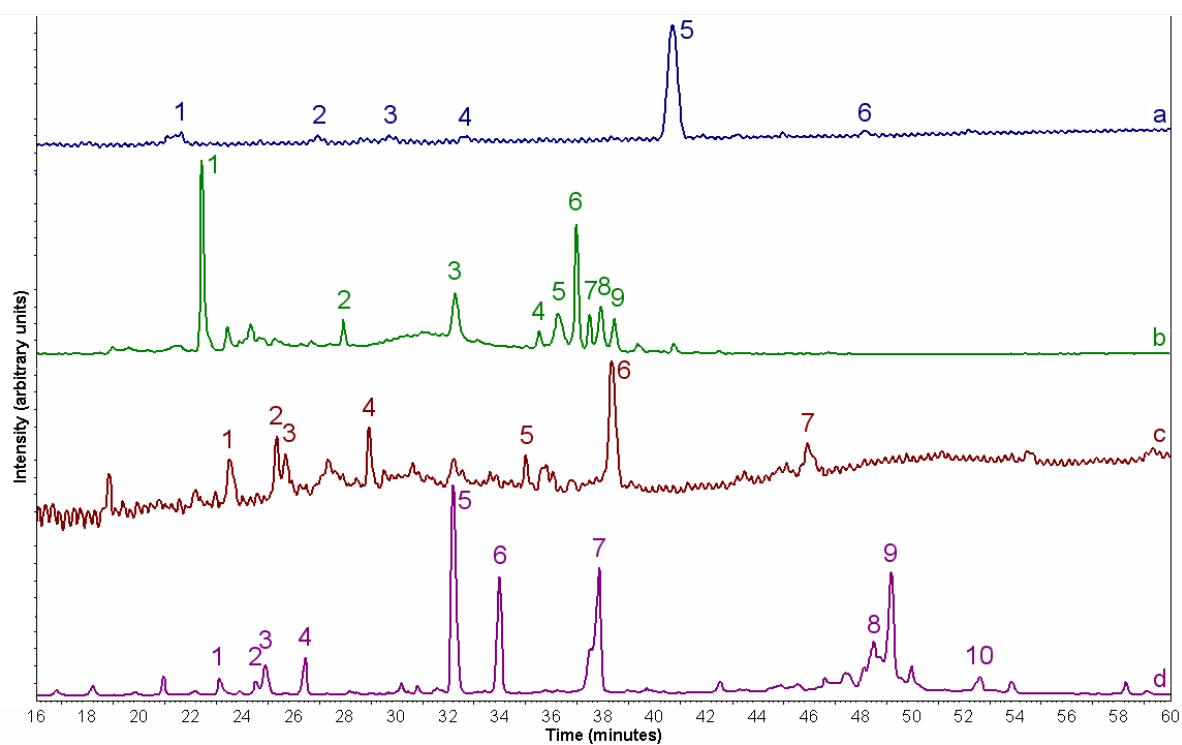


Figure IV.5. Chromatograms at 280 nm of pure protein samples: (a) collagen, (b) ovalbumin, (c) egg yolk and (d) casein.

With the exception of the collagen sample, the chromatograms at 214 nm are unresolved, although sufficient resolved peaks are present for fingerprint identification. In most cases the 280 nm chromatograms are, however, more suitable. The retention factors³ (k) of the components used for identification of the unknowns are listed in table IV.1.

Table IV.1. Retention factors (k) of the peptides used for identification. The peak numbers are shown in figure IV.5.

Peak number	Retention factor (k)			
	Collagen	Ovalbumin	Egg yolk	Casein
1	7.82	7.80	8.61	11.51
2	9.99	9.94	9.38	12.28
3	11.12	11.65	9.50	12.47
4	12.31	12.92	10.83	13.31
5	15.59	13.21	13.31	16.42
6	18.65	13.48	14.68	17.38
7		13.68	17.79	19.48
8		13.86		25.22
9		14.06		25.59
10				27.46

IV.3.2 Paint models and historical paint samples

The results on the historical paint samples are summarised in table IV.2. The attribution of proteins using the peptide analysis method is based on comparison with the chromatograms of paint model samples (using the same or comparable pigments) and those of the pure binders. Each of the historical samples has also been verified with classical amino acid analysis methods, after Scott *et al.* (GC-MS) (28) or Wouters *et al.* (HPLC)(27). Similar sample quantities (scrapings of ca. 100 - 200 µg) were used for all methods. The identified pigments reported in the table are determined by non-invasive (XRF and MRS) and cross-sectional (SEM-EDX and MRS) analysis; these results have been published previously(147,148).

³ The retention factor describes the migration rate of an analyte on a column. It is defined as the ratio of the adjusted retention time (or volume) to the hold-up time (volume). The hold-up time (t_M) is time required to elute a component that is not retained at all. The adjusted retention time of a component is the retention time (t_R) minus the hold-up time. The retention factor of a component is thus $k = t_R - t_M / t_M$.

Table IV.2. Identification of the binders of historical paint samples with amino acid and peptide analysis.

Sample	Description	Pigments	Amino acid analysis (^G = GC-MS, ^H = HPLC)	Peptide analysis (HPLC-DAD)
1	Reliquary of the Virgin's Veil, black	Carbon black (C)	Casein and/or egg?, animal glue?, gum? ^G	Casein, yolk
2	Reliquary of the Virgin's veil, orange	Red lead (Pb ₃ O ₄)	Animal glue ^G	Animal glue
3	Crucifixion with St Catherine and St Barbara, red-brown	Red earth (Fe ₂ O ₃)	Animal glue ^G	Animal glue
4	St Ursula shrine, red	Vermilion (HgS)	gum? (low on hydroxyproline, low correlation coefficient) ^H	Casein

In the case of sample 1, the GC-MS method suggested a complex mixture. In this case, identification with the help of Pearson's *r* correlation coefficients of theoretical mixtures of reference samples is cumbersome and it is difficult to distinguish between casein and egg. The small proportion of animal glue (collagen) found with GC-MS might be caused by a small contamination of ground layer in the analysed paint sample. With HPLC-DAD both casein and yolk peptides were found. As mixtures of casein and egg seemed uncommon and illogical practice, a sample taken from the black background was transformed into a cross section for observation with polarised light microscopy (PLM), fluorescence microscopy and analysis with scanning electron microscopy coupled to an energy-dispersive X-ray detector (SEM-EDX). This revealed not less than four superimposed black layers with a thin isolating layer between the second and third layer (figure IV.6). It seems more probable that casein and yolk are present as binding media in different black paint layers.

Samples 2 and 3 can both be identified as animal glue (collagen) using the 214 nm and the 280 nm (figure IV.7) chromatograms. Amino acid analysis with GC-MS confirmed these results.

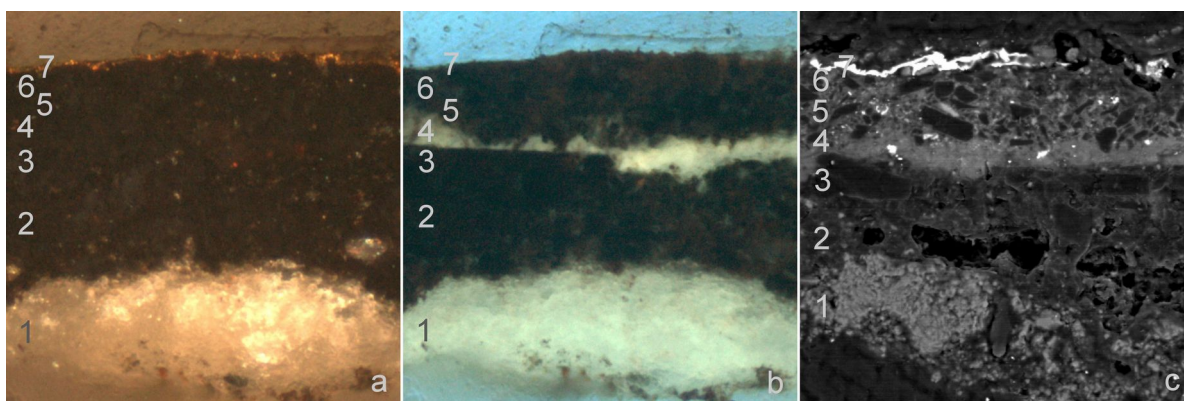


Figure IV.6. Cross section of the gilded black background of the Reliquary of the Virgin's Veil, revealing four black layers (2, 3, 5 and 6) with an isolating layer in between (4). Pictures taken with (a) polarised light, (b) fluorescence and (c) scanning electron microscopy.

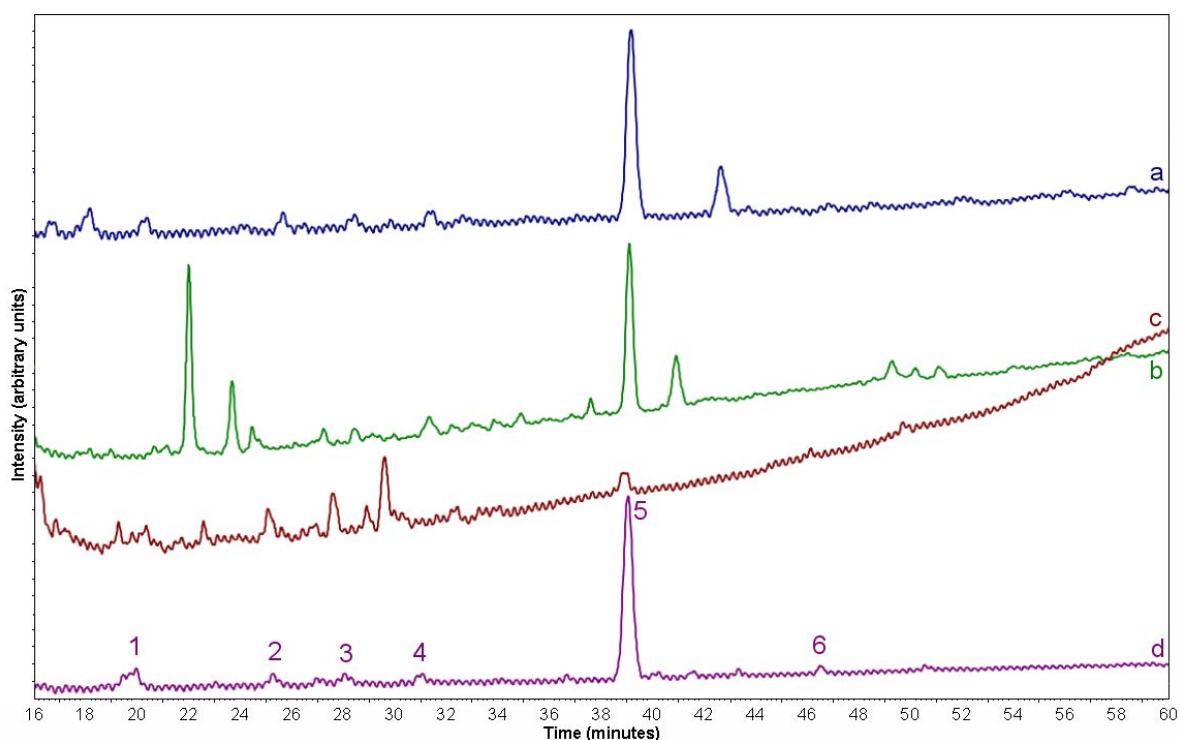


Figure IV.7. Chromatograms at 280 nm of (a) the orange sample from the Reliquary of the Virgin's veil, (b) the red-brownish sample from Crucifixion with St Catherine and St Barbara (retention shift corrected), (c) a model red earth/animal glue paint sample and (d) pure collagen.

HPLC amino acid analysis on sample 4 yielded dubious results. Based on its glycoprotein content, the best match was a gum with low hydroxyproline levels, as the procedure is not able to analyse polysaccharides. The Pearson's r correlation coefficient of the relative amounts of seven stable amino acids between sample 4

and the gum reference, however, was too low to be conclusive. Peptide analysis on the same sample gave a good match with the casein chromatograms (figure IV.8): nearly all peaks of the pure casein reference are present. A prepared paint sample containing vermilion and borax casein binder, shows some of the most important peaks, but is missing other prominent peptides, probably due to the different preparation method of the borax casein binder. It is not impossible that casein was used as a binding medium or as an adhesive during one of the previous (undocumented) restoration campaigns. The interpretation of these results in their historical context should be done.

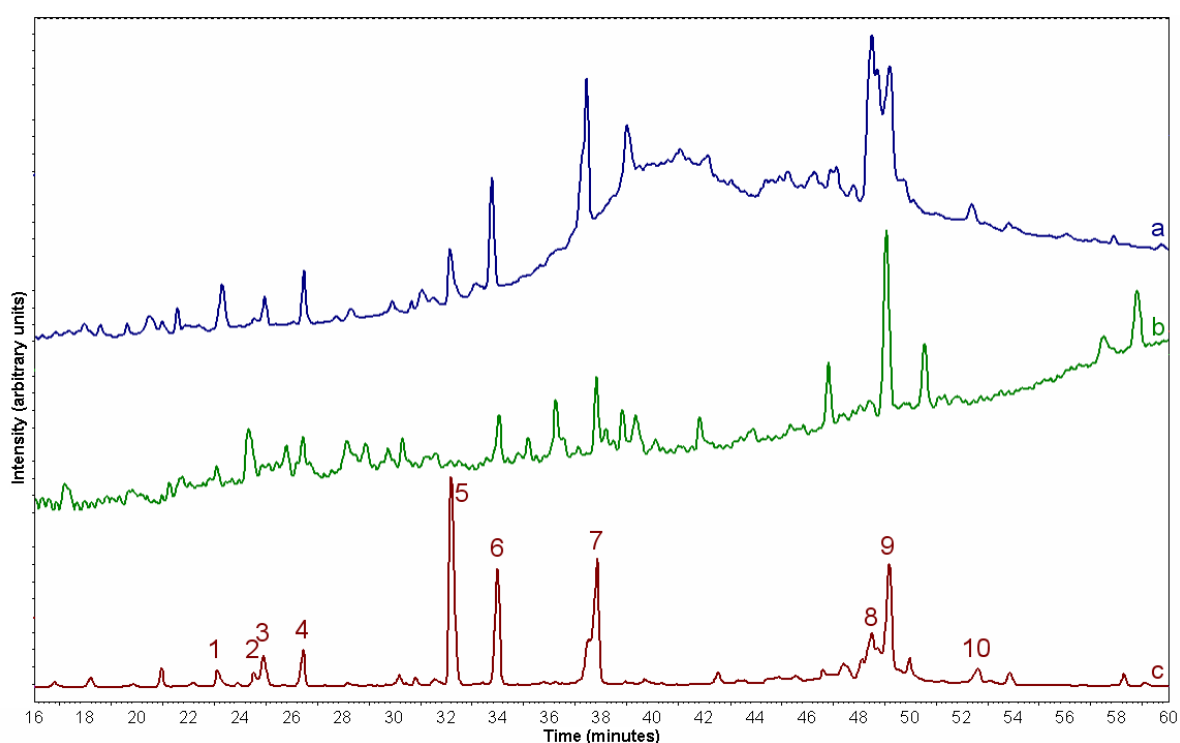


Figure IV.8. Chromatograms at 280 nm of (a) the unknown red sample from the small St Ursula Shrine, (b) a model vermilion/borax casein paint sample and (c) pure casein.

In these examples, not all peaks present in the reference samples are found in the paint sample and some peaks in the sample cannot be attributed to any of the references, probably indicating some kind of degradation effects. Peptides may be destroyed (no peptide peak any more) or altered (peak shift) during ageing or caused by interaction with, amongst others, pigments. A study of these effects, however, is beyond the possibilities of the DAD. For this, mass spectrometry is a more suitable technique; it is able to identify precisely the separated peptides and

to identify precisely peptide modifications between the paint models and ancient samples.

IV.4 Conclusions

Peptide analysis by HPLC-DAD proves to be a straightforward technique with instruments that are already available in many art analysis labs, and can easily compete with or complement the classical amino acid analysis techniques such as HPLC and GC-MS. In respect to these, the novel technique is especially advantageous to distinguish between casein and egg binders and to distinguish between egg white and yolk. Due to its different resistance to trypsin digestion, animal glue (collagens) yields a much weaker chromatogram, but still sufficient for identification. The sample size is about 100 to 200 µg – the same as for amino acid analysis, but a tenfold higher than other, mass spectrometric proteomic techniques.

Identification is achieved by comparing the chromatogram of the unknown binder digest with those of reference products, in which great care has to be taken in maintaining the experimental conditions during all measurements to prevent peak shifts. UV spectra of tryptic peptides do not allow peptide identification, nor to study the alterations in the peptides due to degradation, for which a mass spectrometer should be used.

The following paragraphs are not included in the published scientific paper:

This chapter focused on the initial application of protein digestion. A basic method of separation and detection by HPLC-DAD enabled us to identify relatively easily the different protein binders by fingerprinting. In the next chapter a more robust detection method will be explored: mass spectrometry allows to determine the mass and charge of the tryptic peptides, while detection limits (hence sample consumption) are significantly lower.

A number of new applications using MALDI-TOF-MS were published during and after the publication of this paper upon which this chapter is based: the team of Hynek and Kuckova continued their research on protein binders in paints^(83,150) and masonry^(79,151,152), while others worked on archaeological samples with

both MALDI-TOF-MS(78,80,81,153) and HPLC-MS/MS(86,89,154–156). Chambery *et al.* used the sample preparation method that was developed in this chapter in combination with HPLC-MS/MS to identify milk casein up to species level(88).

V

CLASSIFICATION OF PROTEIN BINDERS IN ARTISTS' PAINTS BY MATRIX-ASSISTED LASER DESORPTION/IONISATION TIME-OF-FLIGHT MASS SPECTROMETRY: AN EVALUATION OF PRINCIPAL COMPONENT ANALYSIS AND SOFT INDEPENDENT MODELLING OF CLASS ANALOGY

Based on the scientific paper:

*W. Fremout, S. Kuckova, M. Chrova, J. Sanyova, S. Saverwyns,
R. Hynek, M. Kodicek, P. Vandenabeele, L. Moens,
Rapid Commun. Mass Spectrom. 25 (2011) 1631-1640.*

The complex nature of paint samples, containing different kinds of pigments and degraded by long term exposure to light, humidity and temperature variations, requires robust analysis and interpretation methods to identify the binding media. Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) has proven to be a powerful technique for protein analysis in proteomics and, more recently, in conservation science. The method extensively described by Hynek and Kuckova is used in this chapter(75–77). The novelty of the research in this chapter, however, is in the application of principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA) in order to distinguish between proteinaceous binders based on animal glues, egg white, egg yolk and

milk casein. The mass spectra recorded from samples of a large set of tryptic-digested paint models are used as a training set. By doing so, the most meaningful peptide peaks for a given protein class are determined and, if possible, assigned to their corresponding amino acid sequence. The methodology is subsequently applied on animal glues from different species: small differences in the MALDI-TOF spectra can allow the provenancing of the glue. Finally, three paint samples from the 16th century altarpiece of St Margaret of Antioch (Mlynica, Slovakia) are analysed.

V.1 Introduction

The analysis of tryptic peptides, as commonly used in proteomics, is emerging as an extremely powerful technique in conservation science when it concerns the identification of proteins, for example in archaeological remains(80,86,157), in construction materials(79,151) and as binding media in paints(75–77,85,87,88). Common protein sources found in paints are animal glues, egg (glair and yolk) and milk caseins. Several proteomics-based techniques, using high-performance liquid chromatography with diode array detection (HPLC-DAD, see chapter IV), MALDI-TOF-MS(75–77) and HPLC with tandem mass spectrometry (HPLC-MS/MS, see chapter VI)(85,87,88), are successfully and increasingly being applied in conservation science. These proteomics techniques are based on the analysis of peptides that are created by proteolysis, typically using trypsin. This way, some of the drawbacks of the more traditional methods (amino acid analysis (AAA) and pyrolysis gas chromatography mass spectrometry (py-GC-MS) used for protein binder identification can be addressed. AAA using GC-MS(30,96,139) or HPLC(30) after complete hydrolysis of the protein content is easily hampered by the presence of contaminants and unknown amino acid sources. Protein binders with the same or highly similar amino acid composition cannot be distinguished. Also, AAA is susceptible to alteration of amino acid ratios, both during sample preparation and due to chemical transformation of paints in time. In py-GC-MS(36,145), on the other hand, small specific marker components are used for identification. Contrary to amino acids and pyrolysis markers, tryptic peptides are highly specific for a given protein binder. They permit an unambiguous identification, even in case of mixtures and in case proteinaceous contaminants

are present, while maintaining or even decreasing the sample size (see section VI.3.1, page 127).

In the recent past MALDI-TOF-MS has proven to be a valuable tool for fast and reliable protein binder identification, preceded by very limited sample preparation. This way sample losses are minimised. Proteomics studies commonly apply library searches comparing the resulting mass spectra with calculated peptide masses based on theoretical protein cleavage. These are, however, not usable in paint binder samples, because the latter are in fact complex mixtures of different proteins, which are all proteolysed into peptides; the mass spectrum is thus a superimposed result of tryptic peptides from many different proteins. In a previously published paper(79) a basic procedure was described comparing semi-automatic peaklists from the unknown sample with those of the reference samples. Variability, caused by several factors, hampers however this data treatment procedure. Firstly, variability is expected in the trypsin digest of different samples due to irreproducibility of the experimental conditions. The trypsin-to-protein ratio for example cannot be respected, since the samples are too small to be weighed and the protein ratio in the samples is unknown. Also miscleavages can occur in certain conditions, causing other (longer) peptides, and the presence of pigments can have an influence on the proteolysis and the resulting components. Secondly, MALDI-TOF is known to exhibit a substantial variability in ionisation due to heterogeneous co-crystallisation of analyte and matrix(53). To reduce this effect a large number of shots on different places of the spot are accumulated in a single measurement. Thirdly, if the wrong isotope is selected by the automatic or manual peak picking procedure, attribution of this component by database search may fail. Finally, in case of mixtures or when a sample is contaminated with unknown protein sources, the homology with the reference peaklists will be superficially low.

In this chapter an analytical protocol for protein binder identification using tryptic peptide analysis with MALDI-TOF-MS is proposed. In a first step the obtained mass spectra will be processed automatically, without manual intervention and without cumbersome peak-picking procedures. Further, PCA and SIMCA will be evaluated to classify the different protein classes. PCA will also allow selecting the most influential and discriminative tryptic peptides. Applying PCA on a broad set of paint models, effects from contaminants (keratins), pigment, proteolysis and

ionisation variability will be neglected, and significant signals will be selected in an unsupervised manner, which will allow for positive identification in nearly all conditions. SIMCA on the other hand will allow a supervised pattern recognition model, based on the proximity of unknowns to the different groups in the training set.

Both methodologies are finally evaluated on a set of testing samples, different animal glues and temperas, and subsequently on three paint samples taken from the 16th century polychrome altarpiece of St Margaret of Antioch (Roman Catholic church consecrated to this saint in Mlynica, Slovakia), attributed to Master Paul of Levoča, a sculptor and author of many carved late Gothic altarpieces in East Slovakia(158).

V.2 Experimental

V.2.1 Reagents

Sequencing-grade modified trypsin was acquired from Promega, trifluoroacetic acid (TFA), 2,5-dihydroxybenzoic acid (DHB) from Sigma, acetonitrile (p.a.) and ammonium bicarbonate from Lachema Brno. Reversed-phase C18 solid phase extraction tips (ZipTip) were supplied by Milipore.

V.2.2 Samples

A set of paint models was made by mixing ten historical pigments with four natural binders. The binders were prepared as follows: rabbit skin glue (KIK/IRPA binder library, source unknown) was allowed to swell in water (8% w/w) for 24 h and subsequently heated to 50°C to completely dissolve the glue. Fresh eggs and soft curd cheese (quark cheese) were bought in a local supermarket. The egg white and yolk were separated. The egg white was stirred into a stable foam and allowed to stand for several hours. The reliquidised fraction was separated from the foam, and was used as glair binder. The egg yolk was suspended above a recipient and the membrane was cut open. Soft curd cheese (25 g) was washed with water on a Büchner funnel to eliminate the soluble gums and sugars. After adding 10 mL water and 1 g of borax (UCB) to the curds, the mixture was stirred for

40 min until a syrupy solution was formed, which was diluted (1:4) and used as borax casein binder. Azurite ($\text{Cu}_3(\text{CO}_3)_2(\text{OH})_2$, KIK/IRPA reference collection, source unknown), lead white ($(\text{PbCO}_3)_2\cdot\text{Pb}(\text{OH})_2$, KIK/IRPA reference collection, source unknown), red earth (Fe_2O_3 in a silicate matrix, Blockx, Terwagne, Belgium), chalk (CaCO_3 , KIK/IRPA reference collection, source unknown), chrome yellow (PbCrO_4 , Winsor & Newton, Harrow, UK), lead-tin yellow type I (Pb_2SnO_4 , reference collection, source unknown), verdigris ($\text{Cu}(\text{OH})_2\cdot(\text{CH}_3\text{COO})_2\cdot 5\text{H}_2\text{O}$, Winsor & Newton), carbon black (C, Kremer Pigmente, Aichstetten, Germany), vermilion (HgS , Winsor & Newton) and Prussian blue ($\text{Fe}_7(\text{CN})_{18}(\text{H}_2\text{O})_x$ where $14 \leq x \leq 16$, Winsor & Newton) were selected as pigments. The sample matrix thus consists of 44 pigmented and unpigmented paint samples. The samples were stored in the dark at room temperature.

Pure animal glues from different species (rabbit skin, bovine skin, deer skin and sturgeon bladder glues) were taken from the KIK/IRPA reference collection, without any sample preparation. Two different egg temperas were made using egg white and yolk binders described above and mixed with linseed oil: egg yolk and linseed oil (1:1) and egg white, yolk and linseed oil (1:1:1). The tempera samples were let to dry and stored in the dark at room temperature.

Three microsamples were collected from the altarpiece of St Margaret of Antioch by gently scraping off a single paint layer using a scalpel. Samples S1, S2 and S3 are taken from the paper underlayer behind the carved ornament from the arch; sample S1 is taken from the blue area (azurite), S2 from the green area (malachite) and S3 contains fragments of the greyish chalk ground layer. The pigments were identified using micro-Raman spectroscopy.

V.2.3 Sample pretreatment

The sample pretreatment has been described previously^(75–77,79,151) and will only be summarised here. The microsamples (ca. 10–100 μg) were immersed in 15 μL diluted trypsin (0.4 $\mu\text{g } \mu\text{L}^{-1}$ in 50mM ammonium bicarbonate) and let to digest for 2 h at room temperature. After wetting (50% acetonitrile) and equilibrating (2% TFA), ZipTip pipette tips were loaded with the sample solution and washed with 2% TFA. The peptides were eluted using 5 μL 0.1% TFA/50% acetonitrile in water. Finally

4 μ L of a solution containing 0.1 M DHB and 0.1% TFA in water/acetonitrile (7:3) was added to the eluted peptides. A small drop of the resulting solution was spotted on a stainless steel MALDI target plate and let to evaporate before analysis.

V.2.4 Instrumentation

A Biflex IV MALDI-TOF-MS from Bruker Daltonics, equipped with a nitrogen laser (337 nm) was used in positive reflectron mode to record the mass spectra. At least 200 laser shots were collected for each spectrum. The software package XMASS (Bruker) was used for the acquisition of the spectra and the freely available open source tool mMass(118–120) for manipulation and interpretation of the data and exporting to ASCII data files.

V.2.5 PCA and SIMCA

The data files were converted into readable ASCII format using mMass. In order to perform PCA on the data, the mass spectra were uniformised in a batch operation using a Python script. The Unscrambler X.1 (Camo Software) was used for calculating the PCA and SIMCA models. The calculation of the principal components is based on the preprocessed paint model data and the series of different animal glues; afterwards the coordinates in the reduced dimensions are calculated for the unknowns, which are thus not taken into account for the factor axis calculations.

V.3 Results and discussion

V.3.1 Determination of the main protein binder class

MALDI-TOF-MS data were collected from a set paint models, based on the four most important protein sources: animal glue, egg white, egg yolk and milk casein. These binders are complex mixtures of different proteins and other organic substances. As a consequence, the resulting spectra contain the superimposed signal from the different components. Other than the expected peaks from the peptides originating from the binders, peaks might be caused by the ten different pigments that were mixed with the binders. Other spectral features might be caused by eventual contamination (e.g. keratins) or might be system peaks. PCA is

therefore an exquisite technique to define the most influential masses for a given binder class.

Before PCA is performed, however, the data need to be uniformised. A script in Python language was written to automate this task. Firstly, the mass spectrum is limited to the range 900-2000 Da, since too many peaks originating from salts and the DHB matrix are masking the mass spectral region below 900 Da, while peaks in the higher mass region are typical for miscleavages, which highly depend on the concentration of trypsin. Secondly, to correct for different numbers of datapoints in different measurements and unavoidable small intermeasurement mass shifts, the number of datapoints is reduced to 1101 points, resulting in one datapoint per Dalton. In doing so, the influence of misalignment on the PCA results will be minimal, since a misaligned peak will always coincide with the aligned mass of one of its isotopes. Finally, the data are normalised on a scale from 0 to 100.

The processed mass spectra of the 44 paint model samples were subjected to PCA. This unsupervised chemometrical method allows us to select the most influential and significant datapoints to be used for identification. Sources of possible variability, such as the varying proteolysis and ionisation conditions and the influence of contaminations and pigments in the samples, are excluded by PCA. The effect and classification capabilities of the different principal components were studied. The two first principal components PC1 and PC2 hold respectively 31% and 25% of the total variability. The score plot (figure V.1) allowed us to classify paint models visually according to the different protein binders: the egg white (E) and milk casein (M) models form tight groups, while the rabbit skin glue (G) and egg yolk (Y) models are more spread along respectively the PC2 and PC1 axis. In all four classes one outlier can be noticed with a remarkable lower value for PC1; these are the four unpigmented models. While the presence of pigments has a small influence on the PCA model, none of the pigments negatively affect the model in any way.

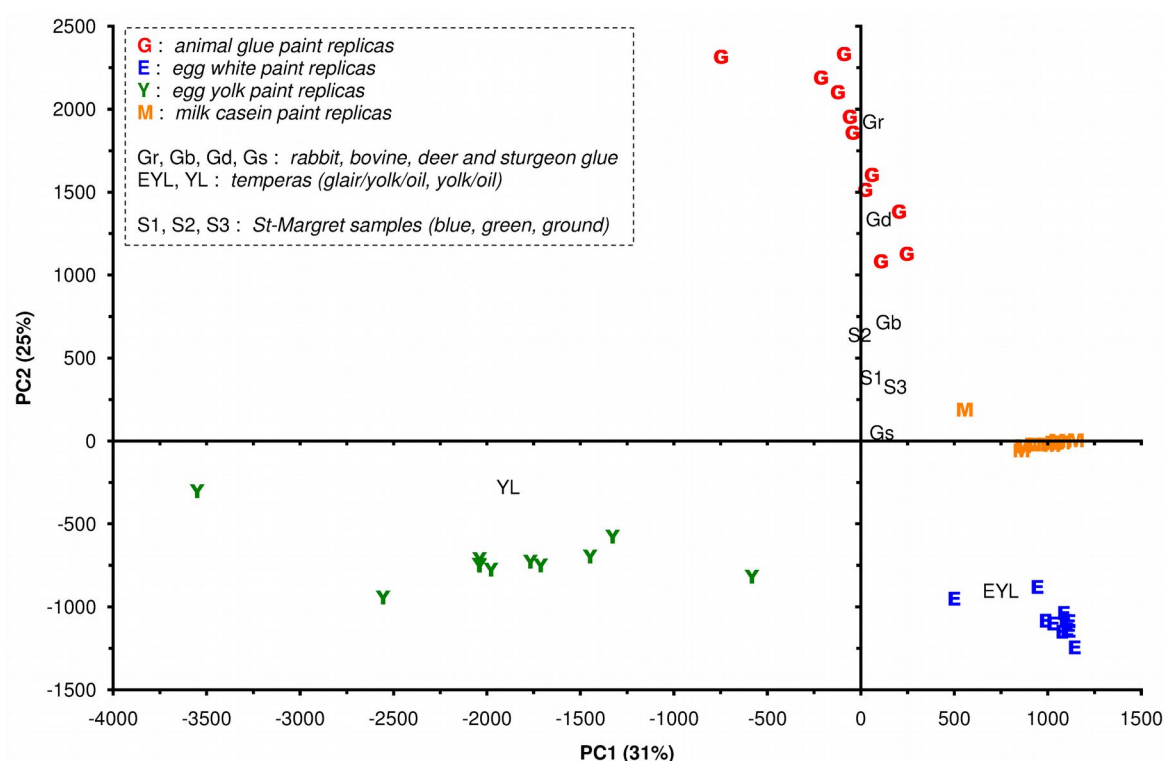


Figure V.1. Score plot of PC1 and PC2 visually separating the main protein binder classes.

The most decisive datapoints (peptide peaks in the MALDI-TOF spectra) in the PCA model can be determined by their factor loadings for PC1 and PC2, which give an insight into the weight of a certain datapoint to classify the samples. As a result, the 1101 datapoints, obtained after preprocessing, can be drastically reduced by only taking into account the most influential datapoints. Because of the large set of samples for each binding medium, with eleven different pigments added, only those datapoints (peaks) that are characteristic for a binder group are automatically retained; datapoints that are due to pigments or to pigment-binder interaction products, as well as to occasional contaminations and variability in the proteolysis and ionisation, only have limited influence on the factor loadings of PC1 and PC2. The retained datapoints are then manually checked not to include different isotopes of the same peptide peak; only the most influential isotope is retained. In most cases this will be the mono-isotopic peak of a peptide, but in case of overlapping isotopic distributions from peptides in different protein binders, PCA will likely have preferred another isotope. As a final result of this procedure, a total of 61 datapoints are selected, of which 19, 8, 25 and 9 are characteristic for respectively rabbit glue (table V.1), egg white (table V.2), yolk (table V.3) and milk

casein (table V.4). Possible peptide sequences are allocated as found by the FindMod tool(117). For a given protein, FindMod calculates theoretical peptides and compares them with the supplied characteristic masses, taking into account possible post-translational modifications. Results were calculated for all proteins that were identified by HPLC-MS/MS (section VI.3.1, page 121). Because rabbit collagen sequences are either incomplete or missing in the Swissprot protein database, bovine collagens were included in the search tasks. Most frequently observed post-translational modifications are multiples of 16 Da (oxidation/hydroxylation). In case of collagen peptides this is likely hydroxylation of proline into hydroxyproline. Besides, in the milk casein proteins phosphorylation (+80 Da) is expected.

