Article

Direct Analysis of Human Hair Before and After Cosmetic Modification Using a Recent Data Fusion Method

Mônica C. Santos[a] and Fabíola M. V. Pereira [iD] *,a,b

aInstituto de Química, Universidade Estadual Paulista (Unesp), 14800-060 Araraquara-SP, Brazil

bDepartment of Chemistry, Idaho State University, 83209 Pocatello, Idaho, USA

The cosmetic modification of hair is a very common procedure used to mask or cover evidence at a crime scene. Deoxyribonucleic acid (DNA) tests are expensive and require good-quality collection of samples and a database profile. To overcome these challenges, direct analysis was performed on a large set of hair strands collected from individuals, denoted original samples, and the data were compared with those of the same samples after cosmetic modification performed by bleaching the samples in the laboratory. A total of 127 samples were evaluated in this study using two analytical techniques, wavelength-dispersive X-ray fluorescence (WDXRF) and laser-induced breakdown spectroscopy (LIBS). Instead of testing many algorithms to develop classification models for the original and bleached samples, a recent method was applied that combines information from 17 classifiers. Data fusion was also evaluated to improve the accuracy of the classification model, which was higher than 99.2%, with no requirements to select eigenvectors or thresholds.

Keywords: hair, WDXRF, LIBS, chemometrics, data fusion

Introduction

Human hair has macroscopic peculiarities, e.g., color and shape,[1,2] and the main molecular approach for hair characterization is deoxyribonucleic acid (DNA) sequencing,[3] which yields a unique fingerprint that differentiates one person from another. Thus, human hair samples have been considered to be reliable evidence at crime scenes over the last years.[4]

Cosmetic modifications are one of the most common procedures used to mask or cover evidence in a crime scene.[5,6] DNA testing is applied to identify the presence of an individual via samples of hair, nails or biological fluids. However, a DNA profiling result requires a reference or database profile, the comparison to which is sometimes inconclusive, being time-consuming, limited to hair samples with intact bulbs and may not be available in some places.[3]

To overcome these limitations, attempts to investigate the chemical composition of human hair have been described in the scientific literature.[7-10] Nevertheless, most studies have been related to the organic composition.[11] Inorganic characterization is seldom performed, and in general, it involves the destruction of the sample using

wet digestion procedures and requires specific analytical techniques with high sensitivity.[12] Two relevant points must be considered before using a destructive analysis with this type of sample: hair is an important analytical matrix with a low risk of deterioration, preserving important information for years; hair also contains historical information regarding toxicology, diseases and the health condition of individuals, in addition to the informational content of DNA.[8]

The direct analysis of this material is a challenging task, considering that the shape of samples can hinder flatness. In our previous study,[13] we obtained good performance of a noninvasive method using wavelength-dispersive X-ray fluorescence (WDXRF) and pellets of strands of human hair without using chemicals or agglutinants. Principal component analysis (PCA) combined with the WDXRF spectra revealed that the chemical elements Cu and Fe were correlated with hair samples that had been straightened or dyed or with a single strand that had undergone both cosmetic procedures.

In our current study, the investigation of human hair was extended to detect other elements related to the chemical composition, such as C, H, N, Na and O, by using laser-induced breakdown spectroscopy (LIBS) on the same pellets used previously.[13] In the scientific literature,[14] it has been reported that the elemental analysis of hair

*e-mail: fabiola.verbi@unesp.br

reveals approximately 50 wt.% carbon, 7 wt.% hydrogen, 22 wt.% oxygen, 16 wt.% nitrogen, and 5 wt.% sulfur, with a slight variation in the percentage values depending on the source of the hair.

In this study, the ability to identify strands of human hair, original and bleached, from the same person was shown. The strategy involved collecting a wider variety of samples. The hair samples were cosmetically modified at the laboratory. A recent chemometric method developed by Brownfield *et al.*[15] for classifying these samples without further preprocessing that allowed the easy fusion of the data from LIBS and WDXRF was performed. The fusion of instrumental data has been shown to provide complementarity and improvements in analytical information for many purposes.[16-18]

In our previous study,[13] light elements such as C, H, O and Na could not be detected by the WDXRF instrument, but the influence of K on the bleached samples was confirmed. Another noteworthy finding was the influence of Ca, Fe and S on the ability to differentiate the samples before the bleaching procedure. The scattering of a Rh source was used as analytical information for evaluating the data to compensate for matrix effects from the light elements. Furthermore, there is a lack of information regarding other chemical elements that constitute hair.[14] The use of LIBS for the tests is complementary in the investigation of hair and maintains the integrity of the samples,[19] which, as mentioned above, is a relevant issue.

