Assessment of Raman microscopy coupled with principal component analysis to examine egg yolk–pigment interaction based on the protein C–H stretching region (3100–2800 cm\(^{-1}\))

Julia Romero-Pastor,\(^a\) Carolina Cardell,\(^a\) Eloisa Manzano,\(^b\) África Yebra-Rodríguez\(^c\) and Natalia Navas\(^b\)∗

This work seeks to identify the slight changes in the characteristic C–H stretching region (3100–2800 cm\(^{-1}\)) of a protein-based binder and fatty acid esters from egg yolk, which may occur in complex paint samples due to the presence of particular pigments. To date, this protein region – where historic pigments do not show characteristic Raman bands – has not been used to identify possible interactions between painting materials, in spite of its potential due to the mentioned feature. This study is based on the investigation of pure egg yolk model samples and tempera model samples prepared by mixing this binder with some historic pigments (cinnabar, raw Sienna, lead white, gypsum, azurite, lapis lazuli and smalt) as binary samples. All samples were analyzed in this region by Raman microscopy (RM) coupled with principal component analysis (PCA) for three color groups (red, white and blue) separately. The results show relevant spectral changes in the C–H stretching region of amino acids and polyunsaturated fatty acids esters of the egg yolk binder, particularly in the azurite, lead white and gypsum-based tempera samples. Lesser interactions were discerned in the tempera samples made with smalt, as well as shift in the region of polyunsaturated fatty acid esters of the egg yolk binder in the cinnabar and raw Sienna-based tempera paintings. No interactions were recognized between the egg yolk and the pigments calcite and lapis lazuli. The effectiveness of applying RM combined with PCA for identifying interaction processes between binders and pigments is demonstrated. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: proteinaceous binding media; historic pigments; C–H stretching region; Raman microscopy; PCA

Introduction

It is highly worthwhile to characterize artworks by nondestructive analytical techniques combined with statistical methods such as multivariate chemometric tools to extract the maximum information concerning composition, and to detect slight disparities in spectroscopic data.\(^{[1,2]}\) One of the most applied chemometric analyses in the field of cultural heritage is principal component analysis (PCA). PCA has been used successfully for classification purposes on pottery and paintings,\(^{[3,4]}\) identification of pigments and binders in historical and model paintings samples,\(^{[5]}\) aging detection on painting materials,\(^{[6]}\) dating purposes,\(^{[7,8]}\) recognition of different sources of raw materials and characterization of alteration products.\(^{[9,10]}\) Such analyses have been applied to data obtained by diverse analytical techniques such as inductively coupled plasma mass spectrometry, X-ray fluorescence, gas chromatography-mass spectrometry, Fourier transform infrared spectroscopy (FT-IR), diffuse reflectance infrared Fourier transform spectroscopy or Raman microscopy (RM). The main benefit of PCA is its capability to reduce data dimensionality to principal components (PCs), which contain more interpretable and representative information of the system under investigation.

At present, RM is one of the most commonly used analytical techniques to study painting materials both organic and inorganic simultaneously in the same sample, including binders, varnishes, pigments and dyes.\(^{[11]}\) The identification of these compounds using RM is based on the detection of characteristic bands in the spectra, particularly in the fingerprint region.\(^{[12,13]}\)

The study of protein-based binding media (e.g. animal glues, casein, egg, etc.) using Raman spectroscopy is centered on the main structure of amino acid compounds (with the general formula H\(_2\)NCHR\(_{2}\)COOH, where R is an organic substitute). This formula varies according to the different amino acids present, and is based on an amine group (N–H bond), a carboxylic acid group (C=O bond) and its characteristic side chain.\(^{[14]}\) According to the most...
recent studies, the main bands assigned to the presence of egg yolk are the amide bands I and III located at approximately 1670 and 1250 cm$^{-1}$, respectively. In addition, egg yolk contains 40% fatty acid esters, with a characteristic Raman feature at 2940 cm$^{-1}$, and phenylalanine (Phe) and tyrosine (Tyr) (aromatic amino acids) both showing the C–H stretching modes at 3060 cm$^{-1}$. All these bands, found at high Raman wavenumbers, have been considered less for study since their spectral details are less than those in the fingerprint region. However, this region has recently proved its importance in the identification of different kinds of proteins and in the discrimination of fresh versus aged protein model samples studied by RM. Moreover, this region has also demonstrated its capability to discriminate complex painting model samples based on the aging state of the protein, using FT-IR spectroscopic data. The importance of studying this characteristic region in complex painting samples with RM arises from the fact that historic pigments do not show characteristic bands here. Thus it provides an exceptional opportunity to track the interactions between the protein binder and diverse pigments.