Table V.1. Characteristic datapoints in the animal glue models and their possible sequence annotations as found by the FindMod tool. () indicates one or more possible oxidations as a post-translational modification (multiples of 16 Da).*

Processed dataset		Possible peptide sequence annotation		
Mass (m/z)	Height (%)	Mass (m/z)	Peptide sequence	Protein
900	9.8	898.45	GPSGDRGPR	Bovine collagen α -2(I)
		898.51	GWGLPGQR +16 Da*	Bovine collagen α -1(I)
1106	40.4	1104.53	GFPGADGVAGPK +32 Da*	Bovine collagen α -1(I)
		1105.57	GVQGPPGPAGPR +16 Da*	Bovine collagen α -1(I)
1202	5.7	1201.61	GHRGFSGLDGAK	Bovine collagen α -1(I)
		1201.58	GEPGNIGFPGPK +32 Da*	Bovine collagen α -2(I)
		1200.60	GPSGPQGIRGDK +32 Da*	Rabbit/bovine collagen α -2(I)
1428	30.3	1427.70	GSAGPPGATGFPGAAGR	Bovine collagen α -1(I)
		1427.79	ALLIQGSNDVEIR	Bovine collagen α -1(II)
		1427.73	GIPGEFGLPGPAGAR +32 Da*	Bovine collagen α -2(I)
1454	86.3	1452.73	SAGVSVPGPMGPSGPR	Rabbit collagen α -1(I)
1467	12.7			
1470	10.5	1468.72	SAGVSVPGPMGPSGPR +16 Da*	Rabbit collagen α -1(I)
1502	57.4	1500.71	SAGVSVPGPMGPSGPR +48 Da*	Rabbit collagen α -1(I)
1534	24.0	1533.79	DGRSGHPGTVGPAGLR	Rabbit collagen α -2(I)
1565	15.9	1567.75	GEVGPAAGSPGSSGAPGQR	Bovine collagen α -1(III)
		1565.78	DGRSGHPGTVGPAGLR +16 Da*	Rabbit collagen α -2(I)

(table continued on next page)

Processed dataset		Possible peptide sequence annotation		
Mass (m/z)	Height (%)	Mass (m/z)	Peptide sequence	Protein
1587	78.3	1586.74	GNSGEPGAPGSKGDTGAK	Bovine collagen α -1(I)
1649	9.3	1649.86	GRPGLPGAAGARGNDGAR	Bovine collagen α -1(III)
		1649.78	GFSGLDGAKGDAGPAGPK +48 Da*	Bovine collagen α -1(I)
		1648.78	AGEDGHPGKPRPGER +32 Da*	Bovine collagen α -2(I)
		1648.83	GSTGEIGPAGPPGPPGLR +32 Da*	Bovine collagen α -2(I)
1656	5.5	1656.82	GLVGEPGPAGSKGESGNK +16 Da*	Bovine collagen α -2(I)
		1655.81	GFPGADGVAGPKGPAGER +16 Da*	Bovine collagen α -1(I)
		1654.82	GERGFPGLPGPSGEPGK +16 Da*	Bovine collagen α -1(I)
		1655.76	GEPGSSGVDGAPGKDGPGR +16 Da*	Bovine collagen α -1(III)
1728	9.2	1726.93	DGLNGLPGPIGPPGPRGR	Bovine collagen α -1(I)
		1725.75	GENGVPGEDGAPGPMGPR +16 Da*	Bovine collagen α -1(III)
1817	16.5	1816.86	GPPGPMGPPGLAGPPGESGR +32 Da*	Bovine collagen α -1(I)
		1816.91	SLSQQIENIRSEGSR +16 Da*	Bovine collagen α -1(I)
		1817.91	GPPGPQGLPGLAGTAGEPGR +32 Da*	Bovine collagen α -1(III)
1834	38.1	1831.88	GEPGPTGIQGPPGPAGEEGK	Bovine collagen α -1(I)
1850	14.8	1847.88	GEPGPTGIQGPPGPAGEEGK +16 Da*	Bovine collagen α -1(I)
1873	11.3			
1889	22.9	1887.83	DGNPGNDGPPGRDGQPGHK +16 Da*	Bovine collagen α -2(I)

Table V.2. Characteristic datapoints in the egg white models and their possible sequence annotations as found by the FindMod tool.

Processed dataset		Possible peptide sequence annotation		
Mass (m/z)	Height (%)	Mass (m/z)	Peptide sequence	Protein
1346	3.5	1345.74	HIATNAVLFFGR	Chicken ovalbumin
1536	8.7	1534.85	SAGWNIPIGTLLHR	Chicken ovotransferrin
1556	14.1	1555.72	AFKDEDTQAMPFR	Chicken ovalbumin
1582	18.7	1581.72	LTEWTSSNVMEER	Chicken ovalbumin
1688	91.5	1687.84	GGLEPINFQTAADQAR	Chicken ovalbumin
1754	7.6	1753.84	NTDGSTDYGLQINSR	Chicken lysozyme C
1774	15.8	1773.90	ISQAVHAAHAEINEAGR	Chicken ovalbumin
1860	94.3	1858.97	ELINSWVESQTNGIIR	Chicken ovalbumin

Table V.3. Characteristic datapoints in the egg yolk models and their possible sequence annotations (as found ^(a) by the FindMod tool or ^(b) in Kuckova et al.(77)). (*) indicates one or more possible oxidations as a post-translational modification (multiples of 16 Da).

Processed dataset		Possible peptide sequence annotation		
Mass (m/z)	Height (%)	Mass (m/z)	Peptide sequence	Protein
922	9.4	923.50	KTSTALMR +16 Da ^{*,a}	Chicken vitellogenin-1
		922.46	YVIQEDR ^a	Chicken vitellogenin-2
990	6.7	988.56	MAVRALSPK +16 Da ^{*,a}	Chicken vitellogenin-1
		988.55	ELLQQVMK ^a	Chicken vitellogenin-2
		989.54	MVVALTSPR +16 Da ^{*,a}	Chicken vitellogenin-2
1003	6.2			
1045	5.2			
1049	60.4	1048.65	LPLSLPVGPR ^a	Chicken vitellogenin-2
1078	54.4	1077.56	VFRFSMFK +16 Da ^{*,a}	Chicken vitellogenin-1
1086	25.7	1088.61	DSKRGKIER ^b	Chicken low density lipoprotein receptor-related protein 1
		1085.61	LSDWKALPR ^a	Chicken vitellogenin-1
		1085.60	SIKAEIPPCK ^a	Chicken vitellogenin-1
		1085.63	QQLTLEVR ^a	Chicken vitellogenin-2
1151	8.5	1150.60	VTVASWMRGK +16 Da ^{*,a}	Chicken vitellogenin-1
		1150.56	MTPPLTGDFR +16 Da ^{*,a}	Chicken vitellogenin-2
1165	37.1	1163.65	ALGNVGHPSIK ^a	Chicken vitellogenin-1
		1163.54	TVDLNNCQEK ^a	Chicken vitellogenin-2
		1164.50	ADTYFDNYR ^a	Chicken vitellogenin-2
1366	12.5	1365.75	VPGVTLYYQGLR ^a	Chicken vitellogenin-1
1402	68.5	1401.71	IANADNLESIWR ^a	Chicken vitellogenin-2
1435	11.5	1435.80	ILGIDSMFKVANK ^{a,b}	Chicken vitellogenin-2
		1434.76	LSQLLESTMQIR +16 Da ^{*,a}	Chicken vitellogenin-2
1440	10.9	1439.64	AANEENYESVWK ^a	Chicken vitellogenin-1
		1439.80	YLLDLLPAAASHR ^a	Chicken vitellogenin-1
		1439.77	GILNMFQMTIKK +16 Da ^{*,a}	Chicken vitellogenin-2
1446	47.9	1445.77	VGATGEIFVVNSPR ^a	Chicken vitellogenin-2
		1445.68	SSHDSRAASWPK +16 Da ^{*,a}	Chicken vitellogenin-2

(table continued on next page)

Processed dataset		Possible peptide sequence annotation		
Mass (m/z)	Height (%)	Mass (m/z)	Peptide sequence	Protein
1464	6.6	1459.64	KPEHEFLVYGK ^b	Chicken low density lipoprotein receptor-related protein 1
		1464.77	KVCLLSYASLCHK ^a	Chicken vitellogenin-1
1561	78.2	1559.87	LEISGLPENAYLLK ^a	Chicken vitellogenin-2
		1560.74	SPQVEEYNGVWPR ^a	Chicken vitellogenin-2
		1560.85	ELPTETPLVSAYLK ^a	Chicken vitellogenin-2
1593	19.5	1591.75	GSAPDVPMQNYGSLR ^a	Chicken vitellogenin-2
1608	23.9	1607.74	GSAPDVPMQNYGSLR +16 Da ^{*,a}	Chicken vitellogenin-2
1664	9.5	1663.88	SDFRLTELLNSNVR ^a	Chicken vitellogenin-1
		1663.89	WLLSAVSASGTETLK ^a	Chicken vitellogenin-2
1778	10.6	1777.91	EALQPIHDLADEAISR ^a	Chicken vitellogenin-2
		1776.87	IGSHEIDMHPVNGQVK +16 Da ^{*,a}	Chicken vitellogenin-2
1839	10.2			
1892	93.3	1891.96	AAVSVEGKMTPLTGDFR +16 Da ^{*,a}	Chicken vitellogenin-2
1914	12.6	1914.10	YRFAVLPLQMLQLIK ^a	Chicken vitellogenin-2
1947	7.8	1947.01	NPVLQQVACLGYSSVVNR ^a	Chicken vitellogenin-2
		1947.03	TGGLQLVYADTDSVRPR ^a	Chicken vitellogenin-2
1974	18.2	1975.74	LSSKLEISGLPENAYLLK ^b	Chicken vitellogenin-2
		1972.97	NSIAGQWTQPVWMGELR ^a	Chicken vitellogenin-2

Table V.4. Characteristic datapoints in the milk casein models and their possible sequence annotations (as found (^a) by the FindMod tool or (^b) in Kuckova et al.(79)). One or more possible post-translational modifications are indicated by (*) for oxidations (multiples of 16 Da) and (°) phosphorylations (multiples of 80 Da).

Processed dataset		Possible peptide sequence annotation		
Mass (m/z)	Height (%)	Mass (m/z)	Peptide sequence	Protein
971	14.5	970.41	NMAINPSK +96 Da ^{*,o,a}	Bovine α _{S2} -casein
1196	36.1	1195.68	NAVPIPTLNR ^{a,b}	Bovine α _{S2} -casein
1253	26.4	1251.75	TKVIPYVRYL ^b	Bovine α _{S2} -casein
		1251.57	EQLSTSEENSK ^a	Bovine α _{S2} -casein
		1251.71	YIPIQYVLSR ^a	Bovine κ-casein
1268	99.8	1267.70	YLGYLEQLLR ^{a,b}	Bovine α _{S1} -casein

(table continued on next page)

Processed dataset		Possible peptide sequence annotation		
Mass (m/z)	Height (%)	Mass (m/z)	Peptide sequence	Protein
1385	31.1	1384.73	FFVAPFPFVFGK ^b	Bovine α_{S1} -casein
1661	7.1	1658.78	LSFNPTQLEEQCHI ^b	Bovine β -lactoglobulin
		1660.79	VPQLEIVPNSAEER +80 Da ^{o,a}	Bovine α_{S1} -casein
1760	48.1	1759.94	HQGLPQEVNLNENLLR ^{a,b}	Bovine α_{S1} -casein
1953	9.1	1951.95	YKVPQLEIVPNSAEER +80 Da ^{o,a}	Bovine α_{S1} -casein
1960	18.7	1958.69	DIGSESTEDQAMEDIK +192 Da ^{*,o,a}	Bovine α_{S1} -casein

To test the models, two unpigmented egg tempera models were measured. Their MALDI-TOF spectra were projected in the PCA score plot (figure V.1, page 100). The first, based on a 1:1 mixture of egg yolk and linseed oil (YL), correlates well with the yolk group. The presence of linseed oil in the sample clearly does not hamper the analysis. The second tempera sample is a 1:1:1 mixture of egg white, yolk and linseed oil (EYL), but visually only correlates to the egg white models in the score plot (figure V.1, page 100). Looking at the previously selected markers for egg white and yolk as shown in figure V.2, high responses are indeed observed for the egg white markers. The egg yolk markers are present too, but as smaller peaks that have limited impact on the PCA model. The inferior outcome of the yolk markers can be explained by the fact that the proteins in yolk are packed in lipoid particles and the enzyme has little chance to cleave them because of their hydrophobicity.

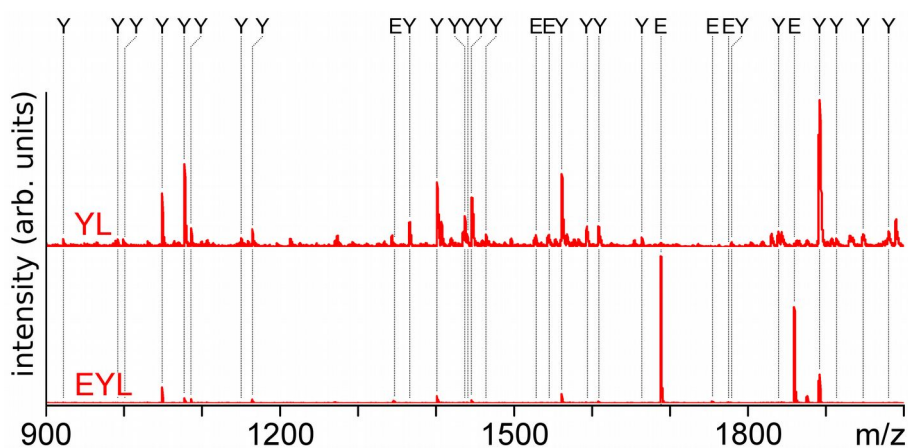


Figure V.2. MALDI-TOF spectra of tempera models. The characteristic peaks for the egg white ("E") and yolk ("Y") paint models are marked with dotted lines.

Whereas PCA failed to visually group EYL in both the egg white and yolk classes, due to the low signal from the yolk peptides, SIMCA trained with the four main protein classes (table V.5) successfully classified both the YL and EYL samples: YL is classified solely into the group of the egg yolk models, while EYL is classified to both the egg white and yolk groups. Clearly the lower abundance of the yolk peptides did not restrain its correct classification.

Table V.5. SIMCA classification based on the main protein binder classes.

Sample	Animal glue paint models	Egg white paint models	Egg yolk paint models	Milk casein paint models
YL			●	
EYL		●	●	
Gr	●			
Gb	●			
Gd	●			
Gs				
S1	●			
S2	●			
S3	●		●	

Four pure animal glue samples from different species were analysed. The results are shown in the score plot (figure V.1, page 100) as passive observations. The scores for bovine glue (Gb), deer glue (Gd) and rabbit glue (Gr) correlate well to the animal glue models. The score for sturgeon glue (Gs) is more isolated. These observations are on a par with the results of the SIMCA classification (table V.5). Herein Gr, Gb and Gd are correctly classified as animal glue, while the model failed to classify Gs. Apparently sturgeon collagen differs too much from that of rabbit, on which the animal glue paint models in the training set were based. Indeed, while most characteristic datapoints for rabbit glue are equally valid for bovine and deer glues, none of the 19 datapoints exhibit a high response in the sturgeon glue sample, as illustrated in figure V.3. Datapoint 1599, according to its loadings a weak marker for egg white, is one of the highest and most characteristic datapoints for sturgeon glue. It is assumed that the mutations between collagens from different mammals are relatively small, while mammal and sturgeon collagens

differ much more from each other and few common tryptic peptides can be observed.

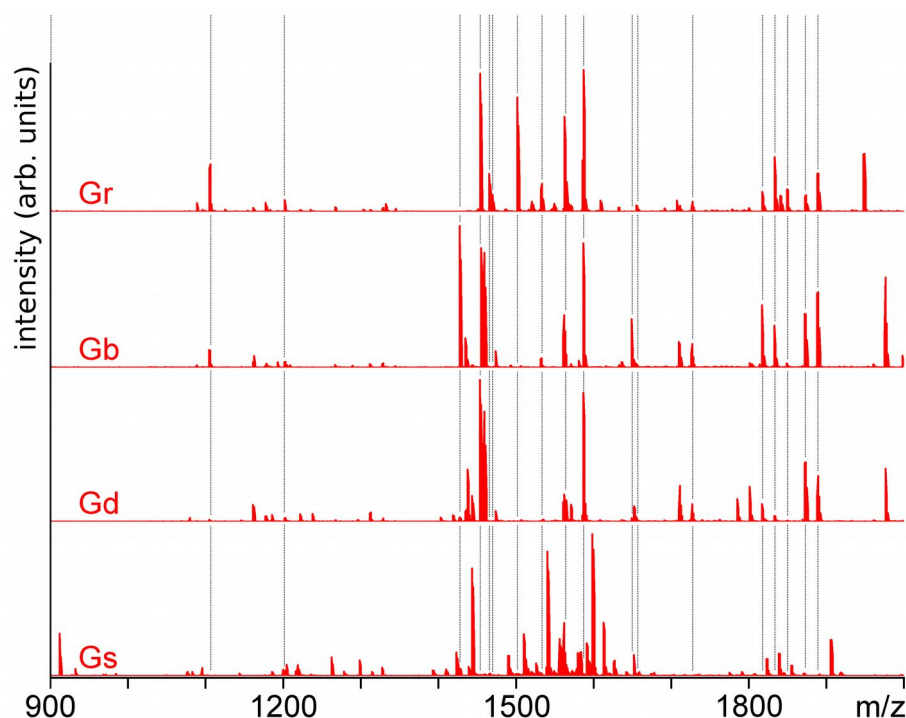


Figure V.3. MALDI-TOF spectra of different glues. The 19 characteristic peaks for the rabbit skin glue paint models are marked with dotted lines.

V.3.2 Species determination of animal glues

The same PCA procedure as described above can be used to evaluate the separation of animal glue samples originating from different animal sources. Figure V.4 shows the score plot of this second PCA experiment, using a series of measurements on rabbit skin (10), cow hide (10), deer hide (3) and sturgeon bladder (8) glues. In this PCA model PC1 accounted for 61% of the total variability. Yet it is mainly PC2 (14% of the variability) that separates the sturgeon glues from the mammalian glues in this model. The high spread in the bovine glue series, mainly along the PC1 axis, was largely due to datapoint 1428 (and its isotopes), present in all mammalian glues, but with significant differences in its relative height. No principal component was found to visually separate the different genetically closely related mammalian glues, restricting this PCA model to two groups: glues based on sturgeon and those based on a mammalian source. Thus, rabbit, cow

and deer glue cannot be discerned. Datapoints with high specificity for mammalian glues are 1428, 1454, 1565 and 1587, while for sturgeon glues are typified by datapoints 1313, 1470, 1598 and 1679. Because no sturgeon collagen sequences are known to the authors, annotations for those peaks could not be made.

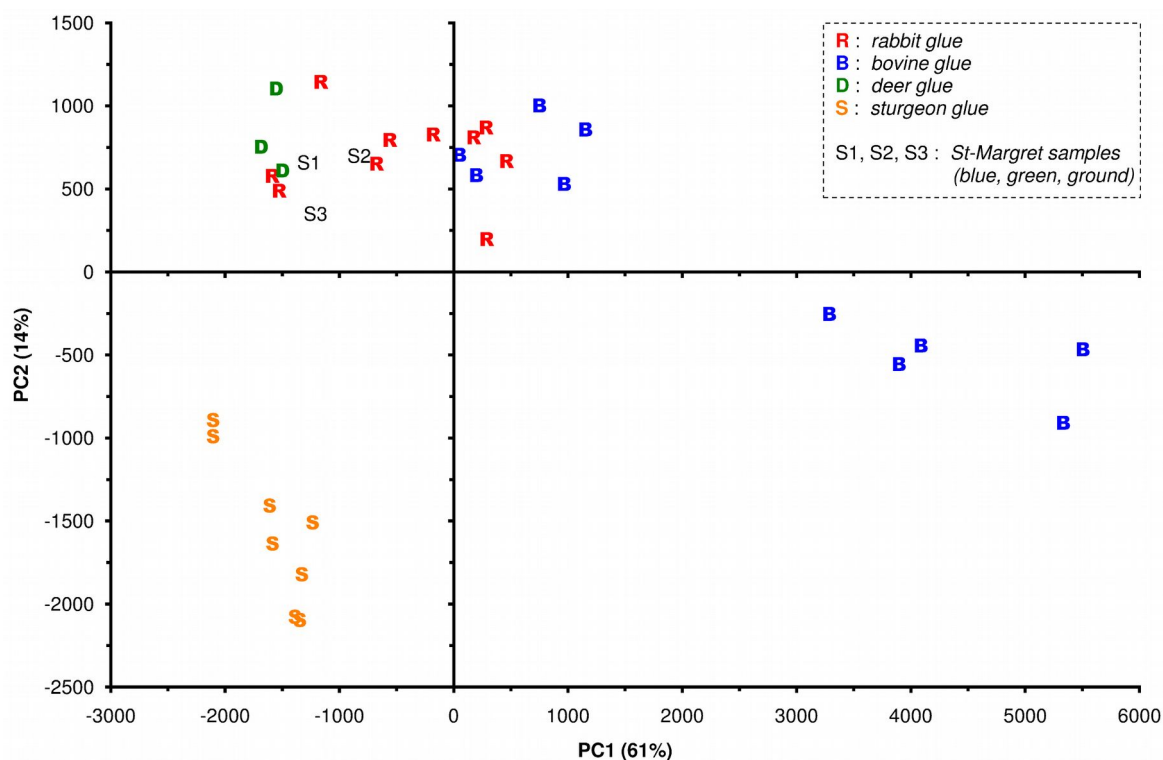


Figure V.4. Score plot of PC1 and PC2 visually separating different animal glues.

V.3.3 St Margaret of Antioch

Three samples were taken from the altarpiece of St Margaret of Antioch from the azurite blue (S1) and malachite green (S2) paints, as well as from the greyish chalk ground layer (S3). In all cases and sample S3 in particular, the mass spectra contain only few peaks and have a low signal-to-noise ratio. Their scores in the PCA model for the main protein binder classes (figure V.1, page 100) are therefore unsurprisingly biased. Although not visually taken up in the class of the animal glue models, they can be seen as an extension to this class, which is characterised by the large spread on the PC2 axis. The few peaks that are present are indeed characteristic markers for animal glue, as seen in figure V.5. The absence of some markers for animal glue in the St Margaret samples is probably the reason for the

shift in the PCA model. SIMCA, using the paint models as a training set, unambiguously classified S1 and S2 as animal glue, but fails for S3 (classified as both animal glue and egg yolk).

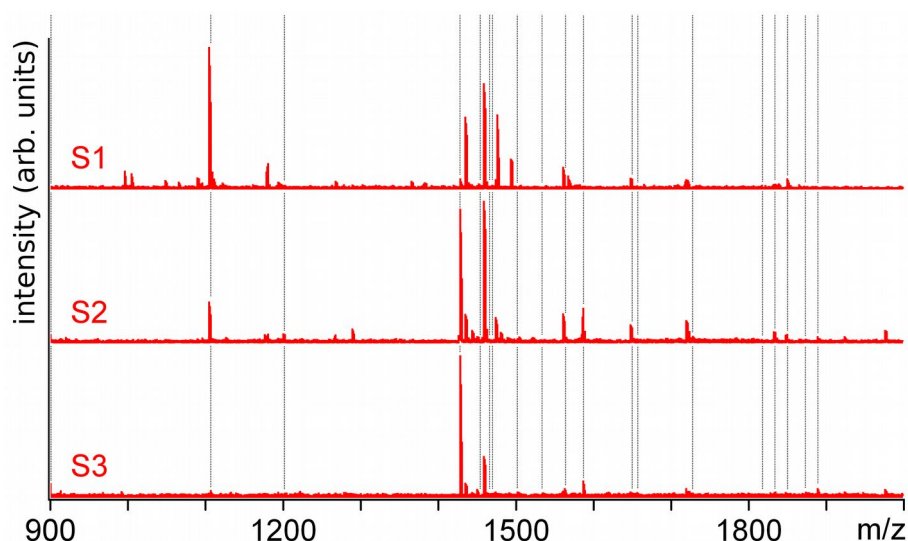


Figure V.5. MALDI-TOF spectra of the samples from the Altarpiece of St Margaret of Antioch. The markers for rabbit skin glue paint models are marked with dotted lines.

In order to determine the animal origin of the supposed glue in the St Margaret samples, the mass spectra of S1, S2 and S3 were also projected in the PCA model of the different animal glues (figure V.4). Herein they are situated within the mammalian glue group. Classification using SIMCA with all mammalian and sturgeon glues as a training set confirms these findings (table V.6).

Table V.6. SIMCA classification based on different animal glues.

Sample	Mammalian glue	Sturgeon bladder glue
S1	●	
S2	●	
S3	●	

Not all expected glue peaks are present in the St Margaret samples and yet others are, but much weaker. This suggests some forms of degradation processes in the aged paint layers and as a result some tryptic peptides are lost. A study with artificially aged paint samples would possibly allow mapping which peptides are

proof against ageing and could further improve the results of PCA and SIMCA with historical paint samples. On the other hand, MALDI-TOF-MS is not the first-rate technique to study degradation effects due to the rather complex mass spectra and limited mass information (in respect to HPLC-MS/MS).

V.4 Conclusions

Chemometrical methods prove to be an effective aid in the identification of proteinaceous binders using MALDI-TOF-MS of tryptic digests. The four most common groups of proteins to be used in historical paint (animal glue, egg white, egg yolk and milk casein) can readily be determined by both PCA and SIMCA. By using PCA on a broad set of model paint samples, the most meaningful peaks for differentiation are selected – nearly all were annotated with their amino acid sequence – while peaks originating from sample preparation, instrument and pigments are neglected.

Also, a study on animal glues from different species allowed for basic classification of the species origin of the glue. This, however, is restrained to sturgeon and mammal origins, due to the very limited evolutionarily changes mammalian collagens have undergone.

Finally, despite the low signal-to-noise ratio, the binding medium of the painted paper background from the altarpiece of St Margaret of Antioch was identified as animal glue, based on a mammalian origin. The results of PCA and SIMCA were, however, slightly biased due to natural ageing of the paint. The study of artificially aged models would likely allow further refining the proposed methodology.

The following paragraphs are not included in the published scientific paper:

While chemometrical data mining methods prove partially effective to determine the source of protein binders, the use of characteristic and unique markers for a given species may further refine the process of provenancing collagen. Buckley and Collins(80,81,159), and Kirby(82) both compile and successfully use libraries of such markers. In (single) mass spectrometric data, however, actually only the mass of the peptides is obtained. To verify possible markers, selective tandem mass spectrometry is often performed(160,161).

This chapter demonstrated both the use and shortcomings of single mass spectrometry for tryptic peptide analysis of paint samples. Failure is likely in case of unexpected proteins, heavily degraded samples or complex mixtures, as different peptides that share the same mass may thwart the interpretation. Nevertheless, several successful applications of MALDI-TOF-MS have been published since the publication of the paper upon which this chapter is based. Van der Werf *et al.* (84,162) developed and applied a combined protocol for both protein and oil binders, using amongst others the species-specific markers determined in this chapter. Romero-Pastor *et al.* (83) studied artificial ageing and pigment-binder interactions using MALDI-TOF-MS and PCA. Finally, Kuckova *et al.* (150) combined MALDI-TOF-MS measurements with staining tests on cross sections.

Potentially, far more information can be retrieved on the tryptic peptides by using tandem mass spectrometry: the fragmentation pattern of a single peptide may be sufficient to characterise unambiguously the protein from which it is derived. The powerful tandem mass spectrometry approach will be explored in depth in the following chapter.

VI

TRYPTIC PEPTIDE ANALYSIS OF PROTEIN BINDERS IN WORKS OF ART BY LIQUID CHROMATOGRAPHY TANDEM MASS SPECTROMETRY

*Based on the scientific paper:
W. Fremout, M. Dhaenens, S. Saverwyns, J. Sanyova, P. Vandenabeele, D. Deforce, L. Moens,
Anal. Chim. Acta 658 (2010) 156-162.*

Tandem mass spectrometry has become one of the cornerstones of modern proteomics. In combination with high-performance liquid chromatography, it allows us to identify tryptic peptides one by one through their unique fragmentation pattern. Opposite to this, in single mass spectrometry, as in the previous chapter, all peptides are measured simultaneously and only characterised by their mass. In this chapter, tandem mass spectrometry was used for the identification of protein binders in historical paints: the proteins were digested enzymatically into peptides using trypsin before being separated and detected by high-performance liquid chromatography electrospray ionisation tandem mass spectrometry (HPLC-ESI-MS/MS). Mascot (Matrix Science) was used to analyse the resulting data and for protein identification. The best extraction strategy was selected based on the number of peptides that were identified in the protein content of paint models using different methods. The influence of pigments on the extraction method was evaluated and the analytical characteristics of the selected method were

determined. Finally, the method was applied to historical paint microsamples of the anonymous early 15th century panel painting *Crucifixion with St Catherine and St Barbara (Calvary of the Tanners)*, the *St Catherine Altarpiece* by Joes Beyaert (ca. 1479) and two paintings by Pieter Brueghel the Younger (1617-1628).

VI.1 Introduction

The characterisation of painting materials, such as binding media, is crucial to their conservation and the study of the painting technique. Binding media are an important component of paint, determining its optical aspect and chemical properties. Ancient binding media include drying oils, proteins, resins, polysaccharides (gums) and waxes. Protein binders, based on animal glue (bone or skin glue), egg yolk, glair or milk casein, played an important role in early painting and are still used today. (163)

Several analytical techniques have been proposed for identifying protein binders, including microscopic staining(71), Fourier-transform infrared (FTIR)(142,164), near-infrared (NIRS)(56) and micro-Raman spectroscopy (MRS)(57–59), but these lack the ability to specify which protein source was used. Other, mainly chromatographic techniques are developed solely for protein identification and can discriminate between animal glue, egg and milk casein. Most of the cited chromatographic methods rely on complete hydrolysis of the protein matter into its amino acids (amino acid analysis) followed by derivatisation and separation/detection with gas chromatography coupled to mass spectrometry (GC-MS)(30,96,139) or with HPLC coupled to a fluorescence detector (HPLC-FD)(30). However, much information about the source of the protein binder, its degradation and the pigment-binder interactions is lost in this highly destructive hydrolysis step. The identification of the medium is based on treatment of the analytical results by principal component analysis (PCA) on the relative amounts of amino acids present. In the case of mixed binding media, identification becomes cumbersome. As an alternative, pyrolysis(36,145) can be used to split the macromolecules into very small fragments that are separated and detected by GC-MS. This approach gives rise to complex pyrograms that are difficult to interpret. The identification of the binders is based on the presence of marker fragments.

More recently, proteomics approaches(74,165,166) have been adapted for use on protein binders(75–77,85,146). Herein, proteins are digested into peptides by use of a protease, usually trypsin, which predominantly cleaves the protein at the carboxyl side of lysine and arginine (except when followed by proline)(167). These peptides are characteristic, specific and unique for a given protein: identifying them means identifying the protein as a whole and as such, the protein binder source it constitutes. HPLC coupled to a diode array detector (DAD, chapter IV) allows for an easy binder identification using chromatogram fingerprinting, but is unsuitable for the identification of complex mixtures since individual peptides cannot be identified due to the limited specificity of the DAD spectra. (Single) mass spectroscopy, as used by Kuckova *et al.*(75–77,146) with matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) allows for peptide mass fingerprinting: mass peaks are compared with a library of known peptide masses from reference samples. Since a peptide is only characterised by its mass, unambiguous peak attributions are not possible. Also, the whole sample mixture is measured in a single mass spectrum, which results in complex mass spectra and relatively low signal-to-noise ratios. More certainty in peak attribution can be attained with chromatographic separation and tandem MS, as shown by Tokarski *et al.*(85) on two egg tempera painting models using HPLC-ESI-MS/MS. In order to use this methodology in routine analysis, further study is needed, on one hand using a broader set of protein reference samples, on the other hand studying the impact of the presence of non-proteinaceous components, such as inorganic or organic pigments, on the results.

In the present study an analytical method is developed using tryptic digestion of the protein binders, separation with nano-HPLC and detection with ESI-MS/MS, avoiding as much as possible sample manipulation, which can lead to sample loss. To reduce the influence of the sample preparation on the results, different variants of the method are tested on a broad range of model paint samples (the most common protein binders and different kinds of historical pigments) in order to select the most appropriate method. Steps were taken to estimate and further increase the efficacy of the developed method. Finally, some microsamples from historical paintings were analysed.

VI.2 Experimental

VI.2.1 Reagents

Tris(hydroxymethyl)aminomethane hydrochloride (Tris), sodium docetyl sulphate (SDS) and phosphoric acid were acquired from Acros Organics (Geel, Belgium), ureum from Fluka (Bornem Belgium) and thiourea from Riedel-de Haën (Bornem Belgium). Dithiothreitol (DTT), iodoacetamide (IAA) and ammonium bicarbonate were supplied by Pierce (Perbio Science, Erembodegem, Belgium). Sequencing-grade modified trypsin was acquired from Promega (Leiden, the Netherlands). Ammonium acetate, ammonia solution (25%) and borax were supplied by Merck (Overijse, Belgium). MilliQ water was obtained using a Millipore system (Brussels, Belgium). Methanol, acetonitrile and formic acid were supplied by Biosolve (Valkenswaard, the Netherlands).

VI.2.2 Samples

Unpigmented reference binders were prepared as follows: rabbit skin glue (KIK/IRPA binder library, source unknown) was allowed to swell in water (8% w/w) for 24 h and subsequently heated to 50°C to completely dissolve the glue. Fresh eggs and soft curd cheese (quark cheese) were bought in a local supermarket. The egg glair and yolk were separated. The egg glair was stirred into a stable foam and allowed to stand for several hours. The reliquidised fraction was separated from the foam, and was used as egg white binder. The egg yolk was suspended above a recipient and the membrane was cut open. 25 g of soft curd cheese was washed with water on a Büchner funnel to eliminate the soluble gums and sugars. After adding 10 mL water and 1 g of borax (UCB) to the curds, the mixture was stirred for 40 min until a syrupy solution was formed, which was diluted (1:4) and used as borax casein binder.