Thus, this study focuses on classifying hair samples after cosmetic modification using the fusion of LIBS and WDXRF data and a recent method of classification.

Experimental

Samples

Sixty-four human hair samples provided by 63 different donors composed the dataset. The cutting procedure was performed according to the instructions of the Society of Hair Testing (SoHT)[20] and The Faculty of Forensic & Legal Medicine of the Royal College of Physicians,[21] which recommend cutting close to the scalp. The 1 cm long strands of hair were stored in decontaminated plastic flasks at room temperature. A decontamination procedure was performed to avoid any external contaminating sources. The decontamination procedure consisted of three steps: rinsing with analytical grade acetone (Commercial Neon, Suzano, SP, Brazil), followed by a Milli-Q™ (Merck, Darmstadt, Germany) water rinse and another acetone rinse. The strands were divided into two samples from each donor, with one kept intact and denoted the original samples and the other subjected to a cosmetic procedure of bleaching and denoted the bleached samples, for a total of 127 samples. The only exception was for two samples with the same features that came from the same person. In this case, the dyeing procedure was the same using henna product from same manufacturer but differed in color; therefore, there was no need to perform decolorizing on both strands of hair. Each sample of 150 mg in mass was then shaped into a pellet using only a manual hydraulic press with no chemicals or agglutinant,[13] as shown in Figure S1 in the Supplementary Information (SI). Figure 1 summarizes the main features of the human hair samples investigated in this study. A hair decolorizing kit (Ivel Indústria de Perfumes e Cosméticos Ltda, Nova Iguaçu, RJ, Brazil) was used for the bleaching procedure.

WDXRF data

The 127 pellets of hair were previously measured before LIBS tests using an ARL Perform'X WDXRF spectrometer from Thermo Scientific (Madison, WI, USA) equipped with a rhodium X-ray tube. The following conditions were used for the bulk measurements of the pellets under a vacuum environment with a 10-mm collimator beam
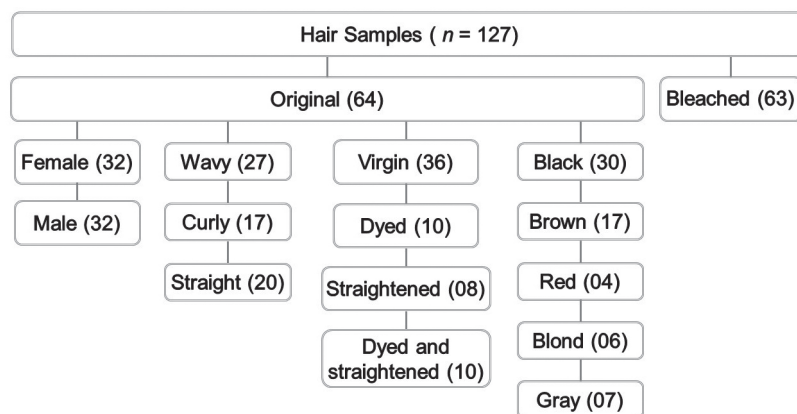

Figure 1. Diagrammatic representation of the assessed hair samples.

using two analyzer crystals: LiF 200 (K to U, 0.17-3.88 Å, 0.01 resolution) and AX03 (O to Si, 3.99-6.89 Å, 0.01 resolution); approximately 15 min *per* sample was used. Three inter-day readings of each sample were performed. The National Institute of Standards and Technology (NIST) database[22] was also accessed to confirm the WDXRF emission lines.

LIBS data

For the LIBS measurements, the same 127 pellets of hair were mapped using a J200 LIBS system from Applied Spectra (Fremont, CA, USA) equipped with a Q-switched Nd:YAG laser (1064 nm) and a 6-channel charge-coupled device (CCD) spectrometer that recorded spectral information from 186 to 1042 nm and generated nanosecond pulses up to 100 mJ. The laser pulse energy applied for the experiments was 80 mJ, with a 0.5-μs delay time and a 125-μm spot size. The estimated fluence of the laser was 650 J cm^{-2}. A total of 924 spectra were acquired *per* sample from a mapped area within 11 lines with 4-mm spacing; approximately 40 laser pulses were recorded for each line measured in the front and back of the pellets. The emission lines were extracted from TruLIBS™ database software (Applied Spectra) and by consulting the database from the NIST.[22]