It is well known that the characterization of proteinaceous binders (such as egg yolk) and their stability in complex paint samples can be conditioned by the presence of particular pigments and extenders. Thus, the role of pigments in the study of proteinaceous binding media in paintings is an important task which must be investigated in depth. In this sense, the benefit of applying RM data coupled with PCA to detect pigments and protein binders in tempera model samples has been demonstrated. Nevertheless, to the authors’ knowledge, PCA has not been applied to RM data from paintings to detect changes and interferences in the C–H stretching region of proteinaceous binders.

This work demonstrates the application of PCA to this characteristic protein region (where the C–H stretching modes of amino acids and esters of fatty acids appear) in three different pigment colors (blue, red and white) by varied model painting samples. The goal is to identify slight changes in these vibrational bands arising from the binder, i.e. natural egg yolk, as resulting from the interaction with some historic pigments, namely cinnabar and raw Sienna as red pigments; lead white, gypsum and calcite as white pigments; and azurite, lapis lazuli and smalt as blue pigments. In this way, it is expected to shed light on the interaction between this binder and the mentioned pigments in their temperas, as well as to probe the usefulness of using multivariate chemometric approaches on Raman spectroscopic data as a valuable tool for investigating and characterizing painting materials.

Materials and Methods

Painting materials

The azurite and smalt pigments were purchased from Kremer Pigments GmbH & Co. KG (Madrid, Spain), while the rest of the samples were supplied by Caremi Pigments (Sevilla, Spain). Cinnabar is mercuric sulfide (HgS) and raw Sienna is a natural iron-rich earth pigment, i.e. a clay-rich material having a complex composition (gypsum, anhydrite, quartz, calcite, dolomite, etc.). Regarding the white pigments, lead white or cerusite is the chemical compound (PbCO$_3$); calcite (or chalk) is a natural white pigment made of calcium carbonate (CaCO$_3$); and gypsum is calcium sulfate (CaSO$_4$·2H$_2$O); the latter two having been used traditionally in ground layers and as fillers or extenders. The blue pigments selected were the hydrated copper carbonate named azurite (Cu$_2$(CO$_3$)$_2$(OH)$_2$), lapis lazuli whose main constituent is lazurite (Na$_{8-10}$Al$_6$Si$_6$O$_{24}$S$_2$–4), and smalt which is a synthetic pigment made of coarsely ground potassium cobalt glass strongly colored with cobalt oxide.

Painting model samples

Nine model paint samples were prepared according to ‘Old Master recipes’ to obtain egg yolk tempera painting standards similar to those used by medieval artists. Tempera is a painting technique in which finely ground pigments are first mixed with water and later blended with a solidifying proteinaceous binding agent, such as egg, animal glue or casein. The first model sample was the pure egg yolk binder. To prepare the binder, the egg yolk was separated from the white and spread out with a paintbrush in a fine coat on a glass slide.

The eight remaining model samples were paint binary mixture samples composed of pure pigments blended with the egg yolk binder; thus the tempera replicate samples were obtained. They were prepared by mixing the pigment with egg yolk. Approximately 0.5 g of each pigment powder was formed as a crater-shaped mass to which a few drops of egg yolk (different number according to each pigment) were added until a fluid paste was attained. Each mixture was applied to a glass slide to be studied directly by RM.

Raman technique

A Renishaw Invia Raman microscope system fitted with a Peltier-cooled CCD detector and a Leica DMLM microscope was used. Samples were analyzed with the 514.5-nm line of an Ar laser (Laser Physics, model 235514), with an average spectral resolution of approximately 1 cm$^{-1}$ over the wavenumber range of 3400–200 cm$^{-1}$. To improve signal-to-noise ratios, spectra from 20-s exposure were averaged ($n = 10$). Spectra were recorded by placing the samples on the microscope stage and observing them with 20× and 50× objectives. Sampled areas were identified and focused using either the microscope binoculars or a video camera. Precautions were taken not to damage the samples (i.e. laser-induced degradation of paint materials). This was done by reducing laser intensity and visually confirming the absence of damage in the sample area with the help of the camera viewfinder. Thus, laser power was kept at 8 mW. Moreover, to avoid sample alteration and obtain the best sample spectrum, the laser power, the number of spectra accumulations and the irradiation exposure times were varied for the diverse samples. The egg yolk model sample was characterized by 20 Raman spectra and the tempera paint samples were characterized by ten spectra. All spectra were obtained from the same location on the model sample to avoid spatial variation.