Paint models were made using the above four binders and ten historical pigments: azurite ($\text{Cu}_3(\text{CO}_3)_2(\text{OH})_2$, KIK/IRPA pigment library, source unknown), lead white ($(\text{PbCO}_3)_2 \cdot \text{Pb}(\text{OH})_2$, KIK/IRPA pigment library, source unknown), red earth (Fe_2O_3 in a silicate matrix, Blockx, Terwagne, Belgium), chalk (CaCO_3 , KIK/IRPA pigment library, source unknown), chrome yellow (PbCrO_4 , Winsor & Newton, Harrow, UK), lead-tin

yellow type I (Pb_2SnO_4 , KIK/IRPA pigment library, source unknown), verdigris ($\text{Cu}(\text{OH})_2 \cdot (\text{CH}_3\text{COO})_2 \cdot 5\text{H}_2\text{O}$, Winsor & Newton), carbon black (C, Kremer Pigmente, Aichstetten, Germany), vermilion (HgS , Winsor & Newton) and Prussian blue ($\text{Fe}_7(\text{CN})_{18}(\text{H}_2\text{O})_x$ where $14 \leq x \leq 16$, Winsor & Newton). The sample matrix thus consists of 44 pigmented and unpigmented paint samples. The samples were stored in the dark at room temperature.

Microsamples were collected by gently scraping off a single paint layer using a scalpel. A sample was taken from the brown border of the anonymous early 15th century panel painting Crucifixion with St Catherine and St Barbara (Calvary of the Tanners, figure IV.2, page 82), which is displayed in the Cathedral of the Holy Saviour at Bruges. A second sample was taken from an adhesive layer for an applied relief of a polychromed statue of the Virgin, part of the St Catherine Altarpiece (figure VI.1) by Joes Beyaert (ca. 1479) and displayed at the St Leonard's Church in the small Belgian town of Zoutleeuw. Two ground layer samples were collected from two paintings by Pieter Brueghel the Younger: Sermon of St John the Baptist (1618-1627, figure VI.2), which is displayed in the Groeningemuseum at Bruges and Crucifixion (1617, with the cooperation of Joos de Momper, figure VI.3) in the Museum of Fine Arts of Budapest.



Figure VI.1. *St Catherine Altarpiece* (Joes Beyaert). (source: KIK/IRPA)



Figure VI.2. *Sermon of St John the Baptist* (Pieter Brueghel the Younger). (source: Christina Currie, KIK/IRPA)



Figure VI.3. *Crucifixion* (Pieter Brueghel the Younger and Joas de Momper). (source: Christina Currie, KIK/IRPA)

VI.2.3 Instrumentation and chromatographic conditions

The tryptic peptide mixtures were separated on an Ultimate 3000 Dual nano-HPLC system (Dionex), consisting of a DGP-3600MB dual ternary low-pressure proportioning micro pump, a WPS-3000TB nano pulled-loop thermostatted autosampler and a FLM-3100B Nano flow manager. The system is controlled via the Chromeleon software package. The solvents used are (A) 0.1% aqueous formic

acid and (B) 0.1% formic acid in 80:20 acetonitrile/water. A 20 μL sample was injected and concentrated in a PepMap C18 trap column (5 μm , 100 $\text{\AA}$, 300 μm I.D. x 5 mm from LC Packings) using a flow of solvent A. Separation of the peptides took place in a PepMap C18 capillary column (3 μm , 100 $\text{\AA}$, 75 μm I.D. x 150 mm from LC Packings) using a linear 45 min gradient at 300 nL min^{-1} in which the solvent B increased from 4% to 100%.

The HPLC was connected to a Waters Micromass ESI-QTOF Ultima mass spectrometer, which was operated in data-directed acquisition mode, with an MS survey mass range from 500 to 1150 Da and an MS/MS mass range from 50 to 1998 Da in 6 channels. The precursor charge state selection for MS/MS analysis was set at 2⁺-3⁺ with a minimum precursor Total Ion Count of 50. Data acquisition was done using MassLynx.

The data were subsequently subjected to the local running software package Mascot MS/MS ions search against the SwissProt sequence database (release 51.6). Carbamidomethylation of cysteine was chosen as fixed modification, *N*-terminal carbamylation, deamidated asparagine/glutamine and oxidation of proline and methionine (depending on the results) as variable modifications and the search was error-tolerant to correct against other possible modifications (or degradations). For a general search, the peptide mass tolerance was set at 0.3 Da, the fragment mass tolerance at 0.7 Da. When trying to define the species from which the animal glue originated, both were set at 0.3 Da. The maximum number of missed cleavages was set at one. The protein mass and taxonomy were unrestricted.

VI.2.4 Analytical procedures

To a 0.5 mL sample tube containing less than 100 μg paint chips, 50 μL of a solution containing 5 M urea, 2 M thiourea and 0.7% SDS (w/w) was added. Multiple short ultrasonic cycles helped the fragmentation and dissolution of the paint chip. 5 μL 110 mM DTT was added and the solution was let to reduce at 60°C for 60 min. Alkylation was performed by adding 5 μL 120 mM IAA; the solution was let to react at room temperature for 15 min.

440 μ L 50 mM Tris was added to dilute the SDS, urea and thiourea that would hamper the digestion. Subsequently, 10 μ L 50 mM ammonium bicarbonate was added containing 10 ng μ L⁻¹ sequencing-grade modified trypsin. The sample was digested overnight at 37°C. Afterwards, the samples were dried in a vacuum concentrator (Labconco Centrivap).

For sample purification and desalting the following solid phase extraction (SPE) products and procedures were tested:

- OASIS HLB 1cc cartridges (Waters, Brussels, Belgium): the hydrophilic-lipophilic balance (HLB) cartridge was conditioned using 1 mL methanol, equilibrated using 1 mL water, loaded with the sample in 0.5 mL water and washed using 1 mL 5% methanol in water. The peptides were collected by elution with 1 mL methanol.
- ICAT cation exchange kit with buffers provided (Applied Biosystems, Foster City, USA): the strong cation exchange (SCX) cartridge was conditioned using 1 mL of the included "clean" buffer, equilibrated using 2 mL "load" buffer, loaded with the sample in 0.5 mL "load" buffer and washed with 1 mL "load" buffer. The peptides were collected by elution with 0.5 mL "elute" buffer.
- Strata-C-X 30 mg cartridges (Phenomenex, Utrecht, the Netherlands): mixed-mode strong cation exchange cartridge was conditioned using 1 mL methanol, equilibrated using 1 mL water, loaded with the sample in 0.5 mL 2% phosphoric acid, washed with 1 mL 0.1% phosphoric acid and with 1 mL methanol. The peptides were eluted using 1 mL 5% ammonia in 1:1 acetonitrile/methanol.
- Glygen Tiptip hydrophilic 10-200 μ L (Bester, Amstelveen, the Netherlands): the releasing buffer was 15 mM ammonium acetate (pH 3.5) in 10% acetonitrile, the binding buffer 15 mM ammonium acetate (pH 3.5) in 85% acetonitrile. The hydrophilic interaction chromatography (HILIC) cartridges were conditioned using 3x 50 μ L releasing buffer, equilibrated using 3x 50 μ L binding buffer, loaded with the sample in 50 μ L binding buffer and washed

using 3x 50 μ L binding buffer. The peptides were eluted using 2x 50 μ L releasing buffer.

After evaporation in the vacuum concentrator, the samples were redissolved in 23.5 μ L 0.1% formic acid.

VI.3 Results and discussion

VI.3.1 Development of the sample preparation

Dissolving centuries old, degraded and crosslinked protein binders might be the most challenging task in adapting a standard proteomics technique for identifying proteins in works of art. Particularly good results were obtained using a combination of fragmentation in an ultrasonic bath and the use of strong denaturing agents. SDS targets the non-covalent bonds of the proteins and causes them to lose their natural conformation. To displace the anionic detergent from the proteins and replace it by uncharged detergents, high concentrations of urea and thiourea are used. Disulphide bonds, which play an important role in protein folding, are reduced by DTT and irreversibly alkylated by IAA. To prevent the SDS and urea from denaturing trypsin itself, their concentrations are lowered to well below the critical concentration by tenfold dilution, after which trypsin is let to hydrolyse the protein into tryptic peptides. After extraction of the latter using two sequential solid phase extractions (Oasis HLB and ICAT cation exchange, see below), they are separated using reversed-phase nano-HPLC and measured using QTOF-MS/MS. That way, a characteristic tandem MS spectrum of each eluting peptide was generated. The results were analysed with the Mascot search engine.

This methodology, summarised in figure VI.4, was applied on the common proteinaceous binding media (rabbit skin glue, chicken egg white, chicken egg yolk and bovine milk casein), as well as on a broad range of model paint samples. Ten historical inorganic pigments were carefully selected to include a broad range of different cations and anions. Mixed with the different protein binders, their potential influence on the sample preparation could be examined. Also, to define the most suitable sample preparation method, several sample preparation variants were compared (data not shown), multiplying the number of experiments.

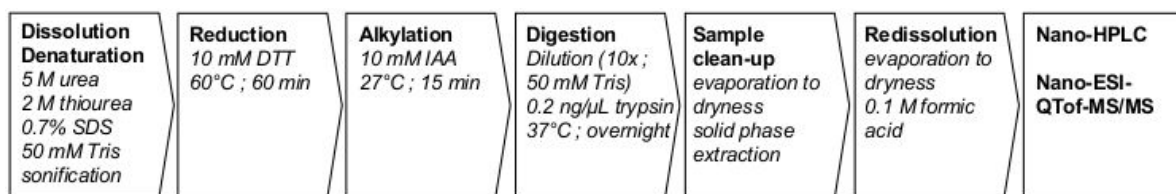


Figure VI.4. Scheme of the sample pretreatment.

The fragmentation of the peptides in the collision cell mainly cleaves the amidobonds, giving rise to y - and b -ions(168). Figure VI.5 A shows the MS/MS spectrum of the rabbit collagen α -2(I) peptide HGNRGEPGPAGSIGPVGAAGPR (with oxidated proline at position 7) found in the animal glue samples as annotated by Mascot. This spectrum shows the annotation of single charged mass fragments y_2 - y_6 , y_8 - y_{16} , b_2 , b_6 - b_8 , b_{10} - b_{14} and the double charged fragments y_8^{2+} - y_9^{2+} , y_{20}^{2+} - y_{21}^{2+} , b_3^{2+} , b_6^{2+} , b_8^{2+} - b_{14}^{2+} , b_{16}^{2+} . Furthermore, some peaks can be attributed to dehydrated b -fragments (b_{12}^0 , b_{12}^{02+} , b_{13}^{02+}) and deaminated y -fragments (y_2^* - y_3^* , y_9^* , y_{15}^* , y_{15}^{*2+}). In this particular case the mass spectrum is identified with a very high-confidence MOlecular Weight SEarch (MOWSE) score of 112, which in our search corresponds to an error rate of e^{-16} .

Figure VI.5 B-D illustrates the more straightforward fragmentation patterns that are found in most non-collagen samples, in which a clear y -ion series makes out the majority of the most intense peaks.

The presence of pigments in paints can drop the efficiency of the sample pretreatment, or even obstruct the analysis. 44 pigmented and unpigmented unaged models were analysed using the above method to rule out pigment influences. Table VI.1 shows the successful determination of the binders in all but two model samples (estimated sample size between 20 and 100 μ g). If at least three peptides are significantly identified, the protein binder is considered to be determined. Only in two cases of the animal glue reference samples the significant determination of the binder cannot be assured. The number of significant identified peptides, however, varies due to several factors other than the influence of pigments: firstly, the sample size can cause a substantial variability since the samples were too small to be weighed. Secondly, the measurements have been run in a longer period of time, in which instrumental conditions and the sensitivity involved have altered (column replacements...). Therefore, the exact reason for

the low number of significant identified peptides in the two failed analyses can possibly be attributed to sensitivity issues rather than to the influence of pigments. In general, we conclude that pigments do not influence the sample preparation to such a degree that interpretation becomes impossible.

Table VI.1. Determination of the binding media in the models. A protein is considered identified if at least three significant peptides could be identified.

Pigment	Binding medium			
	Animal glue	Egg white	Egg yolk	Milk casein
Unpigmented	●	●	●	●
Azurite	●	●	●	●
Lead white	●	●	●	●
Red earth	●	●	●	●
Chalk	●	●	●	●
Chrome yellow		●	●	●
Lead tin yellow	●	●	●	●
Verdigris	●	●	●	●
Carbon black	●	●	●	●
Vermillion	●	●	●	●
Prussian blue		●	●	●

Most analysed peptides show some degree of modification. Cysteine is exclusively found in the form of carbamidomethyl cysteine because of the intentional reaction of the cysteine *SH*-groups with IAA. This modification is thus chosen as a fixed modification in Mascot, which adds a mass of 57 Da to all cysteine groups. Other modifications can be expected either from *in vivo* processes or due to the experimental procedure, but are not a rule of thumb and were thus chosen as variable modifications. These are *N*-terminal carbamylation, deamidated asparagine and glutamine, and oxidated proline. The first is a result of a reaction with isocyanic acid, a breakdown product of urea. The latter is a post-translational modification found in collagen chains in which some of the prolines are converted into hydroxyproline, involving prolyl hydroxylase enzyme and playing a key role in collagen production and stability. Oxidation of methionine can occur both *in vivo* and *in vitro* and is often checked during routine protein identification searches(169). Unexpected or less frequent modifications can be included in the Mascot search results by enabling error-tolerant searches. In this case, however,

the number of selected variable modifications is restricted to two, by reasons of search algorithm; best search results may be obtained by trial and error.

These series of analysis allow us to identify the proteins present in the paint models. When running blanks, several types of human keratin and bovine trypsin, despite its very low concentration, could be identified. Keratins are almost unavoidably introduced by sample handling and will be ignored in this study. A list of the most commonly identified proteins in the paint models is shown in table VI.2. In the animal glue samples different collagen chains were found: α -1(I) (COL1A1) and α -2(I) (COL1A2) α -1(II) (COL2A1), α -1(III) (COL3A1) and α -1(IV) (COL4A1) were found in respectively 91%, 91%, 36% and 36% of the animal glue samples. The most abundant proteins in the egg white paint samples were found to be ovalbumin, lysosyme C and ovotransferrin. In the egg yolk paint samples the main proteins are vitellogenin-2, vitellogenin-1 and apovitellenin-1. Chicken serum albumin (when found in egg commonly designated as α -livetin) was found in few yolk-based models. The egg white protein ovotransferrin was also present in the yolk samples, but to a much lesser extent. This could be caused by a contamination during the preparation of the yolk binder, in which the yolk membrane is cut open and the content is let to pour out. Also, some egg white proteins like ovotransferrin are known to migrate through the membrane in older eggs(170,171), which could be a second explanation. It remains however possible to clearly distinguish between glair, yolk and mixed egg binders, confirming the findings of Tokarski *et al.*(85). Milk casein binder was found to contain mainly α _{s1}-casein, α _{s2}-casein, β -lactoglobulin, β -casein and κ -casein, in accordance with the known substances of fresh milk.

Table VI.2. The most common identified proteins in the paint models. The percentage indicates the number of samples in which the given protein was found.

Animal glue	Egg white	Egg yolk	Milk casein
Collagen α -1(I)	91%	Ovalbumin 100%	Vitellogenin-2 100%
Collagen α -2(I)	91%	Ovotransferrin 91%	Vitellogenin-1 83%
Collagen α -1(II)	36%	Lysozyme C 82%	Apovitellenin-1 17%
Collagen α -1(III)	36%	Ovalbumin-related protein X 82%	Serum albumin 8%
Collagen α -4(IV)	18%	Riboflavin-binding protein 18%	Ovotransferrin 8%
		Ovalbumin-related protein Y 9%	α _{s2} -casein 9%
			Lactadherin 9%

Although species determination was not within the objectives of this chapter, some observations can be made. Species determination depends on mutations of the protein chains between the homologous proteins of the different animals. The protein sequence databases in their present form, in particular the SwissProt database used here, are far from complete: even though most proteins are represented, this is certainly not the case for all species. Rabbit collagens, for example, are only partly known (less than 4% and 39% of COL1A1 and COL1A2 chains, other relevant rabbit collagen sequences are missing in SwissProt). Identification thus depends on the presence of some of the peptides in these known regions; other peptides cannot be used for species determination. Mascot search results consequently identifies not only the rabbit collagen peptides but also a broad range of collagen peptides that are common or highly similar to those of other species. Moreover species determination of glues is further complicated by collagen's evolutionarily relatively well preserved sequence and its repetitive patterns. Consequently, identification of exclusive rabbit collagen fragments is indicative for rabbit glue. On the other hand, when no exclusive rabbit collagen fragments are found, rabbit glue cannot be excluded. More complete databases of rabbit protein sequences and better mass resolution instruments will probably offer new opportunities for this species-specific approach in the near future. It is easier to determine the species from which the egg and milk proteins come from, although less relevant as there is no evidence of the use of eggs or milk from other species than respectively chicken and cows.

VI.3.2 Optimisation of the clean-up and instrumental set-up

SDS, urea and thiourea used during the first reaction steps need to be removed before analysis. In the first experiment the urea and thiourea were successfully removed using the Oasis HLB SPE tubes, but the removal of SDS was unsuccessful. The latter was subsequently removed using the ICAT cation exchange cartridge. The combination of both clean-up steps, however, results in a considerable loss of annotated peptides. Three alternative methods for the removal of SDS, urea and thiourea were thus compared using the unpigmented milk casein model. A 100 µg sample was prepared using the above method; after the digestion the sample was divided into three equal parts for usage with ICAT cation exchange (solely),

Strata-X-C polymeric mixed-mode strong cation exchange and TopTip HILIC solid phase extraction.

The sequence recoveries of the different proteins in the milk casein sample are considered a measure for the efficiency of the different clean-up methods. The averaged sequence recoveries for the five most abundant milk proteins are listed in table VI.3. The highest recoveries in all five cases are obtained with HILIC solid phase extraction, while both strong cation exchange methods are comparable.

Table VI.3. Sequence recoveries (mean) in the unpigmented milk casein models with different clean-up strategies.

Protein	Clean-up method		
	ICAT SCX	Strata-C-X SCX	TopTip HILIC
β -lactoglobulin	2.3%	4.6%	65%
α_{s1} -casein	11%	13%	37%
β -casein	7.0%	6.3%	26%
κ -casein	4.6%	3.0%	44%
α_{s2} -casein	0.0%	0.0%	29%

The HPLC- and MS-methods were optimised as well: the use of a longer column of 250 mm and a slower gradient eluting the peptides by a 90 min linear gradient from 6 to 100% buffer B, further increases the number of identified peptides (data not shown).

VI.3.3 Determination of the required sample size

In literature, little information is found on the actual sample size needed by classical amino acid analysis, due to the difficulty of weighing these extremely small samples. Generally, 100 μ g of paint is considered a common value (see chapter IV)(30). To minimise the damage that is inevitably caused to the art object, the sample size should be further decreased as much as possible.

To estimate the minimal sample size needed for the proposed method, three samples containing 50-100 μ g of milk casein reference sample – the lowest amounts reliably weighable on our balance – were prepared and digested in accordance with the described method. However, before the SPE clean-up procedures, thus at the peptide level to minimise intersample variance caused by the digestion procedure, dilution series were made containing 0.005 μ g to 50 μ g of

the original sample masses. Next, the dilution series samples were cleaned up and analysed. The number of identified significant peptides was plotted against the sample mass (figure VI.6). Even in the lowest concentration, 0.005 μg of the milk casein sample, one peptide was identified: LSFNPTQLEEQCHI with a MOWSE score of 81. Generally, to attain a high confidence level three significant peptides should be identified. It is, however, important to note that unpigmented samples were used: we expect the number of identified significant peptides to be lower in aged paint samples of the same weight. Therefore, a threshold of ten significant peptides was chosen; this is accomplished using samples of 0.5 μg or more. Sample masses higher than 5 μg do not seem to increase the number of identified peptides as most proteotypic peptides are identified at that point. Assuming a good quality protein-bound paint contains 5 to 10% of proteins(164), a real paint sample of 5 μg would suffice.

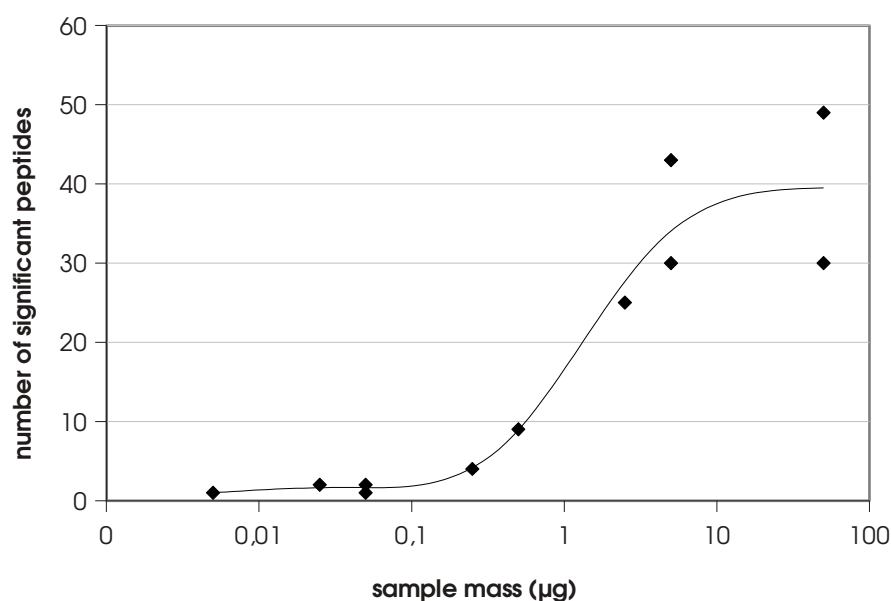


Figure VI.6. The number of identified significant peptides for decreasing sample mass of an unpigmented milk casein sample.

VI.3.4 Proteins in historical samples

The microsamples, too small to be weighed, were subjected to this sample treatment. A Mascot search was performed against the SwissProt database. In all

cases, best results were obtained when oxidation of proline was selected as variable modification (since all samples contained collagens). Other, less frequent modifications were subsequently found in the error-tolerant mode.

In the pre-Eyckian *Crucifixion with St Catherine and St Barbara* three different collagen chains were identified: COL1A1, COL1A2 and COL3A1, unambiguously demonstrating the presence of animal glues. For the first, the total number of identified peptides for dog, bovine and mouse COL1A1 is ten, most of them being shared between the different species. Also human, rat and chicken COL1A1 are presented in the Mascot search results, yet without unique peptide identifications. Particularly high MOWSE scores were reached for GFPGADGVAGPKGPAGER (with oxidated proline at position 3, found in dog, human and bovine sources, MOWSE score 93, figure VI.7 A), SGDRGETGPAGPAGPIGPVGAR (with valine at position 19 substituted with an alanine, found in bovine, MOWSE score 90) and VGPPGPSGNAGPPGPPGPVGK (with oxidated proline at positions 4, 13 and 15, found in mouse and rat, MOWSE score 110, figure VI.7 B). Two (dog) COL1A2 peptides were identified as well as one (mouse) COL3A1 peptide. Based on these results, and taking into account the limited number of different species collagens in the SwissProt database, reliable determination of the species is not possible. One of the most used sources for animal glue is rabbit skin, but only fragments of these collagens are known. Without doubt, a dispersed result like this might indicate the use of a species with no or only limited collagen coverage in the database.

The Mascot search results on the glue microsample from the *St Catherine Altarpiece* identified collagen chains COL1A1, COL1A2, COL2A1 and COL3A1, again demonstrating the use of an animal glue. Here, a bovine origin turns out to be the most convincing animal source. Ten peptides of bovine COL1A1 were identified, of which respectively eight and six are shared with dog and human collagen. The two others, GEPGPAGLPGPPGER (with oxidated proline at positions 3, 9 and 12; MOWSE score 112, figure VI.8) and SGDRGETGPAGPAGPIGPVGAR (MOWSE score 77) are only found in bovine glue. Furthermore, seven bovine COL1A2 peptides were identified, of which four are shared with dog and human collagen. Additionally, not a single non-bovine collagen peptide was identified. This is an indication for a bovine glue.

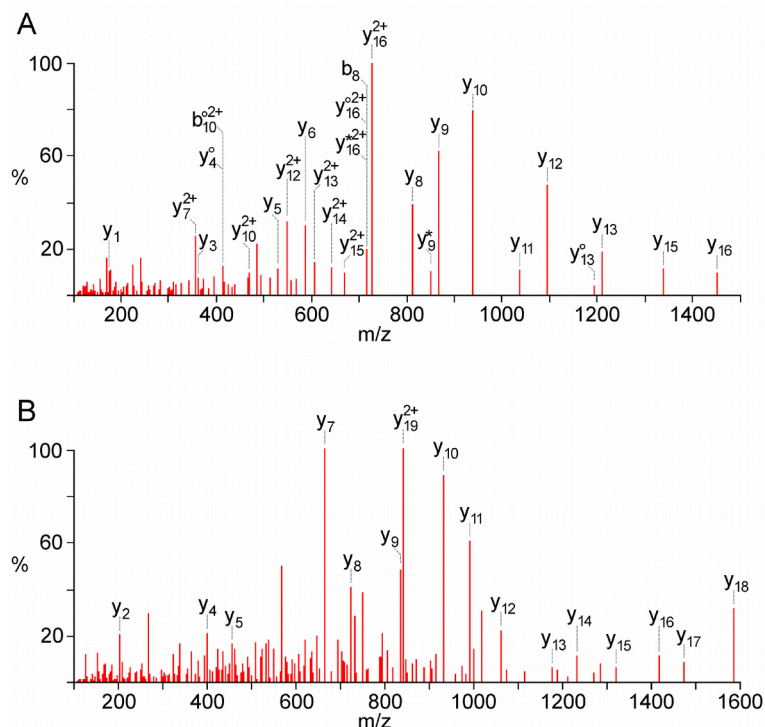


Figure VI.7. MS/MS spectra from the Crucifixion with St Catherine and St Barbara microsample, identifying (A) GFPADGVAGPKGPAGER and (B) VGPPGPSGNAGPPGP-PGPVGK.

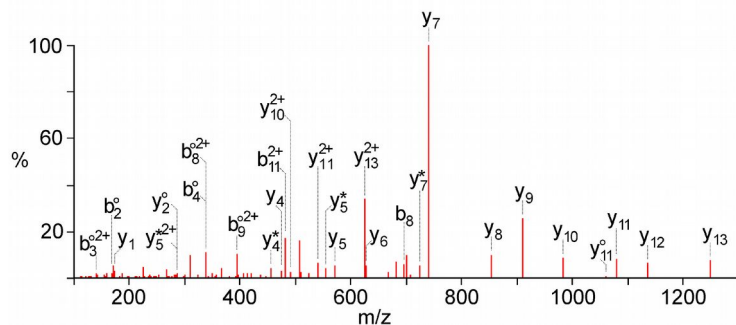


Figure VI.8. MS/MS spectrum from the St Catherine Altarpiece microsample, identifying GEPGPAGLPGPAGER.

The microsamples taken from the two Brueghel paintings were extremely small. As a consequence, only few peptides could be identified. In *Sermon of St John the Baptist*, GSAGPPGATGFPGAAGR (with oxidated proline at positions 6 and 12, MOWSE score 94, figure VI.9) and SGRGETGPAGPAGPIGPVGAR (with valine at position 19 substituted with an alanine, MOWSE score 73) indicate bovine COL1A1, while GIPGPVGAAGATGAR (with oxidated proline at position 3, MOWSE score 81)

indicates bovine or dog COL1A2. In *Crucifixion* only GVQGPPGPAGPR (MOWSE score 47, figure VI.10) could be identified, a peptide occurring in bovine, dog, human, mouse and rat COL1A1. This supports the use of animal glue in both cases. However, with so few identified peptides the species cannot be determined reliably.

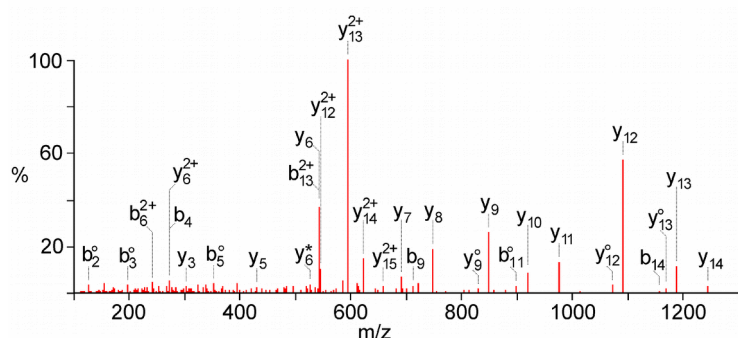


Figure VI.9. MS/MS spectrum from *Sermon of St John the Baptist* microsample, identifying GSAGPPGATGFPGAAGR.

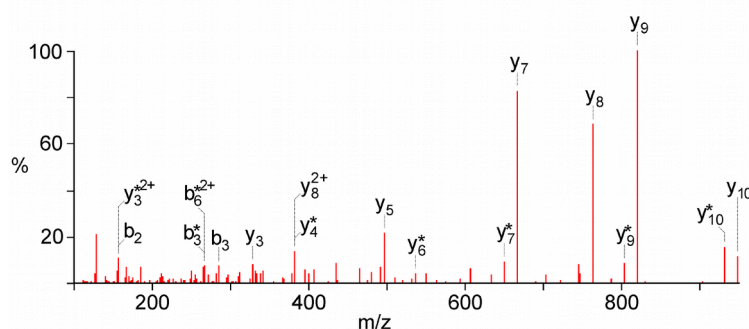


Figure VI.10. MS/MS spectrum from *Brueghel's Crucifixion* microsample, identifying GVQGPPGPAGPR.

VI.4 Conclusions

Analysing tryptic peptides using HPLC coupled to tandem MS proves to be an exquisite choice for the determination of binding media in works of art. The proposed sample preparation method has been developed and optimised for a broad range of binders and with a minimal influence of pigments. In contrast to the classical methods, the novel approach is able to distinguish between egg white and yolk. Because the proteins itself are identified, binary mixtures of proteinaceous binders can be identified next to each other. Due to limitations in

the sequence databases and the way Mascot works and due to the nature of collagen itself, in some cases only indicative species determination of animal glues is possible.

The developed methodology was successfully applied on microsamples from several historical paintings and a polychromed sculpture: in all cases animal glues were identified. In the anonymous pre-Eyckian painting *Crucifixion with St Catherine and St Barbara* three different collagen proteins were identified, but the results indicate that a species might have been used that is not (fully) covered in the sequence databases. The glue in the *St Catherine Altarpiece* has probably of bovine origin, while in the two paintings by Pieter Brueghel the Younger too few peptides were found to draw conclusions about the source species.

The following paragraphs are not included the published scientific paper:

The success of tandem mass spectrometry for the identification of proteins in art and archaeological artefacts is proven by the numerous papers that were published during and after the paper upon which this chapter is based (87–90, 152, 154–156, 160, 161, 172–174).

The tandem MS measurement data most often contain vital clues to the species used to prepare animal glues in the cases described in this chapter, but Mascot is not able to interpret them correctly, due to incomplete sequence libraries. Therefore, valuable information is overlooked or even false conclusions can be made. With the current progress in biotechnology and its analysis software, the species-specific determination of animal glues will steadily improve, even without changing the proposed protocol. However, we believe that using an alternative and complementary data interpretation method based on tandem mass spectral libraries can already lead to more accurate results. Hence, the construction of a dedicated spectral library will be discussed in the following chapter.

VII

A DEDICATED PEPTIDE TANDEM MASS SPECTRAL LIBRARY FOR CONSERVATION SCIENCE

*Based on the scientific paper:
W. Fremout, M. Dhaenens, S. Saverwyns, J. Sanyova, P. Vandenabeele, D. Deforce, L. Moens,
Anal. Chim. Acta 728 (2012) 39-48*

In recent years, the use of liquid chromatography tandem mass spectrometry (HPLC-MS/MS) on tryptic digests of cultural heritage objects has attracted much attention. It allows for unambiguous identification of peptides and proteins, and even in complex mixtures species-specific identification becomes feasible with minimal sample consumption. Determination of the peptides is typically based on theoretical cleavage of known protein sequences and on comparison of the expected peptide fragments with those found in the MS/MS spectra. In this "sequence library search" approach, complex computer programs, such as Mascot MS/MS ions search, perform well identifying known proteins, but fail when protein sequences are unknown or incomplete. Often, when trying to distinguish between evolutionarily well preserved collagens from different species, Mascot lacks the required specificity. Complementary and often more accurate information on the proteins can be obtained using a reference library of MS/MS spectra of species-specific peptides. Therefore, a library dedicated to various

sources of proteins in works of art was set up, with an initial focus on collagen rich materials. This chapter discusses the construction and the advantages of this spectral library for conservation science, and its application on a number of samples from historical works of art.

VII.1 Introduction

Proteins are an important constituent in many art and cultural heritage objects. Notable protein sources are paint binding media(68,85,88) (mainly egg white, egg yolk, animal glue and milk casein), adhesives (mainly animal glues and milk casein), parchment and leather(78), textiles (wool, silk), mortars(151) and archaeological remains(86,89). The knowledge of the materials used in a work of art is crucial, not only to understand degradation processes taking place in aged objects and to develop appropriate conservation treatments, but it also gives an insight in the art-historical context of the object and artist.

Until recently, amino acid analysis was the preferred technique to identify proteins. Herein a microsample containing the proteinaceous matter is hydrolysed into its constituting amino acids, which are subsequently derivatised to enhance detection in either high-performance liquid chromatography (HPLC)(30) or gas chromatography coupled to mass spectrometry (GC-MS)(30,136). This way, protein sources with different amino acid compositions can readily be identified. However, egg white and yolk – in paint binders both used separately or in a mixture – have highly similar amino acid profiles and cannot reliably be distinguished on this basis. Also, mixtures of protein sources and microbiological contamination of samples quickly add to the complexity of the analysis, leading to amino acid profiles that do not correlate to any of the references.