Chemometric evaluation

For all of the data, a lab-made computer code[15] for Matlab R2017a[23] was applied to calculate the fusion of 17 classifiers, as follow: (*i*) partial least squares for discriminant analysis (PLS2-DA); (*ii*) k-nearest neighbor (kNN); (*iii*) the Mahalanobis distance (MD); (*iv*) sinθ; (*v*) Q-residual (Q res); (*vi*) divergence criterion (DC); (*vii*) cosθ; (*viii*) the Euclidean distance; (*ix*) determinant; (*x*) inner product correlation; (*xi*) unconstrained Procrustes analysis (PA); (*xii*) and (*xiii*) constrained PA (for 2 classifiers), and (*xiv*)-(*xvii*) the extended inverted signal correction difference (EISCD) (for 4 classifiers).

Results and Discussion

From the results of our previous study,[13] the dataset was evaluated using eight normalization modes,[24] and the Euclidean norm was considered the best for data discrimination, while first principal component (PC1) had the highest explained variance. For this study, the most important variables were selected by considering WDXRF emission lines with loading values above 0 in PC1 as relevant. As a result, the wavelengths of Fe Kα,

Ca Kα and Kβ, K Kα, Rh Kα and S Kβ, comprised 183 variables as shown in Figure S2 (SI section). The PCA calculated for the same samples with the 183 selected variables confirmed the same clustering tendency reported in our previous paper[13] (see Figures S3a and S3b, SI section). Note that one original sample was found to be part of the bleached cluster. The issue with this sample was that compared to the other samples, the hair was gray in color with a different texture.

Direct analysis using a LIBS system was performed on each face of the hair strand pellets. These measurements were conducted using this procedure because LIBS measures the sample point by point, and a bulk analysis cannot be performed, as is the case with WDXRF. An initial data inspection was performed to detect anomalous spectra, which can occur when the laser bores through a sample. Several parameters were evaluated for each spectrum: the standard deviation, maximum, summation and Euclidean norm.[24] To identify differences between the front and back of the pellets, PCA was carried out (Figures S4a and S4b, SI section) and revealed that there was no tendency for discrimination between the two faces.

Since the signal-to-noise ratio can be influenced by sample heterogeneity and signal fluctuations, the raw data were also assessed using 12 normalization modes.[24] Among these 12 modes, the mean was best able to improve this ratio. The data were tested with PCA, and the data separation mode leading to a better separation in PC1 was chosen as the best one. These were the same criteria applied to the WDXRF data to choose the best standardization mode.

Next, considering the high number of variables (12,288), the selection of the most important variables was performed using PCA. The criterion for excluding non relevant emission lines was variables with loading values of 0 along PC1. Figure 2 shows an average spectrum calculated for all the hair samples after choosing 422 variables.

The results using a PCA calculated with the 127 samples and these 422 variables normalized according to the mean led to a tendency to discriminate between the original and bleached samples, as shown by the scores in Figure 3a and the variables responsible for clustering in Figure 3b. Nevertheless, the overlapping of 2 samples, one original and the other bleached, from different donors was observed, as shown in Figure 3a. These samples were different samples from those that overlapped in the WDXRF data.