Principal component analysis

PCA was performed on the Raman spectral data from 3100 to 2800 cm$^{-1}$ for each color group. Three data matrices were built that included the spectra of the pure egg yolk binder model samples and the spectra of the tempera model samples. In this way, the matrices were formed by 40 spectra for red tempera model samples and 50 spectra for white and blue tempera model samples. The PCs were obtained separately using both the covariance data matrices (scaling by mean-centered data) and the correlation data matrices (scaling by unit variance). The results were clearly better when PCA was performed on correlation data matrices (according to previous work) since the autoscaled
Multivariate Raman study of egg yolk-pigment interaction

Figure 1. Raman spectra of the tempera model samples in the characteristic protein region of 3150–2800 cm⁻¹ of: (a) pure egg yolk; (b) cinnabar; (c) raw Sienna; (d) lead white; (e) chalk; (f) gypsum; (g) azurite; (h) lapis lazuli and (i) smalt.

data normalize relative intensities among the samples. Indeed, this autoscaling procedure assigns the same relevance to each Raman spectral region, and, therefore, may occur to a small region – e.g. 3100–2800 cm⁻¹ – it is very effective at extracting information.[19] Therefore, the results shown and discussed here correspond to these autoscaled data.

Results and Discussion

The tempera model and pure egg yolk samples were analyzed by RM in the wavenumber range 3400–200 cm⁻¹. However, considering the purpose of this work, the protein region between 3100 and 2800 cm⁻¹ (C–H stretching modes region) was further studied to identify small changes that may occur due to the interaction with particular pigments, which otherwise show no absorption band in this spectral region.

With this aim, PCA was particularly performed in this protein Raman characteristic band and in the three color groups separately, where a priori the Raman spectra of all model samples showed an appearance similar to that seen in Fig. 1. Thus, visual inspection of the Raman C–H stretching region revealed no differences in the spectra of the samples. Nevertheless, the presence of a particular pigment could introduce slight changes in this Raman band as a result of the interaction between the protein and the specific pigment, if the pigment interacts with the amino acids involved in this specific Raman region.

From the PCA performed as described previously, the results obtained in terms of explained variance (%) and cumulative explained variance (%) for each group of studied color samples are shown in Table 1. Also, the specific Raman shift intervals analyzed in each case are pointed out.

From the results of Table 1, it is interesting to note that, in the three multivariate analyses performed, the first PC accounted for more that 90% of the variance present in the spectra (96.6% in the case of the red model samples, 93.1% in the case of the white model samples and 98.2% in the case of the blue model samples). This indicated mainly one source of variability in the spectra. Only in the case of the white model samples did the second PC account for 5.1% of the total variance. The remaining PCs explained <2% of the total variance contained in the corresponding spectral data. More detailed study of these results are discussed in the following sections.

Red tempera model samples

The PCA of the red tempera model samples allowed discrimination of the pure egg yolk samples from the red tempera samples. Figure 2(a) shows the score plot of the samples in the plane of

<table>
<thead>
<tr>
<th>Model sample</th>
<th>Raman spectral region (cm⁻¹)</th>
<th>PC</th>
<th>Variance account (%)</th>
<th>Variance accumulated (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red model samples</td>
<td>3120–2800</td>
<td>PC1</td>
<td>96.6</td>
<td>96.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PC2</td>
<td>1.9</td>
<td>98.6</td>
</tr>
<tr>
<td>White model samples</td>
<td>3050–2815</td>
<td>PC1</td>
<td>93.1</td>
<td>93.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PC2</td>
<td>5.1</td>
<td>98.3</td>
</tr>
<tr>
<td>Blue model samples</td>
<td>3070–2800</td>
<td>PC1</td>
<td>98.2</td>
<td>98.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PC2</td>
<td>1.3</td>
<td>99.6</td>
</tr>
</tbody>
</table>

PC1 and PC2. These two PCs accounted for 98.6% of the total spectral variance. As indicated above, the first PC accounted for 96.6% of the total variance; thus it can be inferred that there was a main source of variability in the original data, i.e. in the spectra. Only using this first PC the three kind of samples could be discriminated, i.e. pure egg yolk, raw Sienna tempera, and cinnabar tempera. Therefore, this variability could be associated with the different pigments present in the sample. Since the pigments show no Raman features in the spectral region here studied, it can be inferred that the pigments interact with the protein through the amino acids involved in this Raman spectral region.