In recent years, the analysis of large protein fragments, peptides, has become increasingly popular for the characterisation of proteinaceous materials in cultural heritage objects. Herein the proteins are enzymatically digested into peptides, typically using trypsin. Because of their ability to unambiguously identify peptides, these techniques allow to obtain the most detailed information as yet about the individual proteins that are present in cultural heritage samples, while consuming

very limited amounts of sample, typically less than 10 µg (section VI.3.3, page 127). In complex protein mixtures and contaminated samples, as frequently encountered in art objects, different sources of proteins can be easily discriminated. Two techniques are commonly applied to measure the tryptic peptides: peptide mass fingerprinting (PMF) and tandem mass spectrometry (MS/MS or MS²).

In PMF all tryptic peptides are represented by corresponding peaks in a mass spectrum, which thus acts as a fingerprint of the protein(s). This technique has been discussed in several publications devoted to conservation science and makes use of matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS, chapter V)(75,78,80–82,151). While straightforward and relatively fast, the peptides are only characterised by their mass. The identification is based on comparison of these masses with those obtained on reference samples. This way PMF, originally only used on isolated proteins, can be used on simple mixtures as is the case in protein sources in art objects. Determination becomes harder in case of complex mixtures and contaminated samples. This approach is also successfully applied for species determination, even for the evolutionarily conserved collagen that is found in animal glue (section V.3.2, page 107)(80,82).

MS/MS significantly extends the information obtained on each detected peptide ion. The fragmentation of the peptides occurs mainly along the peptide bonds; tandem mass spectra are hence a fingerprint of the individual peptides and give essential information on the peptide sequence. In conservation science, MS/MS analysis of tryptic digests is usually performed either using MALDI-MS/MS(81) or HPLC-MS/MS(85,87–90).

The common way in proteomics to interpret tandem mass spectrometry data is matching the observed spectra with predicted masses for many given peptide sequences. These known peptides can be predicted by theoretical tryptic cleavage of known protein sequences, as found in sequence libraries such as UniProt, or from translated genetic code. This approach, using software packages such as Mascot, SEQUEST and X!Tandem, generally works out for distinguishing between the main protein sources used in art, as seen in the previous chapter. In

addition, successful species determination of casein-bound paints originating from milk from different species has been reported(88), but this is obviously limited to the protein sequences of the species that are present in the sequence libraries. Species determination of collagen (animal glue, leather) is, however, hampered by several factors. Firstly, a much wider pool of plausible species can be expected for the production of glue or leather, compared with egg or milk based binders or glues. Secondly, existing sequence libraries contain collagen sequences from only a limited number of species. The protein sequences of some species of particular interest to conservation science are either missing completely or covered only partly. Finally, due to the high degree of evolutionary conservation of collagen and its repetitive arrangement, ion search algorithms often get confused, providing false positive identifications or multiple possible annotations with equally high match scores.

In contrast to spectra prediction, traditional spectral library searches as applied for decades in analytical chemistry, did not find much use in peptide identification due to the limited knowledge of the proteome. Nowadays, however, more and more proteomics projects (X!Hunter, NIST, SpectraST) aim to establish tandem mass spectral libraries of peptides (annotated spectrum libraries, ASL)(135), which are believed to quickly yield better, more confident matches for single spectra. Even though these projects have released several public libraries, they all have a strong focus on current biotechnological issues and, as such, none of them suits the specific needs for conservation science.

Nonetheless, spectral libraries hold the promise of tackling the drawbacks observed in the spectra prediction approach, especially for species differentiation of collagens. In this paper the development of a spectral library dedicated to conservation science is discussed. At this point, our focus is on animal glues prepared from different animal sources, although other proteins commonly found in works of art are included as well. A large number of reference samples is measured and the recorded MS/MS spectra are analysed. Identifying reliable marker spectra for the different glues and other proteinaceous matter represents a challenge: spurious spectra from trypsin autodigest products, keratin contamination and other sources are expected in the reference materials measurements. A well-considered selection of candidate markers is therefore based on analysis of a large

number of samples of different sources over a prolonged time span to counter system variability. Moreover, using Mascot analysis on the same data set many of these candidate markers could be verified and annotated. The spectral library discussed in this paper will be submitted to the mass spectral library of the User's Group for Mass Spectrometry and Chromatography (MaSC).

VII.2 Experimental

VII.2.1 Reagents

Tris(hydroxymethyl)aminomethane hydrochloride (Tris), sodium dodecyl sulphate (SDS) and phosphoric acid were acquired from Acros Organics (Geel, Belgium), urea from Fluka (Bornem, Belgium) and thiourea from Riedel-de Haën (Bornem, Belgium). Dithiothreitol (DTT), iodoacetamide (IAA) and ammonium bicarbonate were supplied by Pierce (Perbio Science, Erembodegem, Belgium). Sequencing-grade modified trypsin was acquired from Promega (Leiden, the Netherlands). Ammonium acetate and ammonia solution (25%) were supplied by Merck (Overijse, Belgium). MilliQ water was obtained using a Millipore system (Brussels, Belgium). Methanol, acetonitrile and formic acid were supplied by Biosolve (Valkenswaard, the Netherlands).

VII.2.2 Samples

Reference animal glue samples

Six reference samples were taken from the binding media collection at the Royal Institute for Cultural Heritage (KIK/IRPA), but have no known supplier or preparation date: cattle glue 838 (KIK/IRPA's product number), rabbit skin glues 581 and 588, Siberian sturgeon glues 841 and 843 and deer glue 840. Five products were acquired in 2011 from Kremer Pigmente (Aichstetten, Germany): cattle hide glue 3359, cattle bone glue 3360, rabbit skin glue 3361, sturgeon bladder glue 3362 ("Salianski isinglass") and pig based gelatin 3363.

A set of 16 animal glues was prepared between 2006 and 2010 and generously donated by Julia Schultz (HAWK University of Applied Sciences and Arts, Hildesheim, Germany). Three Siberian sturgeon glues, two carp glues, a trout glue, a rabbit skin

glue, two hare glues, cattle bone and hide glues, two pig bone glues and a deer hoof glue were prepared from fresh animal materials. Also glues based on calf, pig and goat parchment were prepared. The fresh materials and parchments were boiled for 2 h in water. The solution was then strained through a linen cloth and left over night in the fridge, so that the hydrophobic phase could be decanted as much as possible. The glue was liquified by slowly heating, poured on foil and finally dried in the oven at 50°C.

Historical samples

Two samples were taken from a 19th century anatomical model of an ear (figure VII.1) from the *Musée de la Médecine* of the *Université libre de Bruxelles* (ULB, Brussels, Belgium). Sample E1 was taken from a fragment of a transparent glue in between two sheets of papier-mâché, while sample E2 was taken from the papier-mâché itself.



Figure VII.1. An anatomical model of an ear.

The ground layers of two paintings by Pieter Brueghel the Younger were previously studied in section VI.3.4 (page 128), and was shown to be animal glue-bound, but the species could previously not be determined. Sample B1 was taken from *Sermon of St John the Baptist* (1618-1627, figure VI.2, page 118) from the

Groeningemuseum (Bruges, Belgium). Sample B2 was taken from *Crucifixion* (painted in cooperation with Joos de Momper in 1617, figure VI.3, page 118), kept in the Museum of Fine Arts (Budapest, Hungary). The same measurement data were used in this study, using a different data treatment approach.

VII.2.3 Analytical procedures

The sample pretreatment has been extensively described in section VI.2.4, page 119 and will be summarised in what follows. Less than 100 µg sample was dissolved in 50 µL of a freshly prepared solution containing 5 M urea, 2 M thiourea and 0.7% SDS (w/w). Multiple short ultrasonic cycles helped the fragmentation and dissolution of the sample. 5 µL 110 mM DTT was added and the solution was let to reduce at 60 °C for 60 min. Alkylation was performed by adding 5 µL 120 mM IAA; the solution was let to react at room temperature for 15 min. The solution was diluted with 440 µL 50 mM Tris, and subsequently 10 µL 50 mM ammonium bicarbonate was added that contains 10 ng µL⁻¹ sequencing-grade modified trypsin. The sample was digested overnight at 37 °C. Afterwards, the samples were dried in a vacuum concentrator.

Sample purification and desalting was performed with miniaturised solid phase extraction (SPE) using Glygen Toptip hydrophilic 10–200 µL pipette point tips (Bester, Amstelveen, the Netherlands). The SPE cartridges were conditioned using 3× 50 µL releasing buffer (15 mM ammonium acetate (pH 3.5) in 10% acetonitrile), equilibrated using 3× 50 µL binding buffer (15 mM ammonium acetate (pH 3.5) in 85% acetonitrile), loaded with the sample in 50 µL binding buffer, and washed using 3× 50 µL binding buffer. The peptides were eluted using 2× 50 µL releasing buffer. After evaporation in the vacuum concentrator, the samples were redissolved in 23.5 µL 0.1% formic acid.

VII.2.4 Instrumentation and chromatographic conditions

The tryptic peptide mixtures were separated on an Ultimate 3000 Dual nano-HPLC system (Dionex), consisting of a DGP-3600MB dual ternary low-pressure proportioning micro pump, a WPS-3000TB nano pulled-loop thermostatted autosampler and a FLM-3100B Nano flow manager. The system is controlled via the Chromeleon software package. The solvents used are (A) 0.1% aqueous formic

acid and (B) 0.1% formic acid in an 80% aqueous solution of acetonitrile. A 20 μL sample was injected and concentrated in a PepMap C18 trap column (5 μm , 100 $\text{\AA}$, 300 μm I.D. $\times$ 5 mm from LC Packings) using a flow of solvent A. Separation of the peptides took place in a PepMap C18 capillary column (3 μm , 100 $\text{\AA}$, 75 μm I.D. $\times$ 150 mm from LC Packings) using a linear 45 min gradient at 300 nL min^{-1} in which the solvent B increased from 4% to 100%. The HPLC was connected to a Waters Micromass electrospray ionisation (ESI) quadrupole time-of-flight (QTOF) Ultima mass spectrometer, which was operated in data-directed acquisition mode, with an MS survey mass range from 500 Da to 1150 Da and 6 MS/MS acquisitions with mass range from 50 Da to 1998 Da. The precursor charge state selection for MS/MS analysis was set at $2^+ - 3^+$ with a minimum precursor total ion count of 50. Data acquisition was done using MassLynx.

VII.2.5 Data treatment

Mascot

For protein identification based on sequence library search, the software package Mascot Server (Matrix Science) was used, running on a local server. The measured data were subjected to Mascot MS/MS ions search against the SwissProt sequence database (UniProt Consortium, release 57.15). Carbamidomethylation of cysteine was selected as fixed modification. A first query was always run with the "error-tolerant" option to correct against other possible modifications (or degradations). Depending on the results second or more queries were eventually done with appropriate variable modifications (often, but not limited to *N*-terminal carbamylation, deamidated asparagine/glutamine and oxidation of proline and methionine). For a general search, the peptide mass tolerance was set at 0.3 Da, the fragment mass tolerance at 0.7 Da. The protein mass and taxonomy were unrestricted.

NIST

For spectral library search of the reference sample measurements (in Mascot's MGF file format) and creation/management of these libraries MS Search 2.0g (National Institute for Standards and Technology, NIST) was used. The parameters were adjusted to improve the search results on peptide MS/MS spectra: the type of

search was MS/MS identity, with the results limited to those matching the precursor ion mass. Tolerance levels were set at 0.3 Da for precursor ions and 0.7 Da for fragment ions. The precursor mass was ignored in MS/MS spectra and the optimal scoring setting for QTOF instruments was enabled.

Quick scanning of the measurements of the testing set and of the unknowns was done using NIST's batch search tool MSPepSearch 0.9 using the same parameters. All matches with a score exceeding 700 for the reference samples and 600 for unknowns were afterwards verified in MS Search.

VII.3 Results and discussion

VII.3.1 Building and test-running the spectral library

A series of animal glue reference samples originating from different species has been measured with HPLC-MS/MS: different samples of European rabbit (*Oryctolagus cuniculus*), bovine (*Bos primigenius taurus*), Siberian sturgeon (*Acipenser baerii*) glues, but also a limited number of animal glues based on pig (*Sus scrofa*), deer (Cervidae family), goat (*Capra aegagrus*), European brown hare (*Lepus europaeus*), carp (Cyprinidae family) and trout (Salmoninae subfamily of the family Salmonidae). To a lesser extent, other protein sources frequently encountered in arts, have been studied as well: bovine milk casein and chicken (*Gallus gallus*) egg white and yolk, even though their characterisation using Mascot is less challenging (88,87).

A single HPLC-MS/MS measurement often contains over 100 individual MS/MS spectra. Each of these spectra corresponds to an eluted component, although duplicate spectra (corresponding to the same component) often occur. Not all spectra, however, correspond to a peptide; these non-peptidic spectra should be removed. A script was written to semi-automate the selection of MS/MS spectra of a measurement: only those spectra were retained that have either a significant signal at 147.1 Da or 175.1 Da. These are markers for tryptic peptides containing respectively a lysine or arginine residue. Besides, for each spectrum significant annotation data (possible peptide sequence(s), post-translational modifications, protein and species) are extracted from the Mascot result report. Some of the

remaining non-peptidic spectra can be easily classified as noise (low signal-to-noise ratio) or originating from column bleed or polymeric materials (equal peak-to-peak mass increments) and can be manually removed from the dataset. Other MS/MS spectra, however, do have a peptidic nature, but are not characteristic for the measured sample, caused by trypsin autocleavage products, proteinaceous contaminations (mainly keratin) or accumulated peptides in the columns. These are more difficult to pinpoint and as a consequence, not removed from the dataset. In order to be able to discriminate characteristic from these non-characteristic spectra, whenever possible several sources of each of the animal glues were collected and in many cases the samples were analysed multiple times over a period of three years to take into account temporary interferences, such as variability in system sensitivity or instrument and column contaminations. All spectra, including those that were Mascot annotated as keratin peptides or trypsin autocleavage products, were appended to a spectral library. Thus, also undesirable side products can be correctly identified. A scheme summarising the complete data treatment procedure is shown in figure VII.2. At present the library contains over 5000 spectra, of which many are redundant.

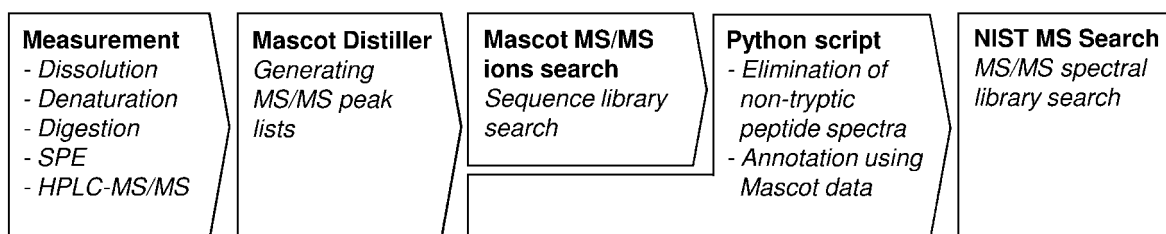


Figure VII.2. Scheme of the data treatment.

When a random measurement is analysed using this library, we expect that some of its tandem MS spectra have several equivalents in the library. The origin of these matched library entries supplies important information on the unknown spectrum and the sample as a whole. A well-considered interpretation of these results is of the utmost importance; different patterns can be distinguished:

- In case an unknown spectrum has several matching entries in the library, all originating from different animal glues of the same species, this spectrum might correspond to a unique peptide that only occurs in that species. The spectrum can thus be considered a species-specific marker. When at least

one of the matching entries is Mascot annotated, this information may be used to verify this hypothesis.

- In case the matching entries all are from animal glues, but originate from different species, this spectrum might correspond to a conserved peptide that is shared between these animals. This spectrum can be considered a species-unspecific marker for animal glue.
- In case the library matches are from totally different sources (e.g. animal glue and egg white), the spectrum is likely to correspond to a contaminant. The latter can have several sources and can be either a peptide (from keratin or a trypsin autocleavage product – this can often be verified if one of the matching entries is Mascot annotated) or non-peptidic.

Careful analysis of these different responses enables a correct determination of the unknown sample. To verify this, in what follows three measurements of animal glue reference samples (cattle, rabbit and sturgeon) were examined.

Cattle hide glue 3359

Because the complete bovine collagen sequences are available in the SwissProt library, many of the peptides in the measurement of Kremer's cattle hide glue 3359 could be annotated correctly in the Mascot results. Many of these peptides are also present in the corresponding proteins of other animals. Hence, Mascot features a "require bold red" function that filters out all proteins without exclusive peptide matches: if a peptide can only be attributed to a protein of one specific animal, this specific protein will be retained in the Mascot search results. Despite this, several collagen peptides from other species are still falsely identified: table VII.1 shows the number of identified peptides in Mascot using the SwissProt library, and the number of unique peptides, which are not present in any of the other proposed proteins. The bovine origin of this glue, however, is clear because of two reasons. Firstly, the higher number of collagen peptides of bovine origin, both unique and common, in comparison with those of the other species. Secondly, most of the unique peptides from other species carry unlikely post-translational modifications and are thus likely mismatched by Mascot.

Table VII.1. Summary of the Mascot/SwissProt search results with total (T) number and number of unique (U) identified peptides for each protein.

Protein chain	Species	T	U
Bovine hide glue 3359			
Collagen α -1(I) (COL1A1)	<i>Bos primigenius taurus</i> (cattle)	9	4
	<i>Mus musculus</i> (house mouse)	4	1
	<i>Gallus gallus</i> (chicken)	2	1
	<i>Mammut americanum</i> (American mastodon)	3	2
Collagen α -2(I) (COL1A2)	<i>Bos primigenius taurus</i> (cattle)	8	1
	<i>Homo sapiens</i> (human)	5	1
	<i>Gallus gallus</i> (chicken)	4	2
Collagen α -1(III) (COL3A1)	<i>Bos primigenius taurus</i> (cattle)	5	4
	<i>Rattus norvegicus</i> (brown rat)	2	1
Rabbit skin glue 581			
Collagen α -1(I) (COL1A1)	<i>Mus musculus</i> (house mouse)	9	1
	<i>Canis lupus familiaris</i> (dog)	8	1
	<i>Homo sapiens</i> (human)	8	1
Collagen α -2(I) (COL1A2)	<i>Bos primigenius taurus</i> (cattle)	3	1
	<i>Canis lupus familiaris</i> (dog)	3	1
	<i>Oryctolagus cuniculus</i> (European rabbit) - fragment	2	1
Collagen α -1(II) (COL2A1)	<i>Homo sapiens</i> (human)	2	0
	<i>Mus musculus</i> (house mouse)	2	0
	<i>Rattus norvegicus</i> (brown rat)	2	0
Collagen α -1(III) (COL3A1)	<i>Mus musculus</i> (house mouse)	2	2
Sturgeon bladder glue 3362			
Collagen α -1(I) (COL1A1)	<i>Rattus norvegicus</i> (brown rat)	4	1
	<i>Mus musculus</i> (house mouse)	4	1
	<i>Bos primigenius taurus</i> (cattle)	4	2
	<i>Homo sapiens</i> (human)	3	1
	<i>Canis lupus familiaris</i> (dog)	3	1
	<i>Xenopus laevis</i> (African clawed frog)	3	2
Collagen α -1(II) (COL2A1)	<i>Cynops pyrrhogaster</i> (Japanese fire belly newt)	5	5

The results of the dedicated spectral library search are summarised in table VII.2. Herein, only those spectra are retained that have at least three different matches (match factor higher than 700) and belong to at least two different glue samples (not necessarily of the same species) that were included in the spectral library. This way the chance of misinterpretation is diminished: spectra due to sample or column contamination, keratins or other accidental reasons are thus excluded. Despite the stringent filtering in the spectral library search, nine mass spectra from the cattle hide glue 3359 sample were matched with spectra from (only) animal glues. These nine spectra belong to seven different peptides, because spectra 17, 22 and 30 were found to be highly similar. For all spectra, matching library spectra from cattle glue were found, but with the exception of spectrum 54, matching spectra of glues from other species were also found. Most likely, these belong to conserved collagen peptides, while the peptide represented by spectrum 54 is exclusively found in bovine collagen.

Table VII.2. Spectral library search results of cattle hide glue 3359 with sequence annotations. (*) *P* indicates a hydroxylated proline (hydroxyproline).

Spectrum	Observed mass and charge	Spectral library query results	Mascot and BLAST results Annotation*	Relevant occurrence	Mass deviation
1	730.41, 2 ⁺	Cattle Deer Pig	GSAGPPGATGFPEGAAGRV COL1A1	Cattle	0.0794
14	634.39, 2 ⁺	Cattle Deer Rabbit	GIPGPVGAAGATGAR COL1A2	Cattle, Rabbit	0.0603
17, 22, 30	549.30, 2 ⁺	Cattle Pig	PGPVGAAGATGAR COL1A2	Cattle, Rabbit	0.0820
50	781.98, 2 ⁺	Cattle Deer	-	-	-
51	659.39, 3 ⁺	Cattle Deer Goat	SGDRGETGPAGPAGPIGPVGA R COL1A1	Cattle	0.1473
54	951.89, 3 ⁺	Cattle	GLTGPIGPPGPAGAPGDKGEA GPSGPAGPTGAR COL1A1	Cattle	0.1442
56	937.92, 3 ⁺	Cattle Deer	-	-	-

If available, the spectra in table VII.2 are annotated with the peptide sequence as proposed in the Mascot search results of both the cattle glue 3359 measurement and the measurements that contain the matching MS/MS spectra in the library. In order to find all proteins (and species) in which these identified sequences are present, they are online queried in the basic local alignment search tool (BLAST) of the National Center for Biotechnology Information (NCBI)(175). Five spectra could be annotated; all stem from collagen α -1(I) (COL1A1) and all are found in bovine, amongst others. The COL1A1 sequences for deer and sturgeon are unknown and less than 4% of the rabbit COL1A1 is sequenced; these species therefore cannot be determined by BLAST for this protein. If the sequence is conserved in evolutionarily closely related species that has its COL1A1 sequenced (e.g. rodents (rat, mouse, guinea pig) instead of rabbit), then BLAST will identify these. This is, however, not the case for the four annotated sequences. The sequence for spectrum 1 (figure VII.3 A) for example, GSAGPPGATGFPGAAGR is present in 12 known COL1A1 proteins from different animals, of which 10 mammals (cattle, dog, human, orang utan, rhesus macaque, gibbon, giant panda, horse, donkey and mastodon), the others being chicken and a lizard species. For readability, table VII.2 lists only those annotations that are relevant in the study of animal glues used in arts; in this case this is only cattle. Spectrum 51 (figure VII.3 C) is annotated as SGDRGETGPAGPAGPIGPVGAR, which is only shared between bovine and giant panda. These findings are in agreement with those obtained with the spectral library search: the sequences do not occur in rabbit (or better sequenced close relatives such as rat, mouse, guinea pig) or any of the fish that have sequenced COL1A1 in the NCBI library used for BLAST. Spectrum 14 (figure VII.3 B), on the other hand matches with library spectra of cattle, deer and rabbit. The corresponding sequence, collagen α -2(I) (COL1A2) GIPGPVGAAGATGAR, occurs in (amongst others) cattle and rabbit. This conserved peptide thus cannot be used to distinguish between these two species.

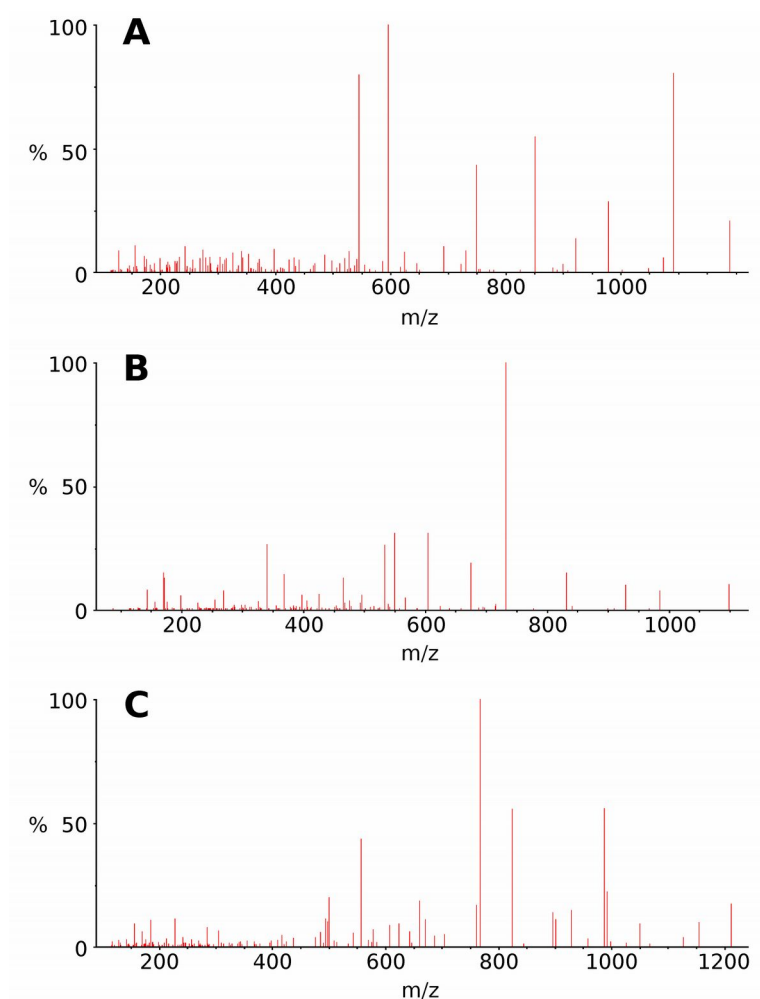


Figure VII.3. MS/MS spectra of cattle hide 3359: (A) spectrum 1, annotated COL1A1 GSAGPPGATGFP-GAAGRV, (B) spectrum 14, annotated COL1A2 GIPGPVGAAGATGAR, and (C) spectrum 51, annotated COL1A1 SGDRGETGPAGPAGPIGPVGAR.

Rabbit skin glue 581

Rabbit skin glue 581 is an old sample labelled *Colle Totin*, which is a former commercial name for rabbit skin glue that still often is encountered in French texts as a synonym for rabbit glue. Mascot results (summarised in table VII.1, page 144) only show one match with a rabbit collagen, namely COL1A2, of which only roughly 38% of the total sequence is present in SwissProt. Apart from the negligible fraction of COL1A1, no other rabbit collagen sequences are present in SwissProt, seriously hampering its determination. Mouse and rat collagen sequences are frequently identified, which may be interpreted as an indication for rabbit skin glue,

because of the relative close evolutionarily relation. This premise, however is scientifically not sturdy.

The spectral library search (table VII.3) is manifestly pointing out a rabbit origin of glue 581: all spectra matched with the library entries from rabbit skin glue. Only three of them, spectra 33, 42 and 46, matched with entries from bovine hide too. These spectra likely correspond with preserved peptides that occur in both species.

Table VII.3. Spectral library search results of rabbit skin glue 581 with sequence annotations. (*) P indicates a hydroxylated proline (hydroxyproline); on N a mass difference of -27 Da is observed.

Spectrum	Observed mass and charge	Spectral library query results	Mascot and BLAST results Annotation*	Relevant occurrence	Mass deviation
16	676.41, 3 ⁺	Rabbit	HGNRGE <u>P</u> GPAGSIGPVGAAGPR COL1A2	Rabbit	0.1872
21	825.02, 3 ⁺	Rabbit	-		
24	801.48, 3 ⁺	Rabbit	-		
33	751.42, 2 ⁺	Rabbit Cattle	GAQGPPGATGF <u>P</u> GAAGR COL2A1	Rabbit Cattle Sturgeon	0.1399
38	650.04, 3 ⁺	Rabbit	NGDRGETGPAGPAGPIGPAGAR COL1A1	Rat Mouse Guinea pig	0.1641
39	802.97, 2 ⁺	Rabbit	GPAGSPGTPGPQGIAGQR, COL1A1	Rat Mouse	0.1965
40	974.56, 2 ⁺	Rabbit	NGDRGETGPAGPAGPIGPAGAR COL1A1	Rat Mouse Guinea pig	0.1562
42	700.93, 3 ⁺	Rabbit, Cattle	-		
43	803.35, 1 ⁺	Rabbit	-		
44	581.87, 2 ⁺	Rabbit	GIPGPPIGP <u>P</u> GPGR COL2A1	Rabbit Cattle	0.1442
46	634.40, 2 ⁺	Rabbit Cattle Deer	GIPGPVGAAGATGAR COL1A2	Rabbit Cattle	0.1395
47	896.49, 3 ⁺	Rabbit	-		
51	670.80, 3 ⁺	Rabbit	-		

Many of these spectra could be annotated using the Mascot results. Spectrum 46, for instance, was identified as GIPGPVGAAGATGAR (figure VII.4 D), which occurs in the COL1A2 chains of cattle, human and dog. BLAST reveals that this peptide is shared between nine mammals. Though this peptide is situated in the missing region of SwissProt's rabbit COL1A2 and thus not found by Mascot, the complete, yet unreviewed, sequence is present in the NCBI library, which is here used by the BLAST tool. As indicated by both the Mascot and NIST approach, spectrum 46 does not allow discerning rabbit and bovine glue, but rules out a sturgeon origin. Spectrum 33 is also not a specific marker for rabbit: Mascot identifies this as GAQGPPGATGFPGAAGR (figure VII.4 B), which is present collagen α -1(II) (COL2A1) from human, rat or mouse. Using BLAST 24 species were revealed sharing this peptide: mammals, chicken, amphibians and fish. These include recent and unreviewed COL2A1 sequences from both rabbit and sturgeon. Contrary to this, spectrum 16, which matched rabbit skin glue samples solely, proved to be a specific marker for rabbit skin glue. Its peptide sequence was revealed by Mascot as rabbit COL1A2 HGNRGEPPAGSIGPVGAAGPR (figure VII.4 A) and using BLAST, no other animals were found sharing exactly the same peptide. Consequently, this spectrum is a good marker for rabbit glue. Spectra 38 (figure VII.4 C) and 40 both represent the same peptide; the first triple charged, the latter double charged. This peptide could be identified by Mascot by an error-tolerant search as NGDRGETGPAGPAGPIGPAGAR (mouse or rat COL1A1) with a -27 Da mass difference at the *N*-terminus. This mass difference suggests a substitution of asparagine with serine in rabbit COL1A1. This still makes it different from its bovine counterpart (SGDRGETGPAGPAGPIGPVVGAR), and can be used as a marker for rabbit glue.

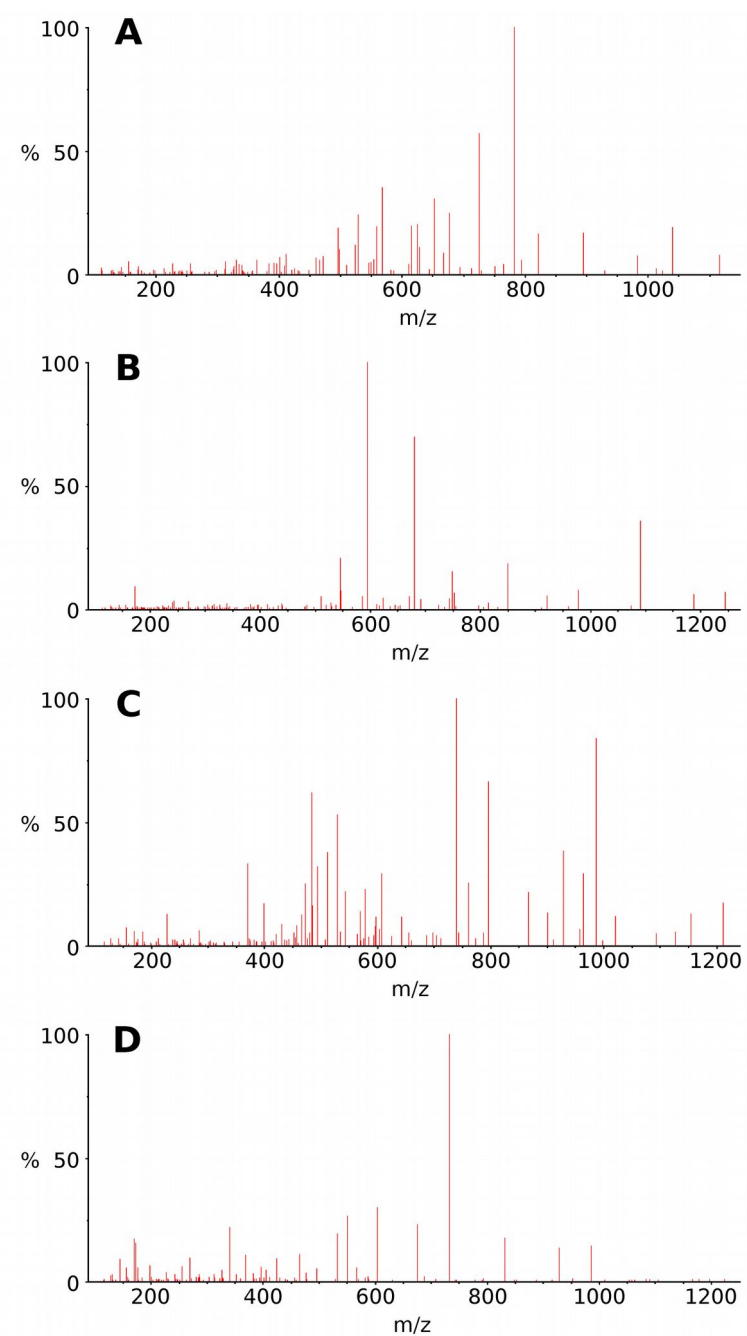


Figure VII.4. MS/MS spectra of rabbit skin glue 581: (A) spectrum 16, annotated COL1A2 HGNRGEPPGAGSIGPVGAAGPR, (B) spectrum 33, annotated COL2A1 GAQGPPGATGFPGAAGR, (C) spectrum 38, annotated COL1A1 NGDRGETGPA-GPAGPIGPAGAR, and (D) spectrum 46, annotated COL1A2 GIPGPVGAAGATGAR.