Human hair is composed of ca. 65 to 95% amino acids, such as glycine, alanine, and arginine. These amino acid components contain organic elements, including C, H and N, in their structures. The detection of minerals, such as Ca and Mg, can be associated with the side chains of proteins
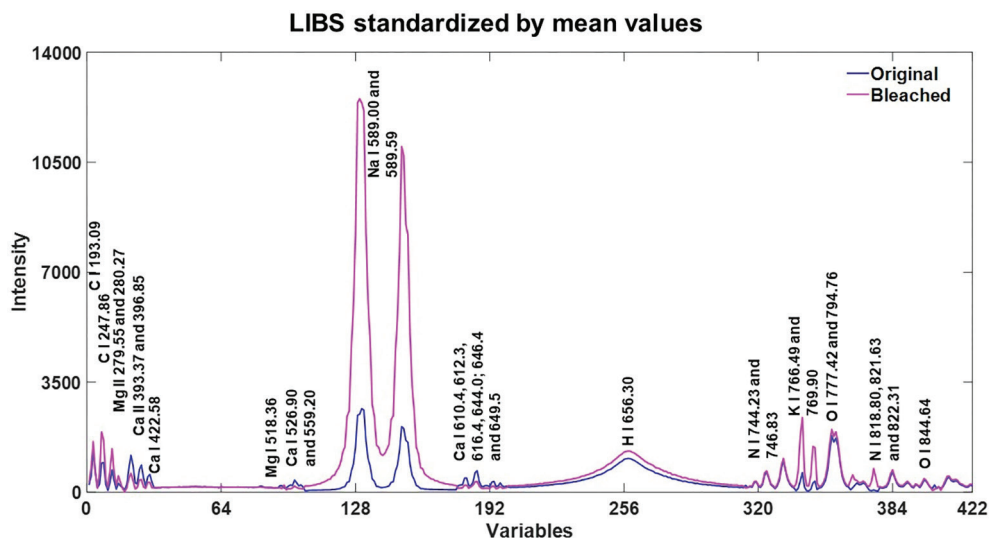
Figure 2. Averaged profiles after the selection of variables of emission lines from the pellets of original (blue line) and bleached (pink line) hair samples obtained using a LIBS system.
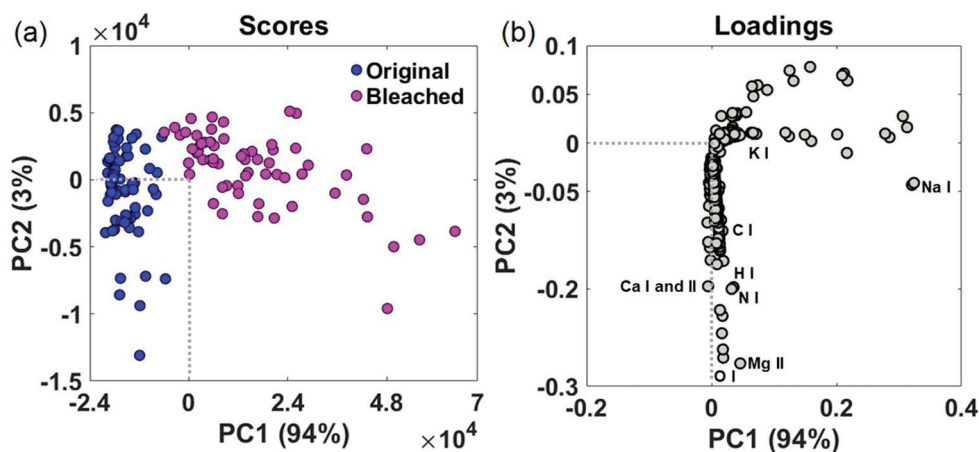

Figure 3. PCA of the LIBS data performed with 422 variables representing (a) scores and (b) loadings for PC1 and PC2.

or with the fatty acid groups of lipids.[10] The mineral content of dried hair samples varies from 0.25 to 0.95% (m/m).[9]

Note that the elements K and Na were correlated with the bleached strands, as verified by the loadings displayed in Figure 3b. The signal ratios between the bleached and original samples for K and Na were 39 and 23, respectively. These two elements are related to the residues of products used to decolorize hair during the bleaching process. Additionally, hair bleach with hydrogen peroxide (H_2O_2) as the oxidizing agent is sold as a kit and can contain mixtures of inorganic compounds, such as persulfates of potassium ($K_2S_2O_8$) and ammonium ((NH_4)$_2S_2O_8$), sodium lauryl sulfate ($CH_3(CH_2)_{10}CH_2(OCH_2CH_2)_nOSO_3Na$), sodium silicate ($Na_2SiO_3$) and ethylenediaminetetraacetic acid disodium salt ($C_{10}H_{14}N_2Na_2O_8$).

In this study, a recent chemometric method for the classification of samples was applied.[15,25,26] Instead of using an algorithm that tests one by one, this method allows the

combination of information from 17 classifiers with no requirements to select a threshold, eigenvector or number of neighbors. In addition, it is possible to easily fuse the data from different instruments.

Basically, a sum value resulting from the fusion of 17 algorithms was used to classify one sample in a specific class, as shown in the diagrammatic representation of Figure 4. First, the calculations were performed for each classifier. Next, each row value was normalized to a unit length that eliminated magnitude differences among the values resulting from each classifier. A large window of multiple values, denoted the tuning parameter window, was used to stack the classifiers in blocks for the eigenvector-based algorithms, e.g., PLS2-DA, kNN, MD, sinθ, Q res and DC, as shown in Figure 4. For each of these classifiers, a tuning parameter window of 61 was applied. For the remaining 11 classifiers, only one value was generated for each. It means for the first six classifiers that is a