The first PC separated the pure egg yolk (with positive scores) from both red tempera samples (raw Sienna-egg yolk sample showed slightly more negative scores) in the same cluster (Fig. 2(a)). The loading plot of PC1 showed a high weight in this whole band (3080 and 2830 cm\(^{-1}\)) in the component (Fig. 2(b)). As stated above, since the raw Sienna and the cinnabar tempera samples showed very different PC1 scores than those from the pure egg yolk samples, it could be concluded that these pigments interacted with the protein within this spectral region where the aliphatic and aromatic C–H\(^x\) (\(x=1,2,3\)) stretching modes of aromatic amino acids appear. In particular, Phe, tryptophan, histidine and Tyr show this band between 3060 and 3080 cm\(^{-1}\) in their spectra.\(^{20,28}\) In addition, high loading at 2830 cm\(^{-1}\) could suggest the oxidation of double bonds in the fatty acid esters by the shifting from 2940 cm\(^{-1}\).\(^{15,16}\) Furthermore, the fact that the scores of each tempera samples were placed in the same cluster – although slightly separated from each other – (Fig. 2(a)) could indicate similar interactions between the pigments and the protein, i.e. both Hg- and Fe-based pigments interact through similar amino acids in the protein chain. The second PC, which accounted for a variance <2% only, expanded the scores through this PC and so was not associated with relevant information.

**White tempera model samples**

The goal in studying the white tempera model samples was similar to that of the red tempera samples, i.e. to obtain knowledge regarding interactions between the white pigments and the egg yolk-based protein using the C–H stretching Raman region where the white pigments do not present any bands in their Raman spectra. Thus, PCA was again performed on the spectral interval 3050–2815 cm\(^{-1}\) but using the white tempera model samples and the pure egg yolk samples. The result of the projection of the original spectral data into the plane of the two firsts PCs, carrying the 98.3% of the total variance, is shown in Fig. 3(a). Examination of the score values of the samples in the first PC, accounting for 93.1% of the total variance, evidenced that lead white tempera and gypsum tempera (with positive scores) clearly differ from the pure egg yolk and calcite tempera (with negative scores; Fig. 3(a)). Thus, this first PC must be related to the interaction of the white pigment and the protein. From the distribution of the scores in the PC, it can be inferred that the presence of calcite in the model sample did not affect the Raman spectra of the protein in the C–H stretching region, whereas the presence of lead white or gypsum affected this characteristic protein Raman interval.

Regarding the loading plot of this first PC (Fig. 3(b)), the highest spectral weight was located at 3050 cm\(^{-1}\), where the amino acids Tyr, Phe and Trp show their C–H stretching mode. In addition, lower loading values were obtained between 2992 and 3010 cm\(^{-1}\) where the C–H stretching modes of threonine (3002 cm\(^{-1}\)), alanine (3005 cm\(^{-1}\)), proline (3005 cm\(^{-1}\)) and cystine (2995 and 3008 cm\(^{-1}\)) are located.\(^{20,28}\) All of this evidence suggests molecular changes in this particular spectral region when lead white and gypsum (i.e. Pb- and Ca-based pigments) were mixed with the binder (egg yolk) in the tempera samples.

The second PC was not useful despite accounting for 5.1% of the total variance. Nevertheless, a closer observation of the cluster...
of pure egg yolk and calcite tempera samples shows a slight separation between the two kinds of samples when using the two first PCs, which are not randomly mixed in the cluster. Thus, it was possible to discriminate all types of samples when using this second PC along with the first PC (Fig. 3(a)).