Sturgeon bladder glue 3362

Mascot search results tend to be particularly confusing when dealing with proteins that are not present in the sequence libraries. Sturgeon collagen, for example, is

missing in SwissProt. Even though some other fish species have their collagen sequenced and present in SwissProt, these are rarely proposed by Mascot in sturgeon glue measurements. Instead, particularly when using Mascot's option "error-tolerant search", a wide variability of proposed species can be observed within the same measurement (table VII.1, page 144), often also including amphibians such as the Japanese fire belly newt. This typical pattern can be considered an empirical, yet scientifically unsatisfactory indication for sturgeon glue.

This ambiguity is in sharp contrast with spectral library search, even though many library entries remain unannotated: table VII.4 shows the results of the spectral library search on sturgeon bladder glue 3362 (Kremer's Salianski isinglass). This is in agreement with previous findings (section V.3.2, page 107), where very low correlation was observed between sturgeon and mammalian based glues due to significant evolutionary change between the two. For eight spectra of glue 3362 matching spectra were found in the library, all of them from sturgeon bladder glue. Limited appeal can be made on the Mascot results for the annotation of the spectra, due to SwissProt lacking sturgeon collagen sequences. Moreover, the peptide sequences proposed when using "error-tolerant search" frequently exhibit unlikely to impossible post-translational modifications. Peptide annotations should therefore be taken with reserve; here, only annotations with hydroxylation of proline into hydroxyproline are taken into consideration. Spectrum 32 (figure VII.5 A) possibly corresponds to GPPGPSGPPGLGGPPGEP, found in COL1A1 of the Japanese fire belly newt. No other species are known to share this peptide. Considering the results of the spectral library search for this peptide, it is rather safe to assign this as a marker for sturgeon glue. The possible annotation of spectrum 35 (figure VII.5 B) on the other hand, COL1A1 GAAGPPGATGFPGAAGR, occurs not only in the Japanese fire belly newt and some fish, but also in certain rodents. As the sequence of rabbit COL1A1 is not present in any public sequence library, it is not possible to know if the same peptide is also present in rabbit, though the absence of rabbit glue entries in the spectral library query suggests that this is not the case.

Table VII.4. Spectral library search results of sturgeon bladder glue 3362 with sequence annotations. (*) *P* indicates a hydroxylated proline (hydroxyproline).

Spectrum	Observed mass and charge	Spectral library query results	Mascot and BLAST results Annotation*	Relevant occurrence	Mass deviation
18	953.50, 2 ⁺	Sturgeon	-		
22	636.03, 3 ⁺	Sturgeon	-		
23	775.10, 3 ⁺	Sturgeon	-		
32	546.01, 3 ⁺	Sturgeon	GPPGPSGPPGLGGPPGEP COL1A1	Fire belly newt	0.0290
35	772.38, 2 ⁺	Sturgeon	GAAGPPGATGFPGAAGR COL1A1	Rat, Mouse, Fire belly newt, Rainbow trout, Grass carp, Goldfish, Zebrafish, Atlantic salmon	0.0553
46	998.30, 5 ⁺	Sturgeon	-		
58	1026.16, 3 ⁺	Sturgeon	-		
60	927.35, 3 ⁺	Sturgeon	-		

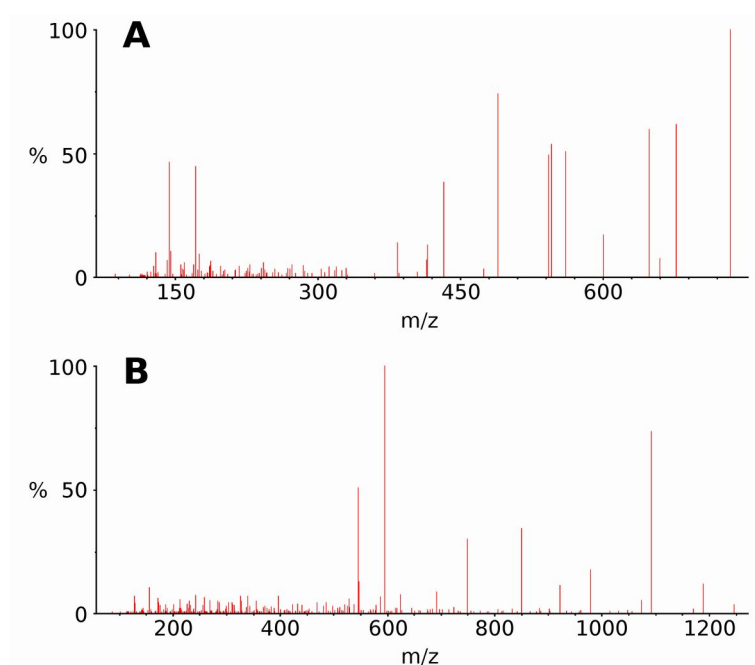


Figure VII.5. MS/MS spectra of sturgeon bladder glue 3362: (A) spectrum 32, tentatively annotated COL1A1 GPPGPSGPPGLGGPPGEP and (B) spectrum 35, tentatively annotated COL1A1 GAAGPPGATGFPGAAGR.

VII.3.2 Application of the libraries on historical samples

Papier-mâché in an anatomical model of an ear

The Mascot results on the two samples from the papier-mâché anatomical model of an ear were not revealing. Collagen peptides were abundantly found in both samples, but from several species: unique matches were found with collagen from human, mouse, rat, rainbow trout, African clawed frog, Japanese fire belly newt and the American mastodon. Most of these are very unlikely to be used for the preparation of animal glues. A large number of these proposed peptides also bear exotic and questionable post-translational modifications at one of the termini. This suggests that the correct sequence, resembling the best match sequence, belongs to a species/collagen that is missing in the library.

Analysis through spectral library search, on the other hand, demonstrates a sturgeon origin of the glue, or at least a species evolutionarily related to sturgeon, but additional sources cannot be entirely excluded in both samples. Ten spectra in sample E1 (table VII.5) have matching spectra with, and only with the Siberian sturgeon glue reference samples. A typical example is shown in figure VII.6 A: E1 spectrum 63 is shown head-to-tail with the best of the seven matching library spectra, all from Siberian sturgeon origin. One spectrum (spectrum 82, figure VII.6 B), on the other hand, has nine matching spectra from the cattle, deer, goat, pig and rabbit reference samples and was identified as GSAGPPGATGFPGAAGR.

In sample E2 (table VII.6) six spectra match exclusively with Siberian sturgeon reference samples, such as spectrum 49 (figure VII.7 A) while three others, such as spectrum 70 (figure VII.7 B), matched mammalian reference spectra.

Table VII.5. Spectral library search results on sample E1.

Spectrum	Spectral library match	Number of matches	Match factor	Mascot annotation
24	Sturgeon	1	809	
30	Sturgeon	1	638	
50	Sturgeon	2	603	
58	Sturgeon	6	637	
63	Sturgeon	7	685	SGDRGETGPAGPSGAPGPAGAR
66	Sturgeon	5	676	GPPGPSGPPGLGGPPGEP (similar to)
71	Sturgeon	4	674	
82	Cattle, Deer, Goat, Pig	9	637	GSAGPPGATGFPGAAGR
106	Sturgeon	1	657	
131	Sturgeon	2	623	
142	Sturgeon	3	636	GFPGTPGLPGVK, GFPGTPGLPGMK (similar to)

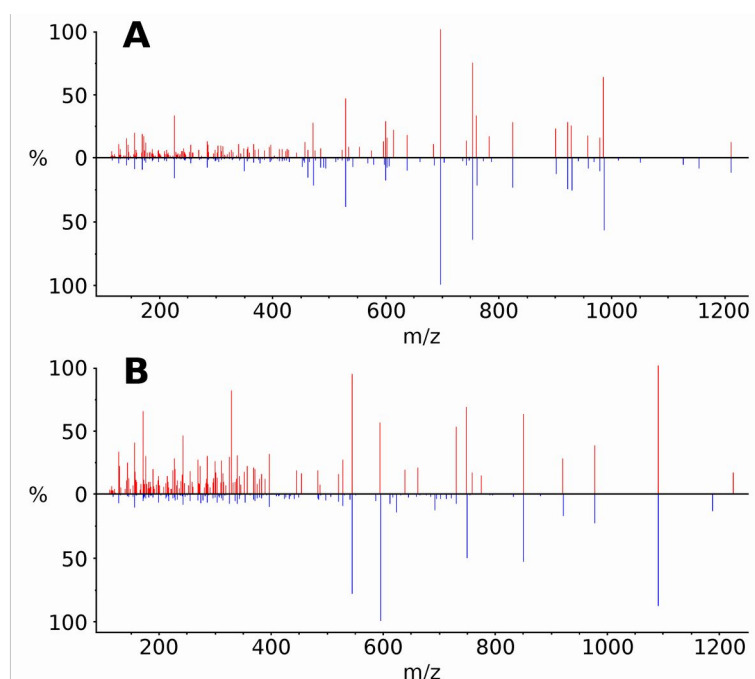


Figure VII.6. Head-to-tail spectra from the anatomical model of an ear sample E1: (A) spectrum 63 (top) compared with a library spectrum (bottom) from sturgeon, and (B) spectrum 82 (top) compared with a library spectrum (bottom) that occurs in cattle, deer, goat and pig.

Table VII.6. Spectral library search results on sample E2.

Spectrum	Spectral library match	Number of matches	Match factor	Mascot annotation
27	Sturgeon	1	678	
30	Rabbit, Deer	2	765	
36	Sturgeon	2	777	
42	Sturgeon	2	602	
49	Sturgeon	6	639	
51	Sturgeon	5	701	GPPGPSGPPGLGGPPGEP (similar to)
70	Cattle, Rabbit, Deer	19	783	GIPGPVGAAGATGAR
105	Sturgeon	4	600	

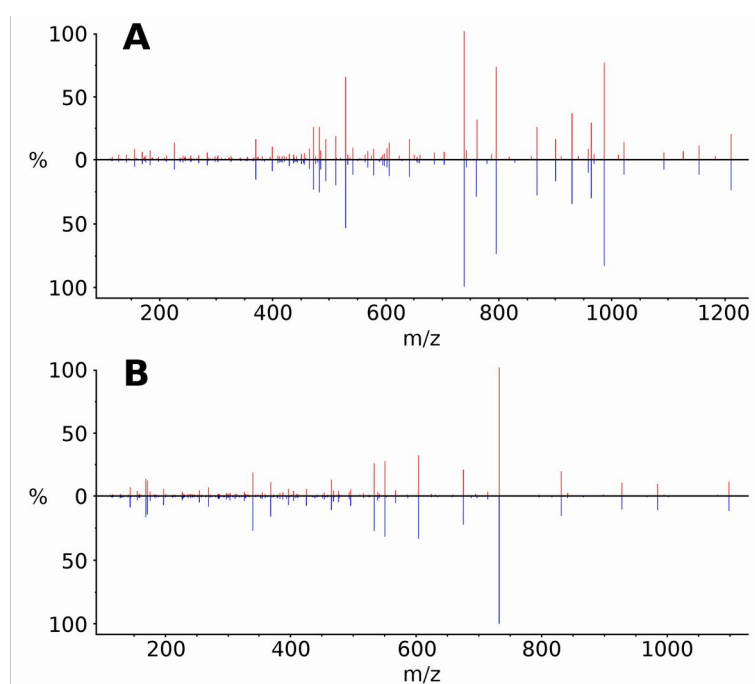


Figure VII.7. Head-to-tail spectra from the anatomical model of an ear sample E2: (A) spectrum 49 (top) compared with a library spectrum (bottom) from sturgeon, and (B) spectrum 70 (top) compared with a library spectrum (bottom) that occurs in cattle, deer and rabbit.

Ground layer in two paintings by Pieter Brueghel the Younger

In the two samples of the grounds of two paintings by Pieter Brueghel the Younger, the presence of animal glue was observed previously in section VI.3.4, page 128. Due to the extremely low sample quantity available for both samples B1 (*Sermon of St John the Baptist*) and B2 (*Crucifixion*), respectively only three and one collagen peptides could be identified; too few to be able to determine the species.

The application of the spectral library approach on sample B1 reveals that six spectra show a very high degree of agreement with spectra in the library (table VII.7). Four of these are exclusively found in rabbit glues. Spectrum 5, for example, has 13 matching rabbit glue spectra and is shown head-to-tail with a rabbit collagen library spectrum in figure VII.8 A. Two other B1 spectra (e.g. spectrum 6, figure VII.8 B) correspond with conserved peptides occurring in rabbit, cattle and deer. This indicates that the ground layer in Brueghel's *Sermon of St John the Baptist* is bound using animal glue based on rabbit collagen. A comparable result is observed in sample B2 (table VII.8) with three exclusive rabbit and two common peptides. Spectrum 1 (figure VII.9 A), for example, matches with a library spectrum from rabbit glue. However, spectrum 2 (figure VII.9 B) in this sample corresponds with reference spectra of sturgeon collagen. Thus, the ground layer in Brueghel's *Crucifixion* is based on rabbit glue, but an additional fraction of fish glue (sturgeon) cannot be excluded.

Table VII.7. Spectral library search results on sample B1.

Spectrum	Spectral library match	Number of matches	Match factor	Mascot annotation
3	Rabbit, deer, cattle	28	928	GSAGPPGATGFPGAAGR
4	Rabbit	16	912	GAQGPPGATGFPGAAGR
5	Rabbit	13	934	NGDRGETGPAGPAGPIGPAGAR (similar to)
6	Cattle, rabbit, deer	26	911	GIPGPVGAAGATGAR
11	Rabbit	3	771	
12	Rabbit	3	730	

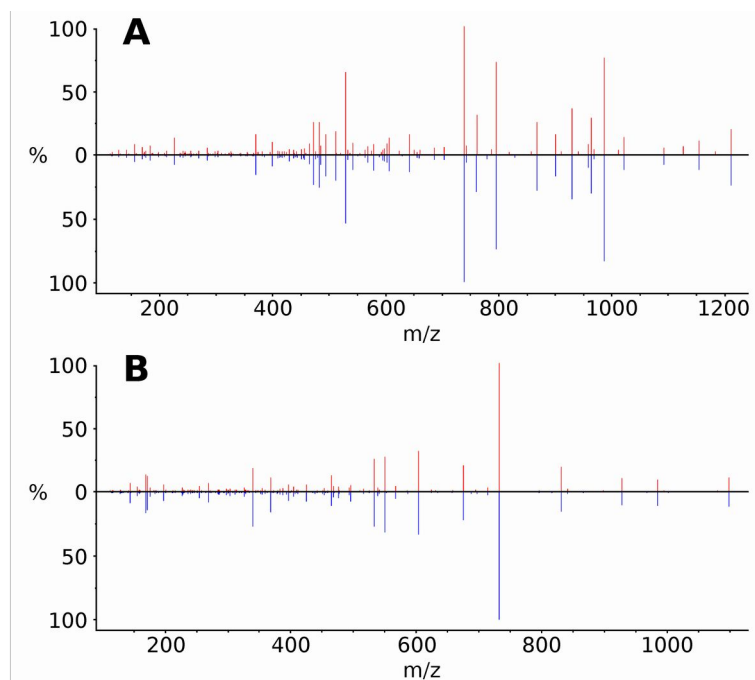


Figure VII.8. Head-to-tail spectra from the Brueghel ground sample B1: (A) spectrum 5 (top) compared with a library spectrum (bottom) from rabbit, and (B) spectrum 6 (top) compared with a library spectrum (bottom) that occurs in cattle, deer and rabbit.

Table VII.8. Spectral library search results on sample B2.

Spectrum	Spectral library match	Number of matches	Match factor	Mascot annotation
1	Rabbit	1	860	
2	Sturgeon	7	899	
3	Rabbit, deer, cattle	29	945	GSAGPPGATGFPGAAGR
4	Rabbit, deer, cattle	26	866	GIPGPVGAAGATGAR
6	Rabbit	3	793	
7	Rabbit	3	732	

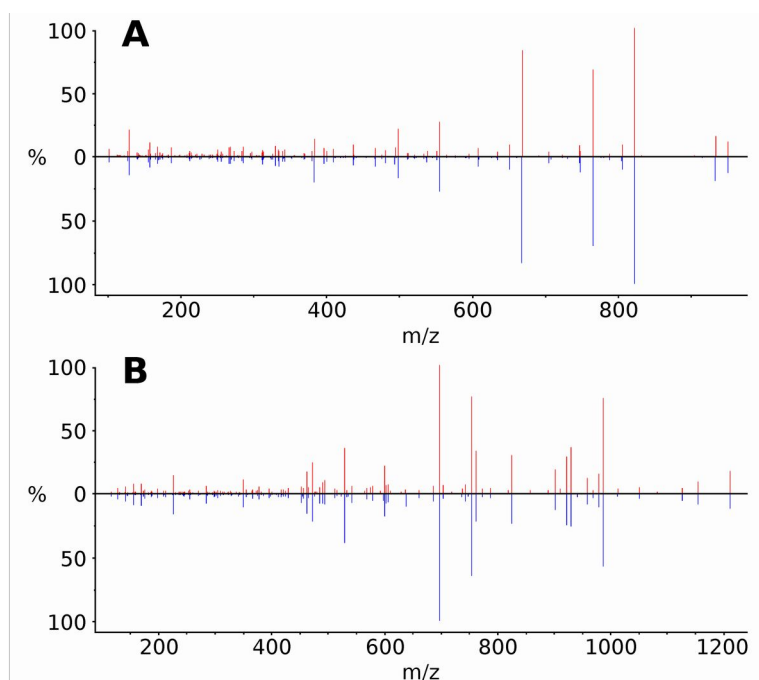


Figure VII.9. Head-to-tail spectra from the Brueghel ground sample B2: (A) spectrum 1 (top) compared with a library spectrum (bottom) from rabbit, and (B) spectrum 6 (top) compared with a library spectrum (bottom) from sturgeon.

VII.4 Conclusions

During the past few years, the use of proteomics techniques has proven to possess solid advantages for the identification of proteinaceous materials in works of art. The approach of sequence prediction, as used in the Mascot MS/MS ions search engine, however, falls short when trying to determine the species origin of animal glue due to incomplete coverage of collagen sequences for the possible sources and the intrinsic characteristics of the collagen sequences. An approach based on MS/MS spectral matching can overcome these problems. A dedicated spectral library, comprising MS/MS data of reference animal glues was built and applied on a test set of known samples from animal glues. The approach used here does not replace Mascot analysis, but combines spectral resemblance with Mascot annotated spectra.

Finally, a number of historical samples with unknown composition were analysed. Two samples from a 19th century anatomical model of an ear, made of papier-

mâché, show the presence of animal glue mainly based on fish glue, most likely Siberian sturgeon or a species that is evolutionarily close. The ground layers of two 17th century paintings from Pieter Brueghel the Younger on the other hand are bound using animal glue of (mainly) rabbit origin.

The spectral library is not finished; more spectra from new samples are continuously added: more of the same species, new species and other types of proteinaceous products that can be expected in art. The spectral library will be submitted to the MaSC mass spectral library in the near future.

The following paragraph is not included the published scientific paper:

Peptide analysis using tandem mass spectrometry has proven to be a unique tool to determine with extreme specificity the protein matter in works of art. Although the development of this and other proteomics-based methodologies continues, they are increasingly used to unravel practical issues in conservation science. In the next chapter proteomics-based techniques, amongst others, will be used to study the peculiar composition of the double ground layers in a painting made by Rembrandt.

VIII

CASE STUDY:

THE DOUBLE GROUND LAYER IN REMBRANDT'S *Portrait of Nicolaes van Bambeeck* ANALYSED BY PROTEOMICS TECHNIQUES

Complex questions regarding paint degradation, conservation treatments and the study of the painting techniques, rarely can be answered with a single analytical technique. Using a diversity of approaches and instruments, the jigsaw puzzle can be laid. The remarkable double ground layers in Rembrandt's *Portrait of Nicolaes van Bambeeck* were studied extensively in this manner. Each technique revealed unique information that is essential to understand the build-up of the ground layers. In this chapter, the application of proteomics-based techniques, developed in the previous chapters, on this grandiose masterpiece is highlighted.

VIII.1 Introduction

Rembrandt van Rijn (Leiden, 1606/07 – Amsterdam, 1669) is widely recognised as one of the greatest painters in European art history. Active in the 17th century, Rembrandt's painting style is considered as baroque, although his work is highly innovative and atypical in many ways.

The *Portrait of Nicolaes van Bambeeck* (Royal Museums of Fine Arts of Belgium, Brussels, figure VIII.1) is a part of a series of two paintings of the Amsterdam-based wool merchant Nicolaes van Bambeeck and his spouse Agatha Bas. She was a daughter of Dirck Bas, director of the Dutch East India Company and mayor of the city of Amsterdam. Both paintings were produced in 1641, when the couple lived in the same street as Rembrandt in Amsterdam. The creation of the two canvasses roughly coincide with that of Rembrandt's master piece, the *Night Watch* (Rijksmuseum, Amsterdam), painted between 1639 and 1642.



Figure VIII.1. Rembrandt's *Portrait of Nicolaes van Bambeeck* after conservation treatments. (source: KIK/IRPA)

The portraits got separated in history. Nowadays, the *Portrait of Agatha Bas* is part of the British Royal Collection in Buckingham Palace. The *Portrait of Nicolaes van Bambeeck*, on the other hand, is part of the collections of the Royal Museums of Fine Arts of Belgium. Irrespective of its aesthetic splendour, the painting is

extraordinary in another way: contrary to most other paintings by Rembrandt, this portrait has undergone almost no restoration during its lifespan. However, at the edges of the canvas, local small tears, loosening of the pictorial layers and extended craquelure urged a conservation treatment. (176)

Rembrandt's canvas paintings are quite variable in their ground structure. However, two basic types are usually encountered(177,178): the so-called double ground and the single quartz ground. In the double ground paintings, the first layer is red to orange, while the second one is light grey, grey or dull greyish yellow, preparing the surface colour and texture for the painting. Single-layer quartz grounds consist of one layer of coarsely sieved quartz (silica sand) tinted with a little brown ochre and contains a low proportion of lead white. The earliest painting with this kind of ground is the Night Watch(179).

VIII.2 Overview of the analytical study

The restoration campaign (2007-2009) at the Royal Institute for Cultural Heritage (KIK/IRPA) was an ideal opportunity to study the painting technique and the materials used by Rembrandt. Two cross sections prepared during a previous investigation of the Rembrandt painting in the KIK/IRPA (1962) were re-examined with current analytical techniques. These were complemented with a limited number of new microsamples, taken during and after the removal of the yellowed varnish layers(180). Optical microscopy (with white light and UV illumination), scanning electron microscopy equipped with an energy-dispersive X-ray detector (SEM-EDX), micro-Raman spectroscopy (MRS), synchrotron radiation Fourier-transform infrared spectroscopy (SR-FTIR) and cluster time-of-flight secondary ion mass spectrometry (TOF-SIMS) were applied on samples transformed into cross sections. Single-layer scrapings were analysed using transmission-FTIR with a diamond anvill cell, gas chromatography mass spectrometry (GC-MS) and high performance liquid chromatography with a diode array detector (HPLC-DAD). To identify eventual proteinaceous binders and their exact nature, two proteomics-based techniques were applied: matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) and HPLC coupled to a tandem mass spectrometer (HPLC-MS/MS). A general overview of the applied methodology and

results(180), detailed results of the TOF-SIMS measurements(181) and an experimental approach of the double ground build-up using reconstructions(178) are published elsewhere and will be summarised below.

Cross sections revealed the presence of two superimposed grey-on-red ground layers in the *Portrait of Nicolaes van Bambeeck*. Similar ground structures are observed in half of Rembrandt's paintings and is a common practice with his contemporaries(177–179). The build-up is as follows:

- The ground layers are applied on a canvas lined with a proteinaceous glue. Traditionally an animal glue is used. TOF-SIMS measurements (figure VIII.2) show the presence of (poly)saccharide and protein markers, indicating the cellulose of the canvas and the animal glue sizing.
- The first, red-orange ground layer is coloured using clay pigments (red earth, umber and/or sienna), mixed with quartz (SiO_2) fragments. No drying oils could be detected by GC-MS, SR-FTIR and TOF-SIMS (figure VIII.2) measurements. Although the latter technique only shows a weak signal for proteins, a (low quantity of) proteinaceous binder is expected.
- The second, light grey ground is coloured using mainly lead white ($2\text{PbCO}_3 \cdot \text{Pb}(\text{OH})_2$) with some charcoal (C), red earth and umber/sienna. Iodine tests revealed the presence of wheat starch as oval granules (figure VIII.3); this was later confirmed with MRS, FTIR and TOF-SIMS. Starch was not used as binder through gelatinisation, since that would imply the granules to have been swollen and burst. It is very likely added as an oil thickener and as an extender to lead white mainly for economical reasons(178). TOF-SIMS (figure VIII.2) furthermore revealed the presence of proteins in the granules, suggesting the use of flour instead of purified starch. FTIR and TOF-SIMS measurements indicate the use of an oil binder; this was ascertained by GC-MS as linseed oil (palmitate-stearate ratio 1.5).
- On top of the ground layers, multiple oil-based pictorial layers are present.

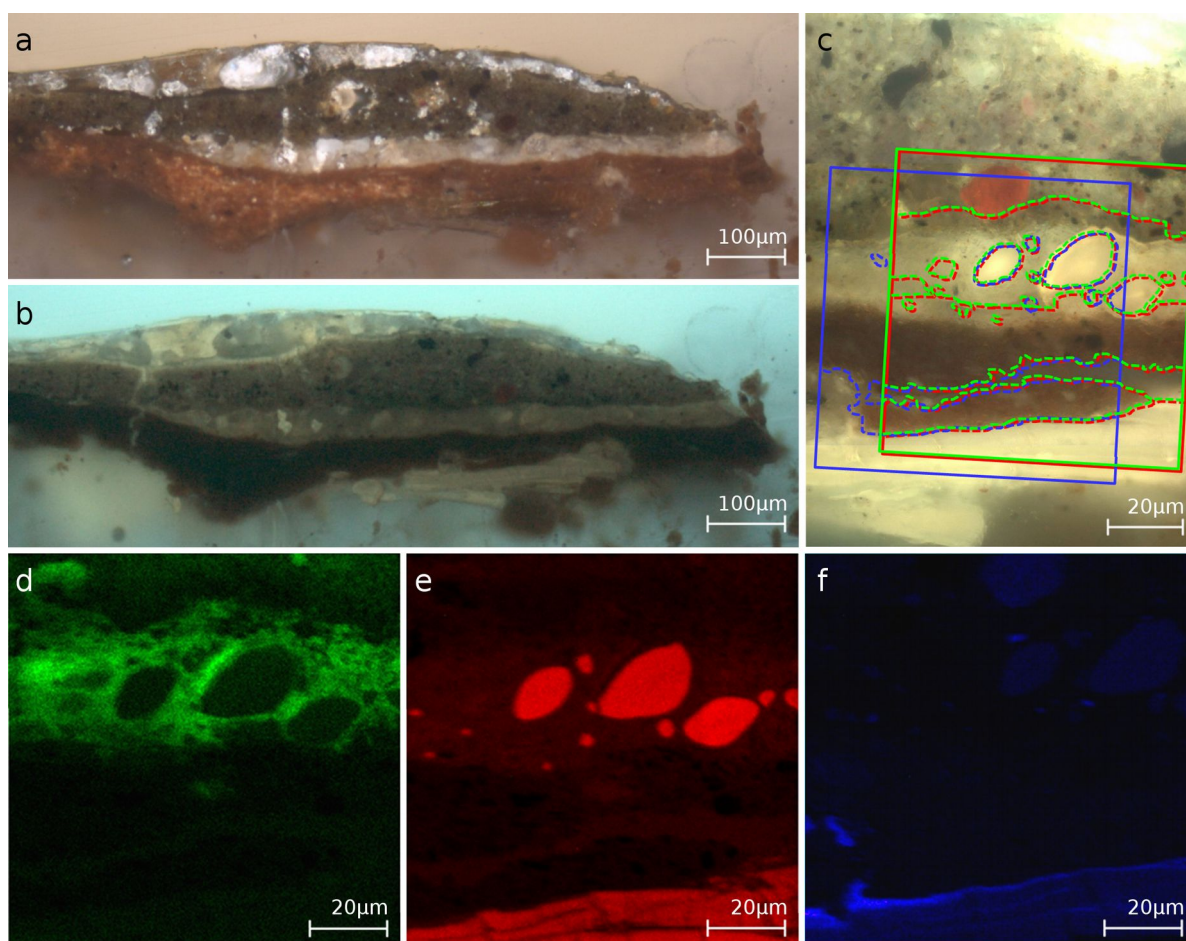


Figure VIII.2. Cross section of sample 2 (1962): (a) microscopy image acquired with 400x magnification under polarised white light and (b) under UV illumination. (c) Detail showing the areas studied by TOF-SIMS: (d) fatty acid markers (m/z 255.96, 284.06, 281.96, 253.92, 183.61, 169.57, 155.52), (e) saccharide markers (m/z -45.00, -59.01, -71.01, -99.01) and (f) protein markers (m/z 18.04, 30.03, 44.05, 70.06, 72.08, 84.08).

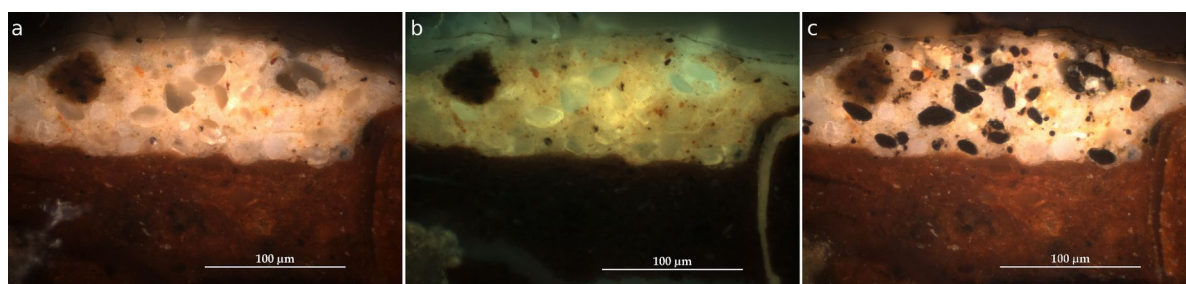


Figure VIII.3. Optical microscopy (500x magnification) on a cross section of sample 8 (2009): (a) under polarised white light and (b) UV illumination. (c) Under polarised white light after application of iodine in an aqueous potassium iodide solution (iodine test).

Most of these results are in agreement with numerous earlier studies made on other Rembrandt paintings, amongst others, in the framework of the Rembrandt Research Project(177,179,182,183). However, the presence of starch in the second

greyish layer of the double ground was unexpected and not reported before. It was probably added as an extender to the lead white and was bound with oil. Also, in the previous studies on Rembrandt paintings with the double ground, little information is given concerning the binding medium of the lower red/orange ground: some sources suggest an oil medium, while others avoid the issue (177,182,183). In the current research, a vague indication for a protein based medium arose.

In this work on Rembrandt's *Portrait of Nicolaes van Bambeeck* two of the proteomics-based techniques were applied to determine the presence and origin of the proteins in the two ground layers: matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS, see chapter V) and high-performance liquid chromatography coupled to tandem mass spectrometry (HPLC-MS/MS, see chapter VI). These techniques were applied in the frame of the present thesis and therefore will be discussed into detail in the following sections.

VIII.3 Experimental

Microsamples were taken at the edge of the painting under the frame by carefully scraping the ground layers one by one using an ophthalmic microscalpel. It is, however, impossible to attain perfect layer-pure samples this way; thus they might also contain traces of neighbouring layers and sizing. To minimise the risk of such contaminations, the samples were also inspected under a stereomicroscope and, if necessary, as much as possible cleared of these contaminants. The sample was divided into several parts for the different layers and techniques (table VIII.1).

Table VIII.1. Samples studied on Rembrandt's *Portrait of Nicolaes van Bambeeck*.

Sample	Description	Technique	A priori results (organic matter)
1	Lower red ground layer	HPLC-MS/MS	Unknown, possibly proteinaceous
2	Upper grey ground layer (analysis showed however a considerable contamination)	MALDI-TOF-MS	Linseed oil, grains of starch or flour
3	Upper grey ground layer	HPLC-MS/MS	

For MALDI-TOF-MS, the sample is enzymatically digested into peptides using trypsin, desalted using ZipTip solid phase extraction (SPE), mixed with 2,5-dihydroxybenzoic acid (DHB, used as a matrix) and measured with a Biflex IV MALDI-TOF-MS (Bruker Daltonics) as described in sections V.2.3 and V.2.4 (page 97). A variant of peptide mass fingerprinting (PMF) is used for the identification of the binders: the resulting mass spectra are evaluated for the presence of relevant marker peaks for reference protein samples.

For HPLC-MS/MS, the proteins in the samples are subjected to a multi-step protocol (sections VI.2.3 and VI.2.4, page 118), consecutively for dissolution (urea, thiourea and sodium dodecyl sulphate solution), denaturation (solutions of dithiothreitol and iodoacetamide) and tryptic digestion into peptides, which are afterwards purified using SPE. The solution containing the tryptic peptides is analysed with HPLC-MS/MS (Waters Q-TOF Premier). The results are treated with a local version of the Mascot search engine, which searches different protein sequence libraries (see chapter VI). For this project the high-quality SwissProt library (version 51.6) and the broader Trembl library (version 34.7) were used. For collagen-based samples, the data-set is queried with the dedicated collagen spectral MS library using the software tools MS Search 2.0g and MSPepSearch 0.9 (National Institute for Standards and Technology, NIST; chapter VII)

VIII.4 Results and discussion

VIII.4.1 Proteins in the lower red layer

TOF-SIMS, GC-MS and FTIR measurements could not unambiguously establish the binding medium used in the lower red ground layer (sample 1) due to its contamination by sizing and also likely because of its low quantity compared to the inorganic filler. Lipid markers were completely absent in all three methods. Protein markers were only observed in TOF-SIMS, albeit in extremely low quantities. To verify this ambiguous finding, HPLC-MS/MS was used for its capabilities to deal with mixtures and to determine the species in case of animal glues, while requiring very little amounts of analyte (10 µg of paint sample, see section VI.3.3, page 127).

Sample 1 was enzymatically digested with trypsin and measured using HPLC-MS/MS. The resulting data was queried with sequence libraries using Mascot MS/MS ions search, which revealed the presence of ten peptides, of which six could be attributed to keratin contaminants. The remaining four peptides (table VIII.2, figures VIII.4) can be attributed to collagen α -1(I) (COL1A1) and α -2(II) (COL1A2). To aid species-specific collagen determination, a spectral library search was done. Mascot potentially classifies spectra 3 (figure VIII.4 A) and 14 (figure VIII.4 C) as COL1A1 peptides from cattle (*Bos primigenius taurus*) with relatively high molecular weight search (MOWSE) scores. The former can alternatively also be attributed to the slightly different homologous peptide in brown rat (*Rattus norvegicus*) COL1A1. This is, however, less probable due to some missing key ions, which reflects in a lower MOWSE score of 54 instead of 66 for the cattle peptide. Also, hydroxyproline in the second place of the GPP-triad is highly unlikely (see section II.3.1, page 31). Mascot's annotation of spectrum 22 (figure VIII.4 D) as another brown rat COL1A1 also suffers from a poor ion coverage and an unlikely hydroxyproline in the GPA-triad, and therefore is most likely mismatched. A single COL1A2 peptide is identified in the Mascot results: spectrum 5 (figure VIII.4 B) corresponds to a peptide found in protein of cattle and, unlikely for the preparation of animal glues, dog (*Canis lupus familiaris*) and human (*Homo sapiens*) in the SwissProt library.

These findings are in agreement with the results of the spectral library search. Because of the relatively high signal-to-noise ratio, query results were enhanced by the use of the "reverse match", which ignores any peaks in the unknown (noise) that are not in the library spectrum. Figure VIII.4 A-C shows the head-to-tail spectra of the unknown spectra 3, 5 and 14 (up) and the best matches in the library (down). Spectra 3, 5 and 14 all correspond to spectra obtained on reference cattle glue samples, although with the exception of spectrum 5 not exclusively (table VIII.2). More importantly, the absence of Siberian sturgeon (and other fish) and rabbit is significant, besides cattle the three most documented sources for animal glues. No matching library entries were found for spectrum 22 (figure VIII.4 D), confirming the hypothesis of a mismatch by Mascot. Working near detection limits, the spectral library approach could not identify other spectra that were not already identified by Mascot. Only spectrum 16 was found to be a double of spectrum 14.

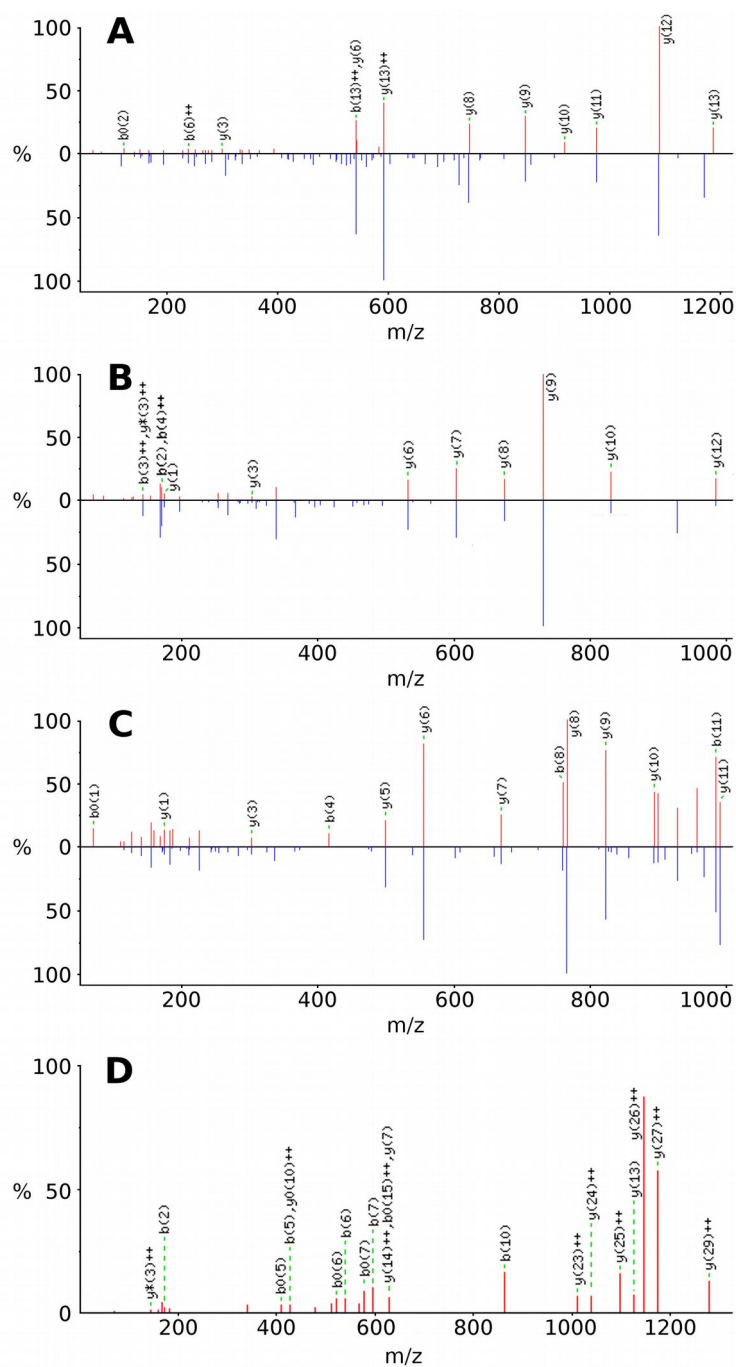


Figure VIII.4. Collagen MS/MS spectra on sample 1 (lower red ground) with Mascot's ion annotations in upright position. When matched, the library spectra is shown below (head-to-tail).
 (A) spectrum 3 (COL1A1 GSAGPPGATGFPGAAGR),
 (B) spectrum 5 (COL1A2 GIPGPVGAAGATGAR),
 (C) spectrum 14 (COL1A1 SGDRGETGPAGPAGPIGPVGAR) and
 (D) spectrum 22 (incorrectly identified by Mascot as COL1A1 GLTGPGPPGPAGAPGDKGETGPSGPAGPTGAR).

Table VIII.2. Summary of the Mascot search results using SwissProt on the lower red ground layer. (*) *P* indicates a hydroxylated proline (hydroxyproline).

Spectrum	Observed mass and charge	Annotation Peptide (*)	Protein	Mass deviation
3	730.44, 2 ⁺	R.GSAGPPGATGFPGAAGR.V R.GAAGPPGATGFPGAAGR.V	COL1A1 COL1A1	0.1800 0.1800
5	634.42, 2 ⁺	R.GIPGPVGAAGATGAR.G	COL1A2	0.1499
14 (16)	659.41, 3 ⁺	K.SGDRGETGPAGPAGPIGPVGAR.G	COL1A1	0.2328
22	961.93, 3 ⁺	R.GLTGPIGPPGPAGAPGDKGETGPSGPA GPTGAR.G	COL1A1	0.3376

The sequence libraries are rapid-growing and form an interesting resource to find, once identified with Mascot or through spectral libraries, in which species an exact (or distinctive) peptide sequence occurs. The three identified peptides were searched using basic local alignment search tool (BLAST) against the NCBI library(175). The results are summarised in table VIII.3. In the three cases, a cattle-based glue seems possible, while only SGDRGETGPAGPAGPIGPVGAR also occurs in rabbit COL1A2. However, apart of a 53 amino acids long fragment, rabbit COL1A1 remains still largely unsequenced. Therefore, no conclusions can be drawn on the other two peptides with regard to a rabbit origin.

Although we generally consider a protein identified only if at least three of its peptides are significantly detected, there are strong indications for the presence of collagen type I, and thus animal glue, in the sample of the lower red ground. Based on as few as three peptides, species-specific determination of the glue is risky. Analysis results favour a cattle-based glue. Mascot and BLAST, however, due to the incomplete rabbit collagen sequences, cannot exclude a rabbit-based glue. Also, the absence of rabbit glue library spectra cannot be interpreted as a hard proof that a cattle-based glue was used. The low number of peptides measured, or the extremely low amount of glue in this sample adds to the question if it is actually present in the lower red ground layer or if small fragments of the glue sizing of the canvas are accidentally sampled with the red ground.

Mascof MS/MS Ions search		Spectral library search	
Occurrence	MOWSE score	Reference sample matches	Match factor
cattle	66	goat hide, deer hoofs, cattle hide, pig skin	610
rat	54		
cattle, dog, human	59	cattle bone, cattle hide	728
cattle	75	cattle bone, cattle hide, goat hide	672
rat	46	-	

Table VIII.3. Species having the exact sequence in their collagen determined with NCBI BLAST.

Sequence	Species
GSAGPPGATGFPGAAGR	African bush elephant, black-capped squirrel monkey, black flying fox, bonobo, bottlenose dolphin, Carolina anole, cattle , common chimpanzee, common marmoset, crab-eating macaque, dog, domestic cat, donkey, giant panda, greater galago, green sea turtle, horse, human, killer whale, naked mole rat, nine-banded armadillo, northern white-cheeked gibbon, olive baboon, rhesus macaque, sheep, Sumatran orangutan, Tasmanian devil, walrus, western lowland gorilla, West-Indian manatee, white rhinoceros, yak
GIPGPVGAAGATGAR	Cattle , domestic cat, giant panda, sheep, yak
SGDRGETGPAGPAGPIGPVGAR	African bush elephant, black-capped squirrel monkey, bottlenose dolphin, cattle , Chinese hamster, Chinese tree shrew, common chimpanzee, common marmoset, crab-eating macaque, dog, domestic cat, European rabbit , giant panda, greater galago, guinea pig, human, killer whale, northern white-cheeked gibbon, olive baboon, rhesus macaque, sheep, Sumatran orangutan, walrus, western lowland gorilla

VIII.4.2 Proteins in the upper grey layer

Based on FTIR, GC-MS and TOF-SIMS measurements, the upper grey ground was found to be bound with linseed oil. Using MRS, staining techniques and SIMS, starch particles were observed. Therefore, no proteins were expected. Two microsamples were nevertheless prepared for proteomics-based protein analysis. Sample 2 was prepared for MALDI-TOF-MS analysis of its tryptic digest. With no protein matter

expected, the mass spectrum (figure VIII.5) was remarkable, exhibiting a high correlation with the animal glue references.

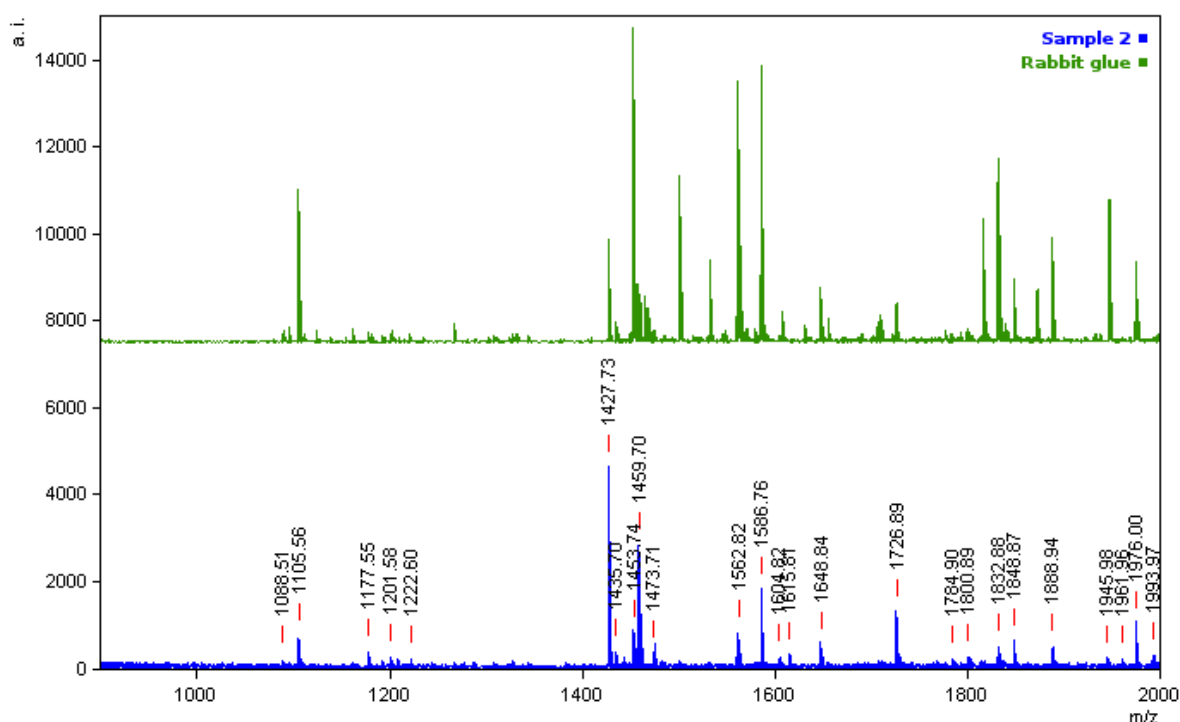


Figure VIII.5. MALDI-TOF-MS spectra of the grey ground layer (sample 2) and a rabbit glue reference.

The data were curated and uniformised, as was done with the reference data (section V.2.3, page 97). Firstly, the mass spectrum is limited to the most revealing range for protein identification (900 - 2000 Da). Secondly, to correct for different numbers of datapoints in different measurements and small intermeasurement mass shifts, the number of datapoints is reduced to one datapoint per Dalton. And finally, the data are normalised on a scale from 0 to 100.

Table VIII.4 lists the datapoints corresponding to the monoisotopic peaks with a peak height over 10% of the highest peak. Twelve out of 14 datapoints can be attributed to characteristic collagen markers (table V.1, page 101). This is confirmed by both principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA). Although the MALDI spectrum of sample 2, projected in the PCA score plot of a training set of the main protein binders (figure VIII.6 A), does not entirely fit within the group of animal glue spectra, it can

be considered closely related. The same phenomenon was observed with the spectra of the St Margaret of Antioch samples (section V.3.3, page 108) and is most likely caused by comparison of age-old samples with a set of unaged reference samples. Degradation phenomena, causing alterations to or breakage of the protein chain, can explain the missing collagen markers and, by consequence, the weak PCA classification. The SIMCA model classified sample 2 as animal glue.

Table VIII.4. The curated datapoints with the highest MALDI-TOF-MS response on sample 2, with possible annotations of the corresponding collagen markers.

Mass (Da)	Height (%)	Collagen marker	Possible peptide sequence annotations
900	11.6	•	GPSGDRGPR, bovine COL1A2 GWGLPGQR, bovine COL1A1
1106	14.5	•	GFPGADGVAGPK, bovine COL1A1 GVQGPPGPAGPR, bovine COL1A1
1428	100	•	GSAGPPGATGFPGAAGR, bovine COL1A1 ALLIQGSNDVEIR, bovine COL2A1 GIPGEFGLPGPAGAR, bovine COL1A2
1454	18.9	•	SAGVSVPGPMGPSGPR, bovine COL1A1
1460	60.8	-	
1476	12.3	•	?
1563	17.3	•?	DGRSGHPGTVGPAGLR, rabbit COL1A2
1587	39.3	•	GNSGEPGAPGSKGDTGAK, bovine COL1A1
1649	12.9	•	AGEDGHPGKPRPGER, bovine COL1A2 GSTGEIGPAGPPGPPGLR, bovine COL1A2 GRPGLPGAAGARGNDGAR, bovine COL3A1 GFSGLDGAKGDAGPAGPK, bovine COL1A1
1727	28.3	•	DGLNGLPGPIGPPGPRGR, bovine COL1A1 GENGVPGEDGAPGPMGPR, bovine COL3A1
1834	10.6	•	GEPGPTGIQGPPGPAGEEGK, bovine COL1A1
1849	14.4	•	GEPGPTGIQGPPGPAGEEGK, bovine COL1A1
1889	10.1	•	DGNPGNDGPPGRDGQPGHK, bovine COL1A2
1977	23.5	-	

Some of the generally more prominent collagen markers (table V.1, page 101) were not present in the measurement of Rembrandt's upper grey ground, such as 1502 Da, 1534 Da and 1817 Da. Interestingly, the former two were tentatively annotated as unique for rabbit collagens, suggesting Rembrandt possibly used

another source for preparing the animal glue. In the PCA score plot of a training set of different animal glues (figure VIII.6 B), the projection of the MALDI spectrum of sample 2 lays within the rabbit, cattle and deer glue data. Contrary to sturgeon glue, these cannot not be distinguished from each other by means of PCA (section V.3.2, page 107). Also, SIMCA classified sample 2 as mammalian glue.

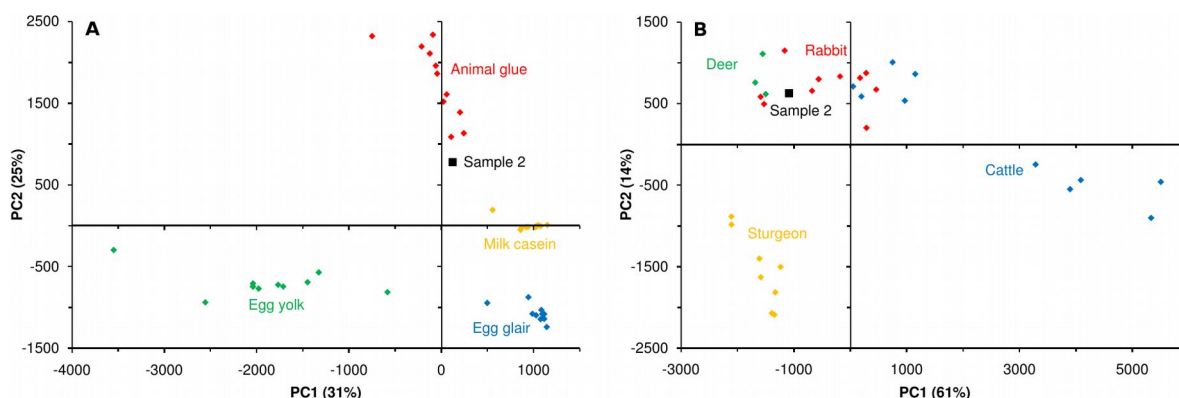


Figure VIII.6. The MALDI spectrum of the grey ground layer (sample 2) projected in the PCA score plots of (A) the main binder classes and (B) different animal glues.

With relative heights of respectively 60.8% and 23.5%, the two remaining datapoints at 1460 Da and 1977 Da without corresponding collagen markers cannot be neglected. These, however, do not correspond to any of the four major protein binders (animal glue, egg white, egg yolk and milk casein) studied in chapter V. Other protein sources, either added to the paint formulation in some form by Rembrandt or through contamination during sample preparation (e.g. keratins), cannot be excluded.

The blatant presence of animal glue in this sample strongly contradicts with the results obtained with FTIR, SIMS and GC-MS. This can be explained by the possible presence of particles of the neighbouring lower red protein-based ground layer (however low on proteins), the canvas sizing or a facing layer in the analysed microsample. MALDI-TOF-MS on the upper grey ground (sample 2) thus did not yield fully satisfying results, partly due to the limitations of the technique to determine the animal source of glues and partly due to the uncertainty of the layer purity of the microsample.

To overcome these issues, a sample was prepared for HPLC-MS/MS measurement (sample 3), taking even more attention to carefully select particles under a microscope from the grey layer exclusively. Mascot MS/MS ions search analysis on this measurement did not show any collagen, confirming that sample 2 used for MALDI-TOF-MS analysis must have been contaminated with collagen and that sample 3 is pure. Other than various unavoidable keratin contaminants, it did however reveal six peptides that could be attributed to different wheat (*Triticum aestivum*) proteins (table VIII.5). This is a surprising finding because of the low quantities of proteins: wheat grains mainly constitute of starch (68 - 72%) and proteins (10 - 12%)(180). About 80% of the protein content is gluten, a protein composite of many multimeric glutenin proteins and monomeric gliadin proteins. The remaining fraction are globulins (e.g. α -amylase inhibitor) and albumins.

Table VIII.5. Summary of the Mascot search results using Trembl on the upper grey ground layer (sample 3).

Spectrum	Observed mass and charge	Annotation	MOWSE score	Mass deviation
29	703.46, 2 ⁺	R.TTTRVPFGVGTGVGG.- LMW glutenin subunit group 3 type II, wheat	52	0.1790
41	784.99, 2 ⁺	R.TTTRVPFGVGTGVGGY.- LMW glutenin subunit group 3 type II, wheat	78	0.1619
42	864.53, 2 ⁺	R.SGNVGESGLIDPGCPR.E α -amylase inhibitor subunit CM3, wheat	105	0.2113
46	555.83, 2 ⁺	R.EMQWDFVR.L α -amylase inhibitor subunit CM3, wheat	49	0.1497
53	839.54, 2 ⁺	R.TLPTMCNVNVPLYR.T LMW glutenin subunit group 3 type II, wheat	72	0.2356
74	850.06, 2 ⁺	R.YFIALPVPSQPVDPR.S α -amylase inhibitor subunit CM3, wheat	43	0.1996

Three MS/MS spectra were identified as wheat glutenin peptides (spectra 29, 41 and 53). Glutenins form complex protein aggregates within the wheat flour granules consisting of a network of high- and low-molecular-weight (HMW and LMW) glutenin subunits. The identified peptides belong to LMW glutenin subunit group 3 type II. The vast sequence libraries (Trembl and NCBIhr) include a large number of these subunit sequences, deviating in only few amino acids. Spectra 29 and 41 (figure VIII.7A) are, for example, variants of the same sequence. Together with spectrum 53 (figure VIII.7B) they only cover 7% of the total sequence of LMW

glutenin subunit group 3 type II at the carboxyl terminus. The rest of the sequence, as do most other glutenin chains, lack suitable lysine and arginine residues for trypsin cleavage into measurable peptides.

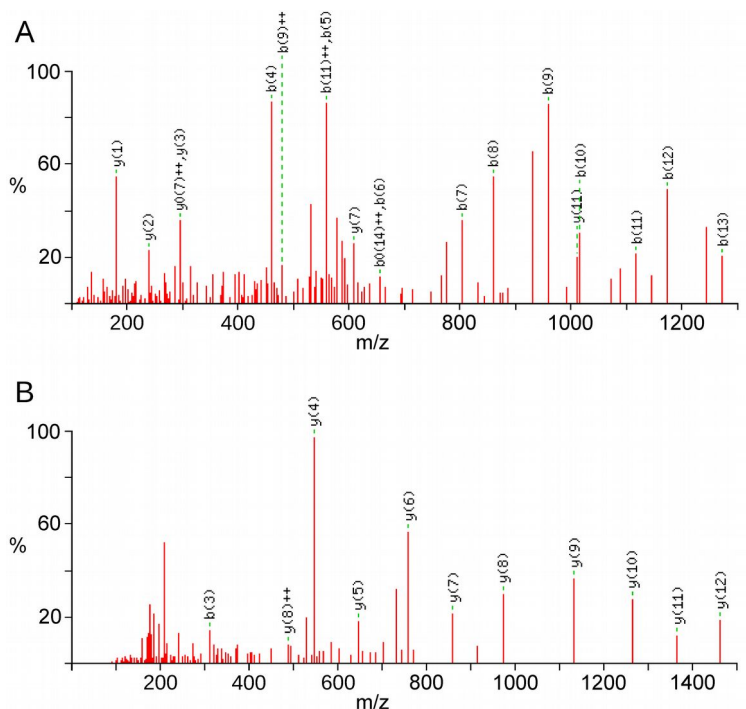


Figure VIII.7. Wheat glutenin MS/MS spectra from sample 3 (upper grey ground), identifying (A) TTRVPFGVGTGVGGY and (B) TLPTMCNVNPLYR.

Spectra 42, 46 and 74 (figure VIII.8) were identified as peptides occurring in the α -amylase inhibitor CM3 chain, one of the subunits of the tetrameric α -amylase inhibitor(184) occurring in wheat grains. This protein inhibits mammalian α -amylase, which plays a major role in the hydrolysis of polysaccharides in the body. The three sequences are identical for common wheat (*Triticum aestivum*) and durum wheat (*Triticum durum*), that both have their CM3 chain sequenced in Trembl. No distinction can be made.

While TOF-SIMS indicated the presence of protein markers in the starch granules, HPLC-MS/MS unambiguously established the presence of glutenin and α -amylase inhibitor. This finding answers the question whether Rembrandt, or more likely his supplier of prepared canvasses(178), indeed used wheat flour instead of starch. The preparation of starch is based on repetitive washing of wheat flour in cold

water, in which the globular proteins dissolve, the gluten disperses and the starch precipitates. The presence of wheat proteins thus means that either wheat flour has been used, or incompletely refined wheat starch. The refinement of starch, however, is a relatively simple and efficient procedure which with present-day methods only leaves 0.3% of proteins in starch(180); there is little reason to believe this value was significantly higher in Rembrandt's epoch.

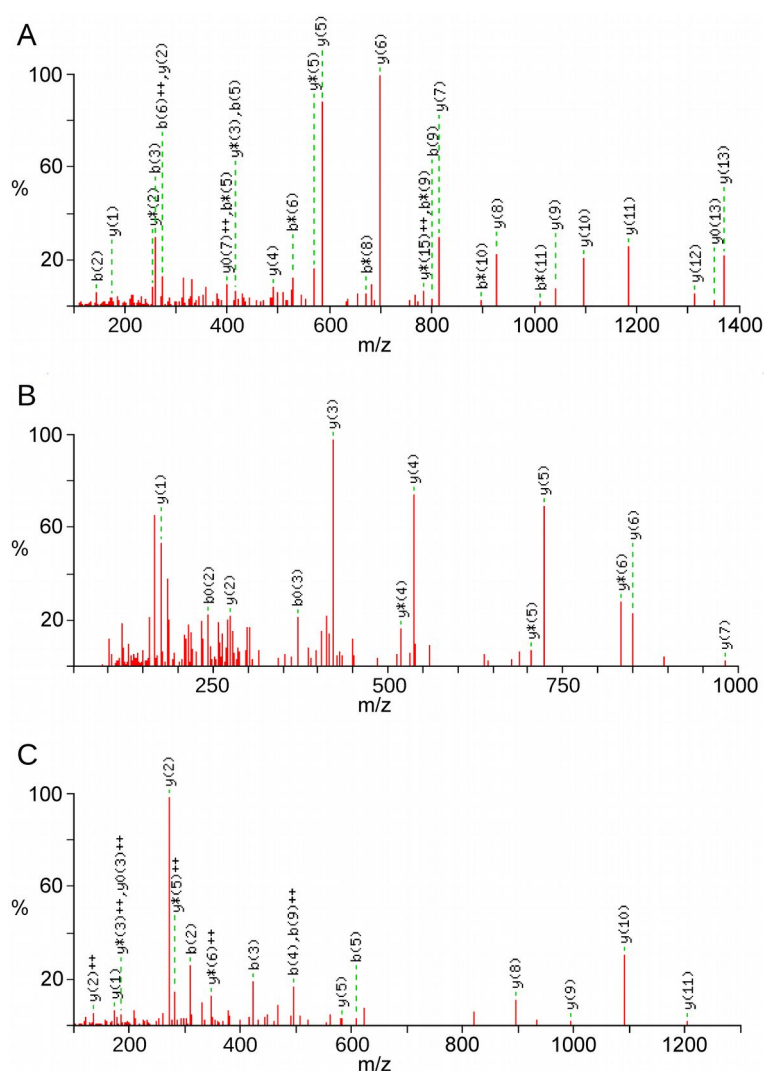


Figure VIII.8. Wheat α -amylase inhibitor MS/MS spectra from sample 3 (upper grey ground), identifying (A) SGNVGESGLIDLPGCPR, (B) EMQWDFVR, (C) YFIALPVPSQPVDPR.

VIII.5 Conclusion

Proteomics-based analytical techniques significantly added to the broad analytical study of the painting technique in Rembrandt's *Portrait of Nicolaes van Bambeeck*. Tryptic digests of the tiny microsamples of the grey-on-red ground layers were analysed with MALDI-TOF-MS and HPLC-MS/MS.

The binding medium of the lower red ground layer was undetermined and poorly documented in similar paintings. Despite the low amounts of peptides that could be identified, HPLC-MS/MS was able to identify animal glue. Sequence library (Mascot, BLAST) and spectral library search approaches suggest a bovine origin, although the frequently used rabbit skin cannot be entirely excluded. Also, it is unclear whether the animal glue is present in the red ground layer or is caused by fragments of the sizing of the canvas in the sample. The faint signal for protein markers observed in the images of TOF-SIMS analysis suggests the former.

Although animal glue was also detected by MALDI-TOF-MS in the sample of the oil-bound upper grey ground layer, HPLC-MS/MS did not. Fragments of the canvas sizing have compromised the MALDI-TOF-MS measurement. Surprisingly, peptides from the wheat grain proteins α -amylase inhibitor and glutenin were observed in the HPLC-MS/MS measurement. This not only confirms the presence of starch that was measured with various other techniques, it also affirms the use of wheat flour. The flour was used as an oil thickener and perhaps as a pigment extender.

IX

CONCLUSIONS AND FUTURE PERSPECTIVES

During the past decade the analysis of protein-bound paints and other proteinaceous materials in arts has been a hot research topic in conservation science. This is because traditional amino acid or pyrolysis based methods often fall short due to the unspecific nature of individual amino acids or their (even smaller) pyrolysis products. This leads to ambiguous identification in case of protein mixtures, contaminated samples or unexpected protein content. Method development seems to focus on two distinct proteomics-based analytical methodologies that are being applied on and adapted to proteinaceous materials in art: immunostaining and peptide analysis following enzymatic cleavage.

In the former, the antibody-antigen reaction allows for very specific determination of proteins up to species level. An attractive feature of some of these techniques is the possibility to use it on paint samples embedded as cross sections. This way, the presence and distribution of a specific protein (binder) along the different layers can be observed.

Peptide analysis, on the other hand, is based on the detection and identification of peptides, created by enzymatic cleavage of the proteins. Trypsin is incontestably the most frequently used enzyme, and up to present the only one known to be used on art materials. For the common proteins expected in paint binders, tryptic

cleavage creates peptides that often range between 5 and 15 amino acids. This is an ideal length: sufficiently long to be unique and thus characteristic for its parent protein, yet sufficiently light to enable high precision (resolution) detection in mass spectrometers. Dedicated single-layer samples, albeit challenging to obtain, are required and it is not possible to visualise the distribution of different types of binder in a sample (as is the case in certain immunostaining techniques). However, peptide analysis offers the lowest detection limits and the highest specificity, while enabling all proteins present in the sample to be identified in a single measurement.

Peptide analysis can be coupled to a range of detection and interpretation methods. In its most simple layout, the peptides can be separated with chromatographic techniques and subsequently measured with a suitable detector (chapter IV). High-performance liquid chromatography with a diode array detector (HPLC-DAD) requires a relatively cheap instrument, already available in many conservation labs. The interpretation of the results is straightforward: ideally each peak in the resulting chromatogram represents a single tryptic peptide, characterised by its retention factor; identification of the protein source is possible through comparison with chromatograms of reference samples. However, the smallest change in experimental conditions may shift retention times, hence complicating the interpretation. Sequence identification is not possible using the UV spectra, nor can this technique be used to study protein degradation processes in ageing works of art. This technique is nevertheless very suitable to complement other analytical techniques such as pyrolysis gas chromatography (py-GC-MS) or amino acid analysis (AAA), while sample consumption (about 100 to 200 µg) is on an equal level.

A more solid interpretation can be achieved using peptide masses instead of retention factors. Using a soft ionisation technique coupled to mass spectrometry, such as matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS), the masses of the (single-charged) individual tryptic peptides are determined. In proteomics, it is common practice to compare the MALDI spectrum with those of reference proteins or with a list of peptides calculated from known proteins. This technique, peptide mass fingerprinting (PMF), generally works well for single-protein samples, but is less suited for multi-protein samples, as is the case for

paint binders. Through the use of principal component analysis (PCA) and soft independent modelling of class analogies (SIMCA) we were able to select the most characteristic peaks for the four main groups of protein paint binders (animal glue, egg white, egg yolk and milk casein), while neglecting peaks originating from pigments, pigment-binder interaction or common contaminants (chapter V). Most of these characteristic peaks were attributed with tryptic peptide sequences, calculated from known protein sequences. Animal glues from fish and mammalian sources could also be distinguished, but no distinction could be made within the mammalian glues: the evolutionarily differences in mammalian collagens were often found too small. The presence of some of the few mutated peptides enabled, however, in some cases a very specific attribution. MALDI-TOF-MS has proven to be a powerful and fast technique for the analysis of enzymatically cleaved protein samples, but identification of peptides on their mass only imposes restrictions on its capabilities: interpretation becomes harder with increasing complexity of the protein mixtures. Moreover, since lists of characteristic markers are only collected for proteins that are known to occur in protein-rich paint binders, unexpected proteins are unlikely to be identified successfully.

Tandem mass spectrometry has the potential to overcome these drawbacks. While in single mass spectrometry only the mass(-to-charge ratio) of the peptides is determined, in tandem mass spectrometry a highly characteristic fragmentation pattern for each peptide is recorded. Since fragmentation favours peptide bonds, a series of smaller peptides are created, which allows to “read” the peptide sequence in its tandem mass spectrum. To do this for all peptides in a tryptic digest of a paint sample, prior chromatographic separation is necessary. Through the combination of nano-HPLC and electrospray ionisation quadrupole time-of-flight tandem mass spectrometry (ESI-QTOF-MS/MS), the individual tryptic peptides and thus all the proteins in a paint sample were identified with high confidence levels (chapter VI). The required paint sample size was estimated to be 5 µg, while no influence of the pigments present in a series of paint models was observed on the results.

Proteomics biochemists often deploy a strategy based on known protein sequences to interpret tandem mass spectrometry results: sequence libraries are used to calculate the possible tryptic peptides and their expected fragments in

tandem mass spectrometry. In this set-up, identification of the peptides is based on comparison of the experimental results (mass spectra) with calculated data. Through the use of public protein sequence libraries (SwissProt, Trembl, NCBI) and the search engine Mascot MS/MS ions search, the proteins constituting animal glue, egg white, egg yolk, milk casein and other protein-rich products in works of art were identified. In nearly all our measurements keratins were observed, probably partly introduced by the lengthy and laborious sample preparation, but also originating from the works of art. These contaminants are identified next to those of the binding media, by no means hampering their interpretation. In theory, this approach would enable species-specific determination of proteins: small mutations in the protein sequence can lead to different MS/MS mass spectra. For animal glues (collagen), however, this rarely works: collagen is evolutionarily well-conserved and has a repetitive amino acid sequence, which confuses the Mascot software. Moreover, for many species, the collagen sequences are missing in the sequence libraries, although this will surely improve over time as the libraries are rapidly growing.