**Blue tempera model samples**

Application of PCA on the blue tempera model samples also used the pure egg yolk samples to obtain the spectral data matrix. The two first PCs were used to project all samples in the new variable space in order to evaluate whether differences could be identified as a result of the presence of the different blue pigments (Fig. 4(a)). The percentage of total variance explained by these two PCs was 99.6%, with 98.2% explained by the first PC. Once again, it can be inferred that the interaction of blue pigment and protein was related to the first PC. The score of the samples on this first PC showed three clusters representing pure egg yolk samples joined to lapis lazuli tempera samples (placed at highest negative score values), smalt tempera samples (placed around zero value) and azurite tempera samples (placed at highest positive score values). The loading plot of the first PC is shown in Fig. 4(b). The high loading values in this whole specific Raman region (3100–2800 cm$^{-1}$) could be associated with the large effect on this region caused by the occurrence of the azurite pigment, since tempera samples prepared with azurite were characterized by the highest positive score values on this component. Also, it suggests that the C–H stretching modes of the amino acids and polyunsaturated fatty acids esters contained in the egg yolk were affected by the contribution of the azurite, i.e. they changed as a result of the interaction with the azurite pigment. The smalt pigment also interacts with the binder, though not in a similar way as azurite does (they are clearly separated, Fig. 4(a)). On the contrary, the presence of the pigment lapis lazuli did not affect the Raman C–H stretching mode of the protein (both samples are in the same cluster); thus it can be concluded that lapis lazuli does not interact with the protein in the same way as azurite (Cu-based pigment) and smalt (Co-based pigment) do.

Although PC2 explained only 1.3% of the total spectral variance, its contribution is useful for a clear observation of the three Raman clusters cited above. The projection of the samples into the plane of PC1 and PC2 allowed a satisfactory discrimination among those three clusters. Also, pure egg yolk samples and lapis lazuli tempera samples could be discriminated in the cluster where they were placed.

**Conclusions**

In this study, an original method to gain insight into interactions between historical pigments and a proteinaceous binder (egg yolk) was successfully demonstrated. The combination of Raman spectroscopic data and a multivariate analysis of those data demonstrated the potential of the approach to discriminate samples on the basis of the interaction between the pigments and the egg yolk-based binder. The novelty of the approach lies in the performance of PCA of a particular Raman characteristic protein region where the pigments do not exhibit any Raman feature. Differences in the Raman C–H stretching region of the protein due to the presence of a specific historical pigment were highlighted by the PCA. In particular, it was possible to infer interactions on the C–H stretching modes of aromatic amino acids and fatty acids esters of egg yolk in the case of cinnabar, raw Sienna, lead white, gypsum, azurite and smalt (Hg-, Fe-, Pb-, Ca-, Cu- and Co-based pigments). By contrast, calcite and lapis lazuli did not interact with the protein through this characteristic Raman spectral region, since no spectral variations with respect to the pure protein were discriminated by the PCA. Therefore these results suggest that the pigments containing (transition or alkali-earth) metals interact with the proteinaceous binder depending on the stability of the complex formed between the metal and the binder’s amino acids. If the complex is more stable than its initial form in the pigment, protein–pigment interaction will occur. This could explain the fact that azurite (Cu) and gypsum (Ca) interact differently to lapis lazuli (Ca, Na) or calcite (Ca). It could be proposed that Cu – a transition metal – interacts in the case of azurite (carbonate-based pigment) because it forms a stable complex with the proteinaceous binder, while this is not the case for lapis lazuli since this silicate is chemically more stable. Similar arguments can be used in the case of the alkali-earth metal Ca: the Ca–protein complex is more stable than CaSO$_4$·2H$_2$O, but less stable than CaCO$_3$.

This approach will be further applied to investigate the interactions taking place after exposing similar model samples to sunlight and air pollutants, in order to gain knowledge about the behavior of historical paintings damaged by environmental exposure. Also, it would be interesting to extend the study to different pigments in order to better understand the roles of metals and anions in the interaction between a pigment and a proteinaceous binder.

**Acknowledgements**

This study was supported by Research Groups RNM325, FQM118 and RNM179 (CICE, JA, Spain). The authors acknowledge financial...
support from the Ministerio de Educación y Cultura, Dirección General de Enseñanza Superior (Spain), Project HUM2006-09262. They also thank ‘Servicios Técnicos de Investigación’ of the University of Jaén, Spain) for Raman spectroscopy experiments and A. Kowalski for the English revision. Finally, the insightful comments and suggestions of one anonymous referee are acknowledged.

References