The use of spectral libraries instead of sequence libraries improves the interpretation of these difficult cases: spectral matching avoids the issues of incomplete understanding of the fragmentation process, while it is not imperative to know the exact protein sequence. Because none of the publicly available spectral libraries fully suits the needs of a conservation science lab, a dedicated tandem mass spectral library for animal glues from different species was created using NIST's MS Search software package (chapter VII). Querying a spectrum of an unknown creates a list of comparable spectra occurring in a number of reference samples. Based on this, a conclusion can be drawn whether this spectrum represents a peptide occurring in animal glue, and if so, a conserved (common) peptide or a peptide carrying a characteristic mutation. The knowledge of which spectra do or do not occur in the studied species is growing with each analysis, also aided and complemented with positive identifications by Mascot. In future, this will help us to select even better those spectra that are specific for a given species.

Three proteomics-based peptide analysis techniques for proteinaceous materials in art are discussed. Throughout the research project these have been applied not only on paint samples, but also on samples from adhesives used in historical

objects. In some cases, depending on the amount of sample available, multiple methods were applied. The results are summarised in table IX.1.

Table IX.1. Overview of the results on works of art obtained with different analysis methods for proteins.

Object	Sample	AAA	HPLC-DAD	MALDI-TOF	HPLC-MS/MS
<i>Reliquary of the Virgin's veil</i> anon. pre-Eyckian	Black paint (multi-layer)	Mixture	Milk casein, egg yolk		
<i>Crucifixion with St Catherine and St Barbara</i> anon. pre-Eyckian	Orange paint	Animal glue	Animal glue		
	Brown/red paint	Animal glue	Animal glue		Animal glue
<i>St Ursula shrine</i> anon. pre-Eyckian	Red overpainting	Gum?	Milk casein		
<i>Altarpiece of St Margaret of Antioch</i> Master Paul of Levoča	Blue paint			Mammal glue	
	Green paint			Mammal glue	
	Ground layer			Mammal glue, egg yolk?	
<i>St Catherine Altarpiece</i> J. Beyaert	Adhesive layer for an applied relief				Cattle(?) glue
<i>Sermon of St John the Baptist</i> P. Breughel II	Ground layer				Rabbit glue
<i>Crucifixion</i> P. Breughel II and J. de Momper	Ground layer				Rabbit glue
<i>Anatomical model of an ear</i> anon. 19 th C	Papier-mâché adhesive				Sturgeon glue
<i>Portrait of Nicolaes van Bambeeck</i> Rembrandt	Lower red ground layer				Cattle(?) glue
	Upper grey ground layer			Mammal glue (cattle?) contaminant	Wheat flour

The identification of protein matter in samples from works of art is often only a small step in the entire analytical chain. A case study on Rembrandt's *Portrait of Nicolaes van Bambeeck* (chapter VIII) clearly demonstrated that often only a multi-analytical study is able to unravel conservation issues or art-historical questions.

All three methods discussed here have their merits and drawbacks. HPLC-MS/MS is without doubt the most powerful method due to the wealth of information obtained through tandem mass spectrometry, yet requiring extremely small sample amounts. However, it is an expensive approach (instrumentation and maintenance) and its laborious sample preparation and data treatment often make it effusive. For many applications, the quicker, cheaper and easier-to-interpret MALDI-TOF-MS method largely suffices. Compared with the two mass spectrometric instruments, HPLC-DAD, is only advisable as a complementary tool to amino acid analysis or py-GC-MS in labs that do not have access to the other techniques. As for other analytical techniques, the best fitted method(s) should be chosen in order to correctly answer the question.

Future perspectives

The road does not end here. An increasing number of research groups get involved in proteomics-based peptide analysis on art and archaeological objects. Many unresolved issues and opportunities remain.

Incomplete knowledge of proteins sequences, for example, frequently hinder the interpretation of the data. This is certainly the case for a niche science, such as conservation science, working outside of the main biotechnological topics. Sequence libraries are steadily and rapidly growing. The first sturgeon collagen sequences appeared only recently in public libraries(185) and tentative, unpublished but complete rabbit collagen sequences have been supplied to us only weeks ago by the research group of Matthew Collins. Species-specific peptides for an increasing number of species are now being enlisted. Notwithstanding the communication between the different research groups, an enormous opportunity exists for a common knowledge base for specific proteomics data used in art and archaeology. This exchange platform, in the form of a website, ideally assembles all available information on the studied materials: species, protein sequences, marker peptide masses and tandem mass spectra, combined with tools to search and interpret measurement data.

To develop the three peptide analysis methods that were discussed in chapters IV to VII, a series of pigmented and unpigmented paint models was used. These were

prepared during the research project and kept in darkness in a closed box in a room with temperature control. Data obtained on the samples from historical art objects seem in agreement with those obtained from the unaged paint models. The structural stability of proteins is also demonstrated by several studies on archaeological samples(78,86,89,153) and the (partly) sequenced mastodon and tyrannosaurus collagen(159,186). We did a pilot study with artificial aged (Atlas SUNTEST CPS+) paint models, but no significant degradation of the protein content could be observed. Nevertheless, a thorough multi-instrument study is needed. In particular, tandem mass spectrometry is an exquisite method to study alterations of protein sequences. Very few (recent) information is available in literature when protein degradation in paints is concerned.

Several research groups are focusing on a range of sources of proteins in art and archaeological objects, protein-bound paints, glues, mortars with proteinaceous additives, hair, leathers and food remains in pottery. Other applications are possible, as was demonstrated by the identification of flour in one of the ground layers of Rembrandt's *Portrait of Nicolaes van Bambeeck*. Flour, added as an additive to the paint, only contains a low amount of proteins, yet these were detected by HPLC-MS/MS. Contrarily, our pilot tests with HPLC-MS/MS on plant gums were unsuccessful, despite the presence of a relatively high amount of glycoproteins. It is yet unclear why this experiment failed. Where these proteins unreachable for trypsin because of their oligosaccharide side chains? Do these proteins have lysine and arginine residues in favourable places that would lead to measurable tryptic peptides, or should we consider other enzymes? Were they washed away in the solid phase extraction? In case the tryptic peptides were measured, can Mascot interpret them anyway, since few plant protein sequences are known? These and other research questions remain open for future research.

Several types of historical textiles exist, some of them with a proteinaceous backbone: wool (keratin) and silk (sericin and fibroin). Keratin is readily digested by trypsin and detected by peptide analysis. It would, however, require extra precautions to avoid keratin contamination during sample handling and preparation. In our preliminary tests we were able to identify silkworm (*Bombyx mori*) fibroin, but with a very low sequence coverage. Due to the distribution of

arginine and lysine residues, few measurable tryptic peptides were formed. Other enzymes (e.g. chymotrypsin) should be considered to improve detectability.

The road does not end here. I am convinced that peptide analysis will become the leading analytical technique for questions concerning proteins in conservation science, with an increasing number of applications.

BIBLIOGRAPHY

- (1) R. Smith, *The Artist's Handbook*, 3rd edition, Dorling Kindersley Publishers, 2009.
- (2) H. Kühn, *Farbmittel, Buchmalerei, Tafel- und Leinwandmalerei*, Reclam, Stuttgart, 1984.
- (3) H.-P. Schramm, B. Hering, *Historische Malmaterialien und ihre Identifizierung*, Akademische Druck- und Verlagsanstalt, Graz, 1988.
- (4) I.M. Bell, R.J. Clark, P.J. Gibbs, *Raman spectroscopic library of natural and synthetic pigments (pre- $\approx$ 1850 AD)*, Spectrochim. Acta, Part A 53 (1997) 2159–2179.
- (5) K. Castro, M. Pérez, M.D. Rodríguez-Laso, J.M. Madariaga, *FTIR Spectra Database of Inorganic Art Materials*, Anal. Chem. 75 (2003) 214A–221A.
- (6) K. Castro, A. Sarmiento, E. Princi, M. Perez-Alonso, M.D. Rodríguez-Laso, S. Vicini, J.M. Madariaga, E. Pedemonte, *Vibrational spectroscopy at the service of industrial archaeology: Nineteenth-century wallpaper*, Trends Anal. Chem. 26 (2007) 347–359.
- (7) S. Bruni, S. Caglio, V. Guglielmi, G. Poldi, *The joined use of n.i. spectroscopic analyses – FTIR, Raman, visible reflectance spectrometry and EDXRF – to study drawings and illuminated manuscripts*, Appl. Phys. A 92 (2008) 103–108.
- (8) M. Bouchard, R. Rivenc, C. Menke, T. Learner, *Micro-FTIR and Micro-Raman Study of Paints Used by Sam Francis*, E-PS 6 (2009) 27–37.
- (9) N. Khandekar, C. Mancusi-Ungaro, H. Cooper, C. Rosenberger, K. Eremin, K. Smith, J. Stenger, D. Kirby, *A Technical Analysis of Three Paintings Attributed to Jackson Pollock*, Stud. Conserv. 55 (2010) 204–215.
- (10) W. Fremout, S. Saverwyns, *Identification of synthetic organic pigments: the role of a comprehensive digital Raman spectral library*, J. Raman Spectrosc. 43 (2012) 1536–1544.
- (11) Z. Kaszowska, K. Malek, M. Pańczyk, A. Mikołajska, *A joint application of ATR-FTIR and SEM imaging with high spatial resolution: Identification and distribution of painting materials and their degradation products in paint cross sections*, Vib. Spectrosc. 65 (2013) 1–11.

- (12) J.S. Mills, R. White, *The organic chemistry of museum objects*, 2nd ed., Butterworth-Heinemann, Oxford, 1994.
- (13) J. Blasko, R. Kubinec, B. Husová, P. Příkryl, V. Pacáková, K. Stulík, J. Hradilová, *Gas chromatography/mass spectrometry of oils and oil binders in paintings*, J. Sep. Sci. 31 (2008) 1067–1073.
- (14) I. Bonaduce, 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 7 (2012).
- (15) V. Pittthard, P. Finch, T. Bayerova, *Direct chemolysis-gas chromatography-mass spectrometry for analysis of paint materials*, J. Sep. Sci. 27 (2004) 200–208.
- (16) S. Prati, S. Smith, G. Chiavari, *Characterisation of Siccative Oils, Resins and Pigments in Art Works by Thermochemolysis Coupled to Thermal Desorption and Pyrolysis GC and GC-MS*, Chromatographia 59 (2004) 227–231.
- (17) G. Chiavari, D. Fabbri, S. Prati, A. Zoppi, *Identification of indigo dyes in painting layers by pyrolysis methylation and silylation. A case study: "The dinner of emmaus" by G. Preti*, Chromatographia 61 (2005) 403–408.
- (18) T. Learner, *Analysis of modern paints*, Getty Conservation Institute, Los Angeles, 2004.
- (19) S.L. Vallance, *Critical Review: Applications of Chromatography in Art Conservation: Techniques Used for the Analysis and Identification of Proteinaceous and Gum Binding Media*, Analyst 122 (1997) 75R–81R.
- (20) I. Bonaduce, H. Brecolaki, M.P. Colombini, A. Lluveras, V. Restivo, E. Ribechini, *Gas chromatographic-mass spectrometric characterisation of plant gums in samples from painted works of art*, J. Chromatogr. A 1175 (2007) 275–282.
- (21) O. Chiantore, C. Riedo, D. Scalarone, *Gas chromatography-mass spectrometric analysis of products from on-line pyrolysis/silylation of plant gums used as binding media*, Int. J. Mass Spectrom. 284 (2009) 35–41.
- (22) C. Riedo, D. Scalarone, O. Chiantore, *Advances in identification of plant gums in cultural heritage by thermally assisted hydrolysis and methylation*, Anal. Bioanal. Chem. 396 (2009) 1559–1569.
- (23) R. Aruga, P. Mirti, A. Casoli, G. Palla, *Classification of ancient proteinaceous painting media by the joint use of pattern recognition and factor analysis on GC/MS data*, Fresen. J. Anal. Chem. 365 (1999) 559–566.
- (24) E. Kouloumpi, G. Lawson, V. Pavlidis, *The contribution of gas chromatography to the resynthesis of the post-Byzantine artist's technique*, Anal. Bioanal. Chem. 387 (2007) 803–812.

- (25) S.M. Halpine, *An investigation of artists' materials using amino acid analysis: introduction of the one-hour extraction method*, in: *Conservation Research 1995*, National Gallery of Art, Washington, 1995: pp. 29–69.
- (26) M.R. Schilling, H.P. Khanjian, L.A.C. Souza, *Gas chromatographic analysis of amino acids as ethyl chloroformate derivatives*, J. Am. Inst. Conserv. 35 (1996) 45–59.
- (27) J. Wouters, M. Van Bos, K. Lamens, *Baroque Stucco Marble Decorations. I. Preparation of Laboratory Replicas and Establishment of Criteria for Analytical Evaluation of Organic Materials*, Stud. Conserv. 45 (2000) 106–116.
- (28) D.A. Scott, N. Khandekar, M.R. Schilling, N. Turner, Y. Taniguchi, H. Khanjian, *Technical Examination of a Fifteenth-Century German Illuminated Manuscript on Paper: A Case Study in the Identification of Materials*, Stud. Conserv. 46 (2001) 93–108.
- (29) J. Peris-Vicente, J.V. Gimeno Adelantado, M.T. Doménech Carbó, R. Mateo Castro, F. Bosch Reig, *Characterization of proteinaceous glues in old paintings by separation of the o-phthalaldehyde derivatives of their amino acids by liquid chromatography with fluorescence detection*, Talanta 68 (2006) 1648–1654.
- (30) P. Přikryl, L. Havlíčková, V. Pacáková, J. Hradilová, K. Štulík, P. Hofta, *An evaluation of GC-MS and HPLC-FD methods for analysis of protein binders in paintings*, J. Sep. Sci. 29 (2006) 2653–2663.
- (31) B. Singer, R. McGuigan, *The Simultaneous Analysis of Proteins, Lipids, and Diterpenoid Resins Found in Cultural Objects.*, Ann. Chim. Rome 97 (2007) 405–417.
- (32) I. Bonaduce, C. Blaensdorf, P. Dietemann, M.P. Colombini, *The binding media of the polychromy of Qin Shihuang's Terracotta Army*, J. Cult. Herit. 9 (2008) 103–108.
- (33) M. Carbinì, R. Stevanato, M. Rovea, P. Traldi, D. Favretto, *Curie-point Pyrolysis-Gas Chromatography/Mass Spectrometry in the Art Field. 2 - The Characterization of Proteinaceous Binders*, Rapid Commun. Mass Spectrom. 10 (1996) 1240–1242.
- (34) G. Chiavari, N. Gandini, P. Russo, D. Fabbri, *Characterisation of standard tempera painting layers containing proteinaceous binders by pyrolysis (/methylation)-gas chromatography-mass spectrometry*, Chromatographia 47 (1998) 420–426.
- (35) S.A. Buckley, A.W. Stott, R.P. Evershed, *Studies of organic residues from ancient Egyptian mummies using high temperature-gas chromatography-mass spectrometry and sequential thermal desorption-gas chromatography-mass spectrometry and pyrolysis-gas chromatography-mass spectrometry*, Analyst 124 (1999) 443–452.
- (36) G. Chiavari, D. Fabbri, S. Prati, *Gas chromatographic-mass spectrometric analysis of products arising from pyrolysis of amino acids in the presence of hexamethyldisilazane*, J. Chromatogr. A 922 (2001) 235–241.

- (37) R. Mateo-Castro, J.V. Gimeno-Adelantado, F. Bosch-Reig, A. Doménech-Carbó, M.J. Casas-Catalán, L. Osete-Cortina, J. De la Cruz-Cañizares, M.T. Doménech-Carbó, *Identification by GC-FID and GC-MS of amino acids, fatty and bile acids in binding media used in works of art*, Fresen. J. Anal. Chem. 369 (2001) 642–646.
- (38) D.L. Nelson, M.M. Cox, *Lehninger Principles of Biochemistry, Fourth Edition*, W. H. Freeman, New York, 2004.
- (39) C. Walsh, *Posttranslational modification of proteins: expanding nature's inventory*, Roberts and Company Publishers, Englewood, 2006.
- (40) Y. Wang, H. Sunwoo, G. Cherian, J. Sim, *Fatty acid determination in chicken egg yolk: a comparison of different methods*, Poult. Sci. 79 (2000) 1168–1171.
- (41) Wikipedia: *Translation (biology)*, [http://en.wikipedia.org/wiki/Translation_\(biology\)](http://en.wikipedia.org/wiki/Translation_(biology)) (accessed: 23 Oct. 2012).
- (42) M.P. Colombini, F. Modugno, *Organic Mass Spectrometry in Art and Archaeology*, John Wiley & Sons, Chichester, 2009.
- (43) Wikipedia: *Collagen*, <http://en.wikipedia.org/wiki/Collagen> (accessed: 23 Oct. 2012).
- (44) E.C.Y. Li-Chan, H.-O. Kim, *Structure and Chemical Compositions of Eggs*, in: Y. Mine (Ed.), *Egg Bioscience and Biotechnology*, Wiley-Interscience, Hoboken, New Jersey, 2008: pp. 1–96.
- (45) T. Johansen, *How To Make Egg Tempera Paint*, http://www.paintmaking.com/grinding_egg_tempera.htm (accessed: 3 Jan. 2014).
- (46) R. Huopalahti, R. López-Fandiño, M. Anton, R. Schade, eds., *Bioactive egg compounds*, Springer-Verlag, Heidelberg, 2007.
- (47) Encyclopedia Britannica: *Low-density lipoprotein (physiology)*, <http://www.britannica.com/EBchecked/media/92254/Cutaway-view-of-a-low-density-lipoprotein-complex-The-LDL> (accessed: 23 Oct. 2012).
- (48) D.S. Horne, *Casein micelle structure: Models and muddles*, Curr. Opin. Colloid Interface Sci. 11 (2006) 148–153.
- (49) P.F. Fox, P.L.H. McSweeney, *Dairy chemistry and biochemistry*, Springer, London, 1998.
- (50) R. Wolbers, G. Landrey, *The use of direct reactive fluorescent dyes for the characterization of binding media in cross sectional examinations*, in: *Preprints of Papers Presented at the Fifteenth Annual Meeting, Vancouver, 1987*, American Institute for Conservation, Vancouver, 1987: pp. 168–202.
- (51) J. Van der Weerd, H. Brammer, J.J. Boon, R.M.A. Heeren, *Fourier Transform Infrared microscopic imaging of an embedded paint cross-section*, Appl. Spectrosc. 56 (2002) 275–283.

- (52) M. Spring, C. Ricci, D.A. Peggie, S.G. Kazarian, *ATR-FTIR imaging for the analysis of organic materials in paint cross sections: case studies on paint samples from the National Gallery, London*, Anal. Bioanal. Chem. 392 (2008) 37–45.
- (53) M.W. Duncan, H. Roder, S.W. Hunsucker, *Quantitative matrix-assisted laser desorption/ionization mass spectrometry*, Brief. Funct. Genomic Proteomic 7 (2008) 355–370.
- (54) A. Rizzo, *Progress in the application of ATR-FTIR microscopy to the study of multi-layered cross-sections from works of art*, Anal. Bioanal. Chem. 392 (2008) 47–55.
- (55) S. Prati, E. Joseph, G. Sciutto, R. Mazzeo, *New Advances in the Application of FTIR Microscopy and Spectroscopy for the Characterization of Artistic Materials*, Acc. Chem. Res. 43 (2010) 792–801.
- (56) A. Jurado-López, M.D. Luque de Castro, *Use of near infrared spectroscopy in a study of binding media used in paintings*, Anal. Bioanal. Chem. 380 (2004) 706–711.
- (57) P. Vandenabeele, B. Wehling, L. Moens, H. Edwards, M. De Reu, G. Van Hooydonk, *Analysis with micro-Raman spectroscopy of natural organic binding media and varnishes used in art*, Anal. Chim. Acta 407 (2000) 261–274.
- (58) A. Nevin, 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*, Anal. Chem. 79 (2007) 6143–6151.
- (59) A. Nevin, 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*, J. Raman Spectrosc. 39 (2008) 993–1000.
- (60) A. Nevin, S. Cather, D. Anglos, C. Fotakis, *Analysis of protein-based binding media found in paintings using laser induced fluorescence spectroscopy*, Anal. Chim. Acta (2006) 341–346.
- (61) A. Nevin, D. Comelli, G. Valentini, D. Anglos, A. Burnstock, S. Cather, R. Cubeddu, *Time-resolved fluorescence spectroscopy and imaging of proteinaceous binders used in paintings*, Anal. Bioanal. Chem. 388 (2007) 1897–1905.
- (62) M. Johnson, E. Packard, *Methods Used for the Identification of Binding Media in Italian Paintings of the Fifteenth and Sixteenth Centuries*, Stud. Conserv. 16 (1971) 145–164.
- (63) L. Kockaert, P. Gausset, M. Dubi-Rucquoy, *Detection of Ovalbumin in Paint Media by Immunofluorescence*, Stud. Conserv. 34 (1989) 183–188.

- (64) B. Ramirez-Barat, S. de la Vina, *Characterization of Proteins in Paint Media by Immunofluorescence. A Note on Methodological Aspects*, Stud. Conserv. 46 (2001) 282–288.
- (65) A. Heginbotham, V. Millay, M. Quick, *The Use of Immunofluorescence Microscopy and Enzyme-Linked Immunosorbent Assay as Complementary Techniques for Protein Identification in Artists' Materials*, J. Am. Inst. Conserv. 45 (2006) 89–105.
- (66) L.S. Dolci, G. Sciutto, M. Guardigli, M. Rizzoli, S. Prati, R. Mazzeo, A. Roda, *Ultrasensitive chemiluminescent immunochemical identification and localization of protein components in painting cross-sections by microscope low-light imaging*, Anal. Bioanal. Chem. 392 (2008) 29–35.
- (67) G. Sciutto, L.S. Dolci, A. Buragina, S. Prati, M. Guardigli, R. Mazzeo, A. Roda, *Development of a multiplexed chemiluminescent immunochemical imaging technique for the simultaneous localization of different proteins in painting micro cross-sections*, Anal. Bioanal. Chem. 399 (2010) 2889–2897.
- (68) L. Cartechini, M. Vagnini, M. Palmieri, L. Pitzurra, T. Mello, J. Mazurek, G. Chiari, *Immunodetection of Proteins in Ancient Paint Media*, Acc. Chem. Res. 43 (2010) 867–876.
- (69) J. Arslanoglu, S. Zaleski, J. Loike, *An improved method of protein localization in artworks through SERS nanotag-complexed antibodies*, Anal. Bioanal. Chem. 399 (2011) 2997–3010.
- (70) P.A. Klausmeyer, R.P. Albertson, M.R. Schmidt, R.T. Woodland, M. Blewett, *Analysis and Treatment of a Painting by Kees van Dongen: FTIR and ELISA as Complementary Techniques in the Analysis of Art Materials*, E-PS 6 (2009) 151–162.
- (71) J. Mazurek, M.R. Schilling, G. Chiari, A. Heginbotham, *Antibody assay to characterize binding media in paint*, in: J. Bridgland (Ed.), *Preprints of ICOM-CC, 15th Triennial Meeting, New Delhi, 2008*, ICOM Committee for Conservation, New Delhi, 2008: pp. 849–856.
- (72) S.P. Gygi, R. Aebersold, *Mass spectrometry and proteomics*, Curr. Opin. Chem. Biol. 4 (2000) 489–494.
- (73) K. Gevaert, J. Vandekerckhove, *Protein identification methods in proteomics*, Electrophoresis 21 (2000) 1145–1154.
- (74) R. Aebersold, M. Mann, *Mass spectrometry-based proteomics*, Nature 422 (2003) 198–207.
- (75) R. Hynek, S. Kuckova, J. Hradilová, 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 Commun. Mass Spectrom. 18 (2004) 1896–1900.

- (76) S. Kuckova, I. Nemec, R. Hynek, J. Hradilova, T. Grygar, *Analysis of organic colouring and binding components in colour layer of art works*, Anal. Bioanal. Chem. 382 (2005) 275–282.
- (77) S. Kuckova, R. Hynek, M. Kodicek, *Identification of proteinaceous binders used in artworks by MALDI-TOF mass spectrometry*, Anal. Bioanal. Chem. 388 (2007) 201–206.
- (78) K. Hollemeyer, W. Altmeyer, E. Heinzle, C. Pitra, *Species identification of Oetzi's clothing with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry based on peptide pattern similarities of hair digests*, Rapid Commun. Mass Spectrom. 22 (2008) 2751–2767.
- (79) S. Kuckova, M. Crhova, L. Vankova, R. Hynek, M. Kodicek, *Towards proteomic analysis of milk proteins in historical building materials*, Int. J. Mass Spectrom. 284 (2009) 42–46.
- (80) M. Buckley, 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 Commun. Mass Spectrom. 23 (2009) 3843–3854.
- (81) M. Buckley, S. Whitcher Kansa, S. Howard, S. Campbell, J. Thomas-Oates, M. Collins, *Distinguishing between archaeological sheep and goat bones using a single collagen peptide*, J. Archaeol. Sci. 37 (2010) 13–20.
- (82) D.P. Kirby, N. Khandekar, J. Arslanoglu, K. Sutherland, *Protein identification in artworks by peptide mass fingerprinting*, in: *Preprints of ICOM-CC, 16th Triennial Conference, Lisbon, 2011 (CD)*, ICOM Committee for Conservation, Lisbon, 2011.
- (83) J. Romero-Pastor, 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*, J. Mass Spectrom. 47 (2012) 322–330.
- (84) I.D. van der Werf, C.D. Calvano, F. Palmisano, L. Sabbatini, *A simple protocol for Matrix Assisted Laser Desorption Ionization- time of flight-mass spectrometry (MALDI-TOF-MS) analysis of lipids and proteins in single microsamples of paintings*, Anal. Chim. Acta 718 (2012) 1–10.
- (85) C. Tokarski, E. Martin, C. Rolando, C. Cren-Olive, *Identification of Proteins in Renaissance Paintings by Proteomics*, Anal. Chem. 78 (2006) 1494–1502.
- (86) C. Solazzo, W.W. Fitzhugh, C. Rolando, C. Tokarski, *Identification of Protein Remains in Archaeological Potsherds by Proteomics*, Anal. Chem. 80 (2008) 4590–4597.
- (87) G. Leo, L. Cartechini, P. Pucci, A. Sgamellotti, G. Marino, L. Birolo, *Proteomic strategies for the identification of proteinaceous binders in paintings*, Anal. Bioanal. Chem. 395 (2009) 2269–2280.

- (88) A. Chambery, 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/Q-q-TOF mass spectrometry: detection of different milk species by casein proteotypic peptides*, *Anal. Bioanal. Chem.* 395 (2009) 2281–2291.
- (89) S. Dallongeville, N. Garnier, D.B. Casasola, M. Bonifay, C. Rolando, C. Tokarski, *Dealing with the identification of protein species in ancient amphorae*, *Anal. Bioanal. Chem.* 399 (2011) 3053–3063.
- (90) G. Leo, 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*, *Anal. Chem.* 83 (2011) 2056–2064.
- (91) M.P. Colombini, 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*, *Stud. Conserv.* 43 (1998) 33–41.
- (92) A. Andreotti, I. Bonaduce, M.P. Colombini, G. Gautier, F. Modugno, E. Ribechini, *Combined GC/MS Analytical Procedure for the Characterization of Glycerolipid, Waxy, Resinous, and Proteinaceous Materials in a Unique Paint Microsample*, *Anal. Chem.* 78 (2006) 4490–4500.
- (93) M.R. Schilling, H.P. Khanjian, *Gas chromatographic analysis of amino acids as ethyl chloroformate derivatives*, *J. Am. Inst. Conserv.* 35 (1996) 123–144.
- (94) M.R. Schilling, H.P. Khanjian, *Gas Chromatographic analysis of amino acids as ethyl chloroformate derivatives - III. Identification of proteinaceous binding media by interpretation of amino acid composition data*, in: *Preprints of ICOM-CC, 11th Triennial Meeting, Edinburgh, 1996*, ICOM Committee for Conservation, Edinburgh, 1996: pp. 211–219.
- (95) M.P. Colombini, F. Modugno, M. Giacomelli, S. Francesconi, *Characterisation of proteinaceous binders and drying oils in wall painting samples by gas chromatography-mass spectrometry*, *J. Chromatogr. A* 846 (1999) 113–124.
- (96) G. Gautier, M.P. Colombini, *GC-MS identification of proteins in wall painting samples: A fast clean-up procedure to remove copper-based pigment interferences*, *Talanta* 73 (2007) 95–102.
- (97) D. Scalarone, M. Lazzari, O. Chiantore, *Ageing behaviour and pyrolytic characterisation of diterpenic resins used as art materials: colophony and Venice turpentine*, *J. Anal. Appl. Pyrol.* 64 (2002) 345–361.
- (98) D. Scalarone, M. Lazzari, O. Chiantore, *Ageing behaviour and analytical pyrolysis characterisation of diterpenic resins used as art materials: Manila copal and sandarac*, *J. Anal. Appl. Pyrol.* (2003) 115–136.

- (99) L. Osete-Cortina, M.T. Doménech-Carbó, *Analytical characterization of diterpenoid resins present in pictorial varnishes using pyrolysis–gas chromatography–mass spectrometry with on line trimethylsilylation*, J. Chromatogr. A 1065 (2005) 265–278.
- (100) H. Khanjian, R. Rivenc, M.R. Schilling, *A procedure for the efficient and simultaneous analysis of Asian and European lacquers in furniter of mixed origin*, in: *Preprints of ICOM-CC, 15th Triennial Meeting, New Delhi, 2008*, ICOM Committee for Conservation, New Delhi, 2008: pp. 1100–1108.
- (101) D.T. Burns, K. Doolan, *A comparison of pyrolysis–gas chromatography–mass spectrometry and Fourier transform infrared spectroscopy for the analysis of a series of modified alkyd paint resins*, Anal. Chim. Acta 422 (2000) 217–230.
- (102) L. Osete-Cortina, M. Doménech-Carbó, *Characterization of acrylic resins used for restoration of artworks by pyrolysis–silylation–gas chromatography/mass spectrometry with hexamethyldisilazane*, J. Chromatogr. A 1127 (2006) 228–236.
- (103) M. Doménech-Carbó, G. Bitossi, L. Osete-Cortina, J. de la Cruz-Cañizares, D. Yusá-Marco, *Characterization of polyvinyl resins used as binding media in paintings by pyrolysis–silylation–gas chromatography–mass spectrometry*, Anal. Bioanal. Chem. 391 (2008) 1371–1379.
- (104) J. Peris-Vicente, U. Baumer, H. Stege, K. Lutzenberger, J.V. Gimeno Adelantado, *Characterization of Commercial Synthetic Resins by Pyrolysis-Gas Chromatography/Mass Spectrometry: Application to Modern Art and Conservation*, Anal. Chem. 81 (2009) 3180–3187.
- (105) R.M. Derrick, S.C. Stulik, *Identification of natural gums in works of art using pyrolysis–gas chromatography*, in: *Preprints of ICOM-CC, 9th Triennial Meeting, Dresden, 1990*, ICOM Committee for Conservation, Dresden, Germany, 1990.
- (106) A. Asperger, W. Engewald, G. Fabian, *Advances in the analysis of natural waxes provided by thermally assisted hydrolysis and methylation (THM) in combination with GC/MS*, J. Anal. Appl. Pyrol. 52 (1999) 51–63.
- (107) I. Bonaduce, M.P. Colombini, *Characterisation of beeswax in works of art by gas chromatography–mass spectrometry and pyrolysis–gas chromatography–mass spectrometry procedures*, J. Chromatogr. A 1028 (2004) 297–306.
- (108) N. Sonoda, *Characterization of Organic Azo-Pigments by Pyrolysis-Gas Chromatography*, Stud. Conserv. 44 (1999) 195–208.
- (109) F. Cappitelli, T. Learner, O. Chiantore, *An initial assessment of thermally assisted hydrolysis and methylation–gas chromatography/mass spectrometry for the identification of oils from dried paint films*, J. Anal. Appl. Pyrol. 63 (2002) 339–348.

- (110) A. Andreotti, I. Bonaduce, M. Perla Colombini, F. Modugno, E. Ribechini, *A diagnosis of the yellowing of the marble high reliefs and the black decorations in the chapel of the tomb of Saint Anthony (Padua, Italy)*, *Int. J. Mass Spectrom.* 284 (2009) 123–130.
- (111) J. McCarthy, F. Hopwood, D. Oxley, M. Laver, A. Castagna, P.G. Righetti, K. Williams, B. Herbert, *Carbamylation of proteins in 2-D electrophoresis—myth or reality?*, *J. Proteome Res.* 2 (2003) 239–242.
- (112) N. Zhang, L. Li, *Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation*, *Rapid Commun. Mass Spectrom.* 18 (2004) 889–896.
- (113) M. Buckley, M. Collins, J. Thomas-Oates, *A method of isolating the collagen (I) α 2 chain carboxytelopeptide for species identification in bone fragments*, *Anal. Biochem.* 374 (2008) 325–334.
- (114) E.S. Grumbach, D.M. Wagrowski-Diehl, P.C. Iraneta, B. Alden, J.R. Mazzeo, *Hydrophilic Interaction Chromatography Using Silica Columns for the Retention of Polar Analytes and Enhanced ESI-MS Sensitivity*, *LCGC North America* 22 (2004) 1010–1023.
- (115) T. Yoshida, *Peptide separation by Hydrophilic-Interaction Chromatography: a review*, *J. Biochem. Bioph. Meth.* 60 (2004) 265–280.
- (116) P. Gates, *Mass Spectrometry Theory Index*, <http://www.chm.bris.ac.uk/ms/theory/theory.html> (accessed: 24 Oct. 2012).
- (117) *FindMod (ExPASy)*, <http://web.expasy.org/findmod/> (accessed: 9 Dec. 2012).
- (118) M. Strohal, M. Hassman, B. Kosata, M. Kodíček, *mMass data miner: an open source alternative for mass spectrometric data analysis*, *Rapid Commun. Mass Spectrom.* 22 (2008) 905–908.
- (119) M. Strohal, D. Kavan, P. Novák, M. Volný, V. Havlíček, *mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data*, *Anal. Chem.* 82 (2010) 4648–4651.
- (120) *mMass - Open Source Mass Spectrometry Tool*, <http://www.mmass.org> (accessed: 22 May 2013).
- (121) M. Dhaenens, *From dysregulated protein expression to histone clipping: the virtue of mass spectrometry: a proteomics tale of serendipity*, dissertation, Ghent University, 2011.
- (122) M.S. Wilm, M. Mann, *Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?*, *Int. J. Mass Spectrom.* 136 (1994) 167–180.
- (123) S.J. Gaskell, *Electrospray: Principles and Practice*, *J. Mass Spectrom.* 32 (1998) 677–688.

- (124) A. Constans, *Q-ToF Mass Spectrometer*, Scientist 20 (2006).
- (125) P. Roepstorff, J. Fohlman, *Letter to the editors*, Biol. Mass Spectrom. 11 (1984) 601.
- (126) UniProt, <http://www.uniprot.org/> (accessed: 31 Oct. 2013).
- (127) NCBI Protein, <http://www.ncbi.nlm.nih.gov/protein/> (accessed: 31 Oct. 2013).
- (128) D.J. Pappin, P. Hojrup, A.J. Bleasby, *Rapid identification of proteins by peptide-mass fingerprinting*, Curr. Biol. 3 (1993) 327–332.
- (129) D. Chiang, *Proteomics 2.0: Why Many Proteomic “Probabilities” Aren’t*, <http://www.proteomics2.com/?p=65> (accessed: 22 May 2013).
- (130) S.E. Stein, *Estimating probabilities of correct identification from results of mass spectral library searches*, J. Am. Soc. Mass Spectr. 5 (1994) 316–323.
- (131) H. Lam, R. Aebersold, *Building and searching tandem mass (MS/MS) spectral libraries for peptide identification in proteomics*, Methods 54 (2011) 424–431.
- (132) NIST Peptide Mass Spectral Reference Data, <http://peptide.nist.gov/> (accessed: 31 Oct. 2013).
- (133) GPMDB Proteome Database, <http://gpmdb.thegpm.org/> (accessed: 31 Oct. 2013).
- (134) Software:SpectraST - SPCTools, <http://tools.proteomecenter.org/wiki/index.php?title=Software%3ASpectraST> (accessed: 31 Oct. 2013).
- (135) X. Zhang, Y. Li, W. Shao, H. Lam, *Understanding the improved sensitivity of spectral library searching over sequence database searching in proteomics data analysis*, Proteomics 11 (2011) 1075–1085.
- (136) M.P. Colombini, F. Modugno, *Characterisation of proteinaceous binders in artistic paintings by chromatographic techniques*, J. Sep. Sci. 27 (2004) 147–160.
- (137) A. Casoli, P.C. Musini, G. Palla, *Gas chromatographic-mass spectrometric approach to the problem of characterizing binding media in paintings*, J. Chromatogr. A 731 (1996) 237–246.
- (138) M.P. Colombini, A. Carmignani, F. Modugno, F. Frezzato, A. Olchini, H. Brecolaki, V. Vassilopoulou, P. Karkanas, *Integrated analytical techniques for the study of ancient Greek polychromy*, Talanta 63 (2004) 839–848.
- (139) K. Sutherland, B.A. Price, I. Passeri, M. Tucker, *A Study of the Materials of Pontormo’s “Portrait of Alessandro de’ Medici,”* in: P.B. Vandiver, J.L. Mass, A. Murray (Eds.), *Materials Issues in Art and Archaeology VII (Mater. Res. Soc. Symp. Proc. 852)*, Warrendale, PA, USA, 2005: pp. 141–151.
- (140) G. Musumarra, M. Fichera, *Chemometrics and cultural heritage*, Chemom. Intell. Lab. 44 (1998) 363–372.

- (141) J.M. Messinger, *Ultraviolet-fluorescence microscopy of paint cross sections*, J. Am. Inst. Conserv. 31 (1992) 267–274.
- (142) M.T. Doménech Carbó, F. Bosch Reig, J.V. Gimeno Adelantado, V. Periz Martínez, *Fourier transform infrared spectroscopy and the analytical study of works of art for purposes of diagnosis and conservation*, Anal. Chim. Acta 330 (1996) 207–215.
- (143) G. Chiavari, D. Fabbri, S. Prati, *In-situ pyrolysis and silylation for analysis of lipid materials used in paint layers*, Chromatographia 53 (2001) 311–314.
- (144) E.S.B. Ferreira, J. van der Horst, J.J. Boon, *Chemical aspects of the binding media of the Oranjezaal ensemble: an insight into 17th century Netherlandish materials and methods*, in: *Preprints of ICOM-CC, 14th Triennial Meeting, the Hague, 2005*, ICOM Committee for Conservation, the Hague, 2005: pp. 774–782.
- (145) G. Chiavari, D. Fabbri, G.C. Galletti, S. Prati, N. Scianna, *Use of pyrolysis-gas chromatography/mass spectrometry to characterise binding media and protectives from a Coronelli's terrestrial globe*, J. Cult. Herit. 7 (2006) 67–70.
- (146) S. Kuckova, R. Hynek, M. Kodicek, *Analysis of art object using mass spectrometry*, in: *9th International Conference ART2008 Proceedings (CD)*, Jerusalem, 2008.
- (147) W. Fremout, S. Saverwyns, F. Peters, D. Deneffe, *Non-destructive micro-Raman and X-ray fluorescence spectroscopy on pre-Eyckian works of art-verification with the results obtained by destructive methods*, J. Raman Spectrosc. 37 (2006) 1035–1045.
- (148) D. Deneffe, F. Peters, W. Fremout, *Pre-Eyckian Panel Painting in the Low Countries, 1 Catalogue*, Brepols, KIK/IRPA, Brussels, 2009.
- (149) E. Dresner, M. Schubert, *The comparative susceptibility to collagenase and trypsin of collagen, soluble collagens and renal basement membrane*, J. Histochem. Cytochem. 3 (1955) 360–368.
- (150) S. Kuckova, I.C.A. Sandu, M. Crhova, R. Hynek, I. Fogas, S. Schafer, *Protein identification and localization using mass spectrometry and staining tests in cross-sections of polychrome samples*, J. Cult. Herit. 14 (2013) 31–37.
- (151) S. Kuckova, R. Hynek, M. Kodicek, *Application of peptide mass mapping on proteins in historical mortars*, J. Cult. Herit. 10 (2009) 244–247.
- (152) M.C. Krizkova, S.H. Kuckova, J. Santrucek, R. Hynek, *Peptide mass mapping as an effective tool for historical mortar analysis*, Constr. Build. Mat. 50 (2014) 219–225.
- (153) K. Hollemeyer, W. Altmeyer, E. Heinzle, C. Pitra, *Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry combined with multidimensional scaling, binary hierarchical cluster tree and selected diagnostic masses improves species identification of Neolithic keratin sequences from furs of the Tyrolean Iceman Oetzi*, Rapid Commun. Mass Spectrom. 26 (2012) 1735–1745.

- (154) S. Stevens, S. Wolverton, B. Venables, A. Barker, K. Seeley, P. Adhikari, *Evaluation of microwave-assisted enzymatic digestion and tandem mass spectrometry for the identification of protein residues from an inorganic solid matrix: implications in archaeological research*, *Anal. Bioanal. Chem.* 396 (2010) 1491–1499.
- (155) A. Barker, 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*, *J. Archaeol. Method Theory* 19 (2012) 407–439.
- (156) A. Shevchenko, 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*, *J. Proteomics* (in press).
- (157) M. Collins, M. Buckley, H.H. Grundy, J. Thomas-Oates, J. Wilson, N. van Doorn, *ZooMS: the collagen barcode and fingerprints*, *Spectroscopy Europe* 22 (2010) 6–10.
- (158) F. Šmigrovský, *Roman-Catholic church of St. Margaret of Antioch (Magrita Antiochijská) in Mlynica: the main altar of St. Margaret of Antioch*, in: *Collection of the Lectures of the 6th International Seminar on Restoration*, Obec Reštaurátorov Slovenska, Bardejov, Slovenska, 2006: pp. 87–91.
- (159) M. Buckley, N. Larkin, M. Collins, *Mammoth and Mastodon collagen sequences; survival and utility*, *Geochim. Cosmochim. Ac.* 75 (2011) 2007–2016.
- (160) T. Tripković, C. Charvy, S. Alves, A.Đ. Lolić, R.M. Baošić, S.D. Nikolić-Mandić, J.C. Tabet, *Identification of protein binders in artworks by MALDI-TOF/TOF tandem mass spectrometry*, *Talanta* 113 (2013) 49–61.
- (161) C. Hong, H. Jiang, E. Lü, Y. Wu, L. Guo, Y. Xie, C. Wang, Y. Yang, *Identification of Milk Component in Ancient Food Residue by Proteomics*, *PLoS ONE* 7 (2012) e37053.
- (162) I.D. van der Werf, C.D. Calvano, R. Laviano, A. Simonetti, L. Sabbatini, *Multi-technique chemical characterisation of a 12–13th-century painted Crucifix*, *Microchem. J.* 106 (2013) 87–94.
- (163) M.T. Doménech-Carbó, *Novel analytical methods for characterising binding media and protective coatings in artworks*, *Anal. Chim. Acta* 621 (2008) 109–139.
- (164) E. Manzano, N. Navas, R. Checa-Moreno, L. Rodríguez-Simón, L.F. Capitán-Vallvey, *Preliminary study of UV ageing process of proteinaceous paint binder by FT-IR and principal component analysis*, *Talanta* 77 (2009) 1724–1731.
- (165) M. Oh-Ishi, T. Maeda, *Separation techniques for high-molecular-mass proteins*, *J. Chromatogr. B* 771 (2002) 49–66.

- (166) B. Cañas, C. Piñeiro, E. Calvo, D. López-Ferrer, J.M. Gallardo, *Trends in sample preparation for classical and second generation proteomics*, J. Chromatogr. A 1153 (2007) 235–258.
- (167) H.P. Kasserra, K.J. Laidler, *Mechanisms of action of trypsin and chymotrypsin*, Can. J. Chem 47 (1969) 4031–4093.
- (168) R.S. Johnson, S.A. Martin, K. Biemann, J.T. Stults, J.T. Watson, *Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine*, Anal. Chem. 59 (1987) 2621–2625.
- (169) J.M. Froelich, G.E. Reid, *The origin and control of ex vivo oxidative peptide modifications prior to mass spectrometry analysis*, Proteomics 8 (2008) 1334–1345.
- (170) J. Williams, *A comparison of conalbumin and transferrin in the domestic fowl*, Biochem. J. 83 (1962) 355–364.
- (171) Y. Sugimoto, S. Sanuki, H.R. Ibrahim, T. Aoki, T. Kusakabe, K. KOGA, *Occurrence of ovalbumin in ovarian yolk of chicken during oogenesis*, Biochim. Biophys. Acta Gen. Subj. 1526 (2001) 1–4.
- (172) G. Corso, M. Gelzo, C. Sanges, A. Chambery, A.D. Maro, V. Severino, A.D. Russo, C. Piccioli, P. Arcari, *Polar and non-polar organic binder characterization in Pompeian wall paintings: comparison to a simulated painting mimicking an “a secco” technique*, Anal. Bioanal. Chem. 402 (2012) 3011–3016.
- (173) S. Kuckova, I.C.A. Sandu, M. Crhova, R. Hynek, I. Fogas, V.S. Muralha, A.V. Sandu, *Complementary cross-section based protocol of investigation of polychrome samples of a 16th century Moravian Sculpture by optical, vibrational and mass spectrometric techniques*, Microchem. J. 110 (2013) 538–544.
- (174) S. Dallongeville, M. Richter, S. Schäfer, M. Kuhlenthal, N. Garnier, C. Rolando, C. Tokarski, *Proteomics applied to the authentication of fish glue: application to a 17th century artwork sample*, Analyst 138 (2013) 5357–5364.
- (175) *Basic Local Alignment Search Tool (NCBI)*, <http://blast.ncbi.nlm.nih.gov/Blast.cgi> (accessed: 9 Dec. 2012).
- (176) H. Dubois, F. Rosier, L. De Belle, *Le Portrait de Nicolaes van Bambeeck par Rembrandt: Technique d’une oeuvre magistrale*, Bulletin de l’Institut Royal Du Patrimoine Artistique / Bulletin van Het Koninklijk Instituut Voor Het Kunstpatrimonium 33 (2013) 101–114.
- (177) K. Groen, *Grounds in Rembrandt’s workshop and in paintings by his contemporaries*, in: *A Corpus of Rembrandt Paintings IV: The Self-Portraits*, Stichting Foundation Rembrandt Research Project, Dordrecht, 2005: pp. 318–334.

- (178) J. Sanyova, M. Postec, K. Groen, *Fonction de la farine de blé dans la préparation du Portrait de Nicolaes van Bambeeck de Rembrandt* (Musées royaux des Beaux-Arts de Belgique, Bruxelles). *Approche expérimentale*, Techné 35 (2012) 14–23.
- (179) K.M. Groen, *Earth Matters: The origin of the material used for the preparation of the Night Watch and many other canvases in Rembrandt's workshop after 1640*, Art Matters 3 (2005) 138–154.
- (180) J. Sanyova, C. Glaude, W. Fremout, S. Saverwyns, *Le Portrait de Nicolaes van Bambeeck par Rembrandt: Analyse des matériaux*, Bulletin de l'Institut Royal Du Patrimoine Artistique / Bulletin van Het Koninklijk Instituut Voor Het Kunstpatrimonium 33 (2013) 115–126.
- (181) J. Sanyova, S. Cersoy, P. Richardin, O. Laprevote, P. Walter, A. Brunelle, *Unexpected Materials in a Rembrandt Painting Characterized by High Spatial Resolution Cluster-TOF-SIMS Imaging*, Anal. Chem. 83 (2011) 753–760.
- (182) P. Coremans, J. Thissen, *Het wetenschappelijk onderzoek van het Zelfportret van Stuttgart, Bijdrage tot de Rembrandt vorsing*, Bulletin de l'Institut Royal Du Patrimoine Artistique / Bulletin van Het Koninklijk Instituut Voor Het Kunstpatrimonium 8 (1964) 187–195.
- (183) D. Bomford, J. Koller, U. Baumer, *Nur eine Mischung aus Leinöl mit verschiedene Pigmenten? Die Bindmittel auf Rembrandts Ganymed aus der Galerie Alter Meister in Dresden*, in: *Rembrandt van Rijn: Die Entführung Des Ganymed*, Michel Sandstein und Staatliche Kunstsammlungen Dresden, Dresden, 2006: pp. 67–76.
- (184) M. Kusaba-Nakayama, M. Ki, M. Iwamoto, R. Shibata, M. Sato, K. Imaizumi, *CM3, one of the wheat α -amylase inhibitor subunits, and binding of IgE in sera from Japanese with atopic dermatitis related to wheat*, Food Chem. Toxicol. 38 (2000) 179–185.
- (185) S. Mizuta, C. Asano, Y. Yokoyama, M. Taniguchi, *Molecular species of collagen in muscular and vertebral parts of white sturgeon Acipenser transmontanus*, Fish Sci 78 (2012) 399–406.
- (186) J.M. Asara, M.H. Schweitzer, L.M. Freimark, M. Phillips, L.C. Cantley, *Protein Sequences from Mastodon and Tyrannosaurus Rex Revealed by Mass Spectrometry*, Science 316 (2007) 280–285.

DUTCH SUMMARY

SAMENVATTING

De analyse van kunstenaarsverven gebonden met proteïnes en andere proteïneuze materialen in kunstvoorwerpen vormde tijdens het laatste decennium het onderwerp van vele onderzoeksprojecten. De traditionele analysemethoden, gebaseerd op aminozuuranalyse (AZA) of pyrolyse, schieten vaak tekort door de lage specificiteit van individuele aminozuren en de nog kleinere pyrolyseproducten. Dit leidt vaak tot dubieuze resultaten wanneer mengsels van proteïnes, gecontamineerde stalen of stalen met onverwachte proteïnes bestudeerd worden. De ontwikkeling van nieuwe methodes voor de identificatie van proteïnes in kunstobjecten blijkt voornamelijk gericht op twee verschillende technieken die gebruikt worden in *proteomics*: immunokleuring en peptideanalyse.

Immunokleuring (Eng.: *immunostaining*) is gebaseerd op een specifieke antilichaam-antigen-reactie, waarin gekleurde antilichamen slechts aan één welbepaalde proteïne (tot op soort-niveau) binden. Een aantal varianten van deze techniek hebben de interessante eigenschap dat ze rechtstreeks op dwarsdoorsneden kunnen toegepast worden. Zo kan de aanwezigheid en verdeling van een specifieke proteïne (en dus het bindmiddel) in de opeenvolgende verflagen gevisualiseerd worden.

Peptideanalyse is gebaseerd op de detectie en identificatie van peptiden. Deze worden verkregen door voorafgaandelijke enzymatische afbraak van de proteïnes uit het verfstaal. Trypsine is hiervoor het meest gebruikte enzym en tot op vandaag is het het enige dat gebruikt wordt voor de analyse van staaltjes van kunstobjecten. Het blijkt namelijk uiterst geschikt voor de proteïnes die voorkomen in bindmiddelen in verf: het gros van de ontstane peptiden bestaat uit 5 tot 15 aminozuren. Dit is lang genoeg om uniek te zijn (karakteristiek voor het proteïne

waar ze uit gemaakt zijn) en kort genoeg om een detectie met hoge precisie toe te staan met behulp van massaspectrometrie. Peptideanalyse staat hierdoor garant voor de laagste detectielimieten, de hoogste specificiteit en de mogelijkheid om alle proteïnes in een enkel staal in één enkele meting te bepalen. Hiertegenover staat dat verfstaaftjes van een enkele laag vereist zijn. Niet enkel zijn deze moeilijk te bemonsteren, de verdeling van de proteïnes over de verflagen kan niet worden bestudeerd.

Peptideanalyse kan uitgevoerd worden met behulp van verschillende types detectoren en bijpassende manieren van interpretatie van de meetgegevens. In deze doctoraatsverhandeling werden drie technieken uitgewerkt of verbeterd. In de eenvoudigste opstelling (hoofdstuk IV) worden de peptiden gescheiden met hoge performantie vloeistofchromatografie en vervolgens gedetecteerd met een eenvoudige diodedetector (HPLC-DAD). Deze instrumentatie is goedkoop en aanwezig in de meeste laboratoria waar kunstvoorwerpen worden bestudeerd. De interpretatie van de gegevens is eenvoudig: in het ideale geval stelt elke piek in het chromatogram een tryptische peptide voor, gekarakteriseerd door zijn retentiefactor. De identificatie van het proteïnehoudende product is mogelijk door vergelijking met de chromatogrammen van referentiestalen. De geringste verandering van de experimentele omstandigheden kan echter de retentietijden beïnvloeden en zo de interpretatie bemoeilijken. Identificatie van de peptidesequentie is niet mogelijk met UV spectra, noch kan deze techniek gebruikt worden voor de studie van verouderingsprocessen van proteïneuze bindmiddelen in kunstobjecten. Toch vormt deze techniek een nuttige aanvulling op andere analytische technieken zoals pyrolyse-gaschromatografie of aminozuuranalyse. De benodigde hoeveelheid staal ligt in de lijn van deze technieken (zo'n 100 tot 200 µg).

Het bepalen van de massa van de peptiden in plaats van enkel de retentiefactor vereenvoudigt de interpretatie gevoelig. Door gebruik te maken van een zachte ionisatietechniek gekoppeld aan een massaspectrometer, zoals *matrix-assisted laser desorption/ionisation time-of-flight* massaspectrometrie (MALDI-TOF-MS), wordt de massa van de enkelvoudig geladen individuele tryptische peptiden bepaald. In *proteomics* worden deze MALDI spectra vergeleken met de berekende peptiden voor proteïnen waarvan de sequentie bekend is. Deze techniek, gekend

onder de term *peptide mass fingerprinting* (PMF), is erg geschikt voor stalen die slechts één enkele proteïne bevatten. Verfstalen zijn echter meestal een complex mengsel van meerdere proteïnes. Daarom worden de MALDI-spectra voor bindmiddelonderzoek gewoonlijk vergeleken met deze van referentiestalen van bindmiddelen. Door gebruik te maken van hoofdcomponentenanalyse (*principal component analysis*, PCA) en *soft independent modelling of class analogies* (SIMCA) konden de meest karakteristieke pieken bepaald worden voor de vier hoofdgroepen van proteïneuze bindmiddelen (dierlijke lijm, eiwit, eigeel en melkcaseïne), terwijl deze van eventueel aanwezige pigmenten, pigment-bindmiddel-complexen en contaminanten uitgesloten werden (hoofdstuk V). Van het merendeel van deze karakteristieke pieken werden de corresponderende peptidesequenties berekend op basis van gekende proteïnesequenties. Dierlijke lijmen bereid uit vis en zoogdieren konden van elkaar worden onderscheiden, maar binnen de groep van de zoogdieren was dit niet het geval: de evolutionaire verschillen in hun collageensequenties waren vaak te klein. Aan de hand van deze specifieke mutaties kon een dierlijke lijm in enkele gevallen toch toegewezen worden aan een diersoort. MALDI-TOF-MS is een krachtige en snelle techniek voor de analyse van tryptische digesten, maar de identificatie van peptiden enkel op basis van hun massa kan een beperking zijn: de interpretatie wordt moeilijker met de toenemende complexiteit van de proteïnemengsels. Aangezien lijsten met karakteristieke massa's enkel aangelegd worden voor de proteïnehoudende producten waarvan hun gebruik in kunstobjecten gekend is, zullen onverwachte proteïnes wellicht niet eenvoudig kunnen worden geïdentificeerd.

Tandemmassaspectrometrie heeft het potentieel om deze euvels te verhelpen. Terwijl in enkelvoudige massaspectrometrie enkel de massa(-tot-lading) van de peptiden wordt bepaald, worden in tandemmassaspectrometrie karakteristieke fragmentatiepatronen bepaald voor elke peptide. Gezien deze fragmentatie voornamelijk plaatsvindt op de peptidebinding, ontstaat een serie van kleinere peptiden, wat toelaat om de peptidesequentie te "lezen" in het tandemmassaspectrum. Om dit te kunnen doen voor elke peptide in een tryptisch digest van een verfstaal, is een voorafgaande chromatografische scheiding noodzakelijk. Door de combinatie van nano-HPLC en electrospray ionisatie quadrapool *time-of-flight* massaspectrometrie (ESI-QTOF-MS), werden de

individuele tryptische peptiden, en als dusdanig de proteïnes in een verfstaal, geïdentificeerd met zeer hoge confidentieniveaus (hoofdstuk VI). De hoeveelheid verfstaal geconsumeerd door deze techniek werd geschat op zo'n 5 µg. Geen significante invloed van pigmenten kon worden vastgesteld in de metingen op een serie van gepigmenteerde verfstalen.

Proteomics biochemici maken vaak gebruik van een strategie op basis van gekende proteïnesequenties om tandemmassaspectrometrische meetgegevens te interpreteren: aan de hand van sequentiebibliotheken worden mogelijke tryptische peptiden en hun verwachte fragmenten berekend. Deze worden dan vergeleken met de gemeten fragmenten in de tandemmassaspectra en derwijze worden de peptiden en de proteïnes geïdentificeerd. Door gebruik te maken van publieke bibliotheken met proteïnesequenties (SwissProt, Trembl, NCBI) en de zoekmachine Mascot MS/MS ions search, werden de proteïnes aanwezig in dierlijke lijm, eiwit, eigeel, melkcaseïne en andere proteïnehoudende producten in kunstobjecten geïdentificeerd. In bijna alle metingen werden daarnaast ook keratines geïdentificeerd, waarschijnlijk geïntroduceerd tijdens de langdurige en arbeidsintensieve staalvoorbereiding, maar misschien ook afkomstig uit de kunstobjecten zelf. Deze contaminanten werden onafhankelijk van de bindmiddelen geïdentificeerd, zonder dat de interpretatie hierdoor verstoord werd. In theorie is tevens een soortspecifieke bepaling van proteïnes mogelijk: de kleine mutaties in de aminozuursequenties leiden namelijk tot afwijkende tandemmassaspectra. Voor dierlijke lijmen (collagenen) is dit echter zelden succesvol: collageen is evolutionair goed geconserveerd en heeft een repetitieve sequentie die Mascot dikwijls in de war brengt. Bovendien zijn voor veel diersoorten de collageensequenties nog niet gekend en kunnen ze dus niet verwerkt worden door Mascot; deze situatie zal zeker verbeteren gezien de huidige snelle aangroei van de sequentiebibliotheken.

Het gebruik van spectrale bibliotheken in plaats van sequentiebibliotheken verbetert op korte termijn de interpretatie van deze moeilijke gevallen. Enerzijds vermijdt deze aanpak de onvolledige kennis over de fragmentatieprocessen en anderzijds is het niet strikt noodzakelijk dat de exacte proteïnesequenties volledig gekend zijn. Helaas zijn geen van de publiek verkrijgbare spectrale bibliotheken specifiek gericht op het onderzoek naar de proteïnes in kunstobjecten. Daarom

werd met behulp van software van het NIST een eigen bibliotheek voor tandemmassaspectra opgesteld gericht op de soortspecifieke bepaling van dierlijke lijmen (hoofdstuk VII). Het zoekprogramma berekent voor een onbekend spectrum de meest gelijkende spectra die voorkomen in een aantal van de referentiestalen van dierlijke lijmen. Hieruit kan vervolgens besloten worden of dit spectrum toebehoort aan een geconserveerd peptide, of aan een peptide met een mutatie. De kennis van welke spectra voorkomen in welke diersoorten en welke niet, wordt voortdurend bijgewerkt, tevens geholpen door identificatie met behulp van Mascot.

Doorheen deze doctoraatsverhandeling kwamen drie technieken voor peptideanalyse gebaseerd op proteomics aan bod. Deze werden specifiek ontwikkeld voor de analyse van verfstalen, maar werden ook gebruikt voor de bepaling van lijm in diverse historische objecten. Afhankelijk van de beschikbare hoeveelheid staal, werden ze soms met verschillende technieken gemeten. De resultaten zijn samengevat in tabel 1.

De identificatie van proteïnehoudende materie in stalen van kunstobjecten is vaak slechts één stap in een veel ruimere analytische keten. De studie van Rembrandts *Portret van Nicolaes van Bambeeck* (hoofdstuk VIII) vormt een mooi voorbeeld van hoe men door de combinatie van verschillende analytische technieken conservatiespecifieke of kunsthistorische vragen kan oplossen.

De drie besproken analytische methodes hebben elk hun voor- en nadelen. HPLC-ESI-MS is met voorsprong de krachtigste methode door de schat aan informatie die het door tandem massaspectrometrie oplevert, terwijl extreem weinig staal vereist is. Echter, de dure instrumentatie en de arbeidsintensieve staalvoorbereiding en gegevensverwerking kunnen niet voor elk staal verantwoord worden. In deze gevallen volstaat de snellere, goedkopere en eenvoudig te interpreteren MALDI-TOF-MS metingen. Vergeleken met deze twee technieken, is HPLC-DAD enkel aan te raden als een complementaire methode voor aminozuuranalyse en pyrolyse en wanneer geen andere peptideanalysemethoden beschikbaar zijn. Zoals in alle disciplines hangt het correct beantwoorden van een analytische vraag af van een weloverwogen keuze van de analysetechniek(en).

Tabel 1. Overzicht van de resultaten op kunstobjecten met behulp van verschillende technieken voor de analyse van proteïnes.

Object	Staal	AZA
<i>Reliekschrijn met de voile van de H. Maagd</i> anoniem pre-Eyckiaans	Zwarte verf (meerdere lagen)	Mengsel
<i>Kruisiging met St.-Katarina en St.-Barbara</i> anoniem pre-Eyckiaans	Oranje verf Roodbruine verf	Dierlijke lijm Dierlijke lijm
<i>St.-Ursulaschrijn</i> anoniem pre-Eyckiaans	Rode overschildering	Gom?
<i>Retabel van St.-Margaretha van Antiochië</i> Master Paul of Levoča	Blauwe verf Groene verf Grondlaag	
<i>Retabel van St.-Katarina</i> J. Beyaert	Lijm voor persbrokaat	
<i>Epistel van Johannes de Doper</i> P. Breughel de Jonge	Grondlaag	
<i>Kruisiging</i> P. Breughel de Jonge en J. de Momper	Grondlaag	
<i>Anatomisch model van een oor</i> anoniem 19 ^{de} eeuw	Lijm voor papier-maché	
<i>Portret van Nicolaes van Bambeeck</i> Rembrandt	Onderste rode grondlaag Bovenste grijze grondlaag	

HPLC-DAD	MALDI-TOF	HPLC-MS/MS
Melkcaseïne, eigeel		
Dierlijke lijm		
Dierlijke lijm		Dierlijke lijm
Melkcaseïne		
	Zoogdierlijke lijm	
	Zoogdierlijke lijm	
	Zoogdierlijke lijm, eigeel?	
		Runder(?)lijm
		Konijnenlijm
		Konijnenlijm
		Steurlijm
		Runder(?)lijm
	Runder(?)lijm contaminant	Tarwebloem

ACKNOWLEDGEMENTS

DANKWOORD

Hoe langer een doctoraatsonderzoek duurt, des te meer mensen een plaatsje verdienen in het dankwoord. Mijn idee om dit in een vijftal regeltjes af te handelen kan ik dus wel opbergen – geen excuses, ik heb dit mezelf aangedaan! Toch ben ik hen allen erg dankbaar, want zonder de steun van deze mensen had ik dit nooit voor mekaar gekregen.

Ik ben het Federaal Wetenschapsbeleid erg erkentelijk. Het gros van de financiële steun kwam uit hun portemonnee. Zonder centen geraak je niet ver, en al zeker niet in het wetenschappelijk onderzoek.

Wellicht heb ik er het imago van de onzichtbare en eeuwige doctoraatsstudent. De S12 en daarvoor het INW heeft mij de jongste jaren weinig gezien. En toch kon ik al die tijd op de steun en het vertrouwen van de vakgroep rekenen. Ik wil dan ook Karel, en bij uitbreiding iedereen in de vakgroep bedanken.

Op het KIK hebben ze mij dan misschien weer te vaak gezien. Myriam, Christina en Hilde – het matriarchale triumviraat van het KIK – hebben mijn “projectje” vanaf de conceptie tot het bittere einde gesteund. Een dikke merci ook voor alle collega's in de labo's. Ik waag me niet aan een opsomming, want de kans dat ik mensen vergeet, is te groot. Ik geef toe, ik was maar al te graag het manusje-van-alles, vooral bij technische mankementjes, maar jullie waren er ook steeds voor mij.

Ik had het privilege om maar liefst vier begeleiders mee te krijgen. Eerlijk verdeeld, twee in Gent, twee in Brussel. Luc, dank je voor de grote interesse in mijn werk en de voortdurende steun ondanks je ongelooflijk drukke agenda. Steven, jij was bijgevolg het slachtoffer van mijn meeste verzuchtingen. In tegenstelling tot wat de chaos op en rond je bureau doet vermoeden, had je altijd een helder beeld,

een strategie voor ogen. Een wonder lijkt het mij. Peter, jouw steun was onontbeerlijk. Je kon de druk opvoeren als de beste, maar ik besef dat je daarmee wou verhinderen dat het zou aanslepen en tenslotte verzanden. Et finalement, Jana, l'inspiratrice originale du sujet de ma thèse, j'admire ta connaissance des matériaux et les processus de dégradation dans les œuvres d'art. De weg die ik met mijn onderzoek ben ingeslagen was nieuw voor ons allemaal; ik vermoed en hoop dat het ook voor jullie een leerrijke trip geworden is.

Tijdens mijn onderzoek heb ik met veel mensen mogen samenwerken. Je veux remercier Mathieu Douineau et Marie Poellaer (en tant que étudiants de conservation-restauration à la Cambre) pour les préparations des peintures protéineuses. Julia Schultz (HAWK University of Applied Sciences and Arts) has very kindly shared her self-prepared set of animal glues. In de loop der jaren heb ik op heel wat interessante kunstvoorwerpen mogen werken; bedankt aan alle toeleveranciers.

En dan moet ik eigenlijk nog talloze anderen bedanken voor de kleine of minder kleine hulp. Daarvan springen er twee in het oog. Zonder hen zou dit een flinterdun boekje geworden zijn. Contrasting to the bitter winter coldness during my stay, was the warmth and hospitality of the people of the department of biochemistry and microbiology in the VŠCHT, Prague. Thanks go to Prof. Kodicek, Radovan and all others in the MALDI lab. Stepanka, you definitely have become one of my tutors in the field of proteomics. And a friend. Thanks. Het enthousiasme waarmee ik ontvangen ben in het labo voor farmaceutische biotechnologie (UGent, Farmaceutische Wetenschappen) ga ik nooit vergeten. Reeds enkele minuten na mijn mailtje kreeg ik van Prof. Deforce, Dieter, al het bericht dat we welkom waren. En dat ben ik altijd gebleven. Maarten, jouw gedrevenheid werkte aanstekelijk, jouw kennis over peptiden en mass specs deden mij het licht zien. Een hele dikke merci, maat! Bedankt ook aan alle anderen in de bokaal voor hun gastvrijheid, interesse en hulp gedurende de vele dagen dat ik er te gast was.

Een dergelijke onderneming vereist een solide basis. En dat was er, onder de vorm van familie en vrienden. Ik heb jullie nooit veel verteld over mijn onderzoek, maar heb vaak verstek moeten geven omdat ik weer eens bezig was met mijn trage

schrijfproces. Merci voor het geduld en de steun. Een bijzonder woordje van dank gaat uit naar mijn ouders. De kinderdroom, de professor-in-alles uit de stripwereld zal wel nooit werkelijkheid worden, maar ik weet dat jullie heel trots zijn. Dat mag gerust, want dit heb ik bereikt dankzij de goede opvoeding die ik van jullie gekregen heb .

Tot slot een tip voor (toekomstige) doctorandi: een fulltimejob in Brussel, een doctoraat als bezigheid tijdens de koude winteravonden én Bob-de-verbouwer tijdens de weekends (bij voorkeur ruim anderhalf uur verwijderd van de werkplek): niet doen! Dank je, Klaas, om op mij in te praten als ik er weer eens de brui aan wou geven, voor alle hulp en om er altijd voor mij te zijn. Jij was, en bent, mijn kracht en doorzettingsvermogen.

Dank je. Thank you. Merci.