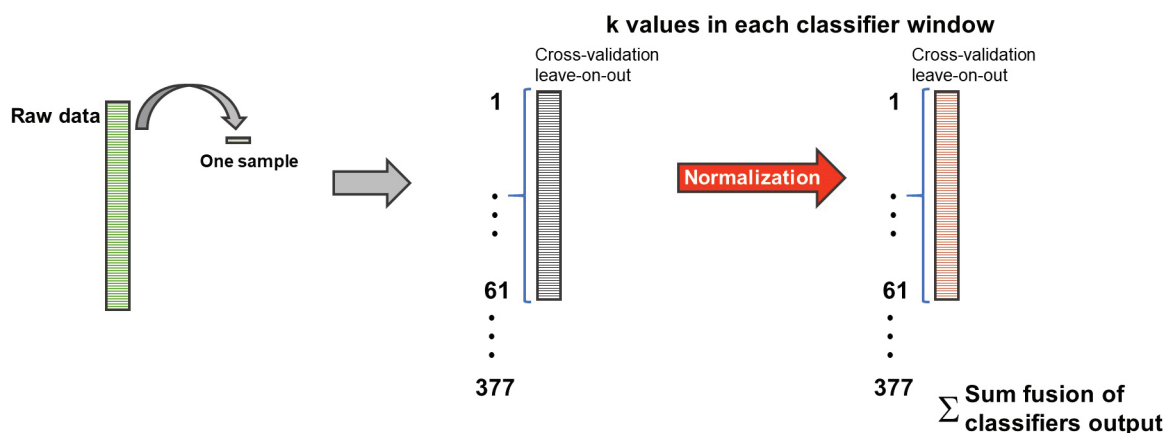## Calculations for one sample in a class

Figure 4. Diagrammatic representation of the sum fusion classification.

window with 61 output values *per* each (367 output values total, from 1 to 366) and for remaining classifiers, eleven individual output values were generated (from 367 to 377).

As shown in Figure 5, for either the original (range 1) or bleached (range 2) sample data, 366 row values for the eigenvector-based algorithms and 11 more for the other classifiers were generated for each instrument, achieving 377 row values. Then, one sum was obtained for each range. A cross-validation leave-one-out was calculated until all the samples had been processed. The smallest sum determined whether the sample belonged to one range or the other (original or bleached). In the example depicted in Figure 5, the sample tested is an original one.
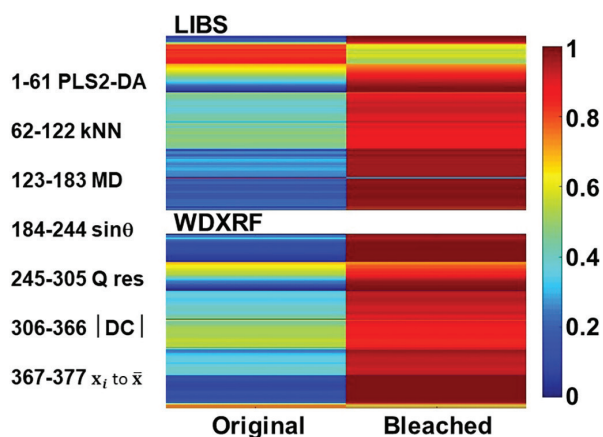

Figure 5. Representation of the classifiers built into the fusion input matrix from two instruments (LIBS and WDXRF) for classifying a range 1 sample (original) into one of 2 ranges.

The ability to classify samples was evaluated using figures of merit such as accuracy, specificity, and sensitivity. These merits were computed according to the occurrence of true positive, false positive, true negative, and false negative events. The advantage of simplicity was verified

for this fusion method since there was no need for training or validation sets due to the large tuning window and the cross-validation process. Further preprocessing of the data was also not needed. Figure 5 shows how the sum fusion performed when the calculations were carried out for instrument 1 (LIBS) with dataset 1 and subsequently for instrument 2 (WDXRF) with the same data, and rows with the raw values for the 17 classifiers were thus generated for these samples. In the same fashion, calculations for dataset 2 were computed, and other rows were formed. Note that for the fusion classification, each row was normalized to a unit length. As is clear from Figure 5, the smaller values were verified for the original samples, and the column sums of the original data were 151.1 and 124.6, respectively, for LIBS and WDXRF. The bleached sample data had values of 331.0 (LIBS) and 345.8 (WDXRF). Furthermore, the total value of the sum of the column for the original samples was 275.7, and for the bleached samples was 676.8. Consequently, the samples belonged to the original range.

As mentioned earlier, either technique showed promising results, as indicated by the plots for LIBS (Figure S5, SI section) and WDXRF (Figure S6, SI section). A good accuracy of 99.2% was achieved with separate calculations for the sum fusion classification. The data studied here pose a binary classification problem: the accuracy was the same for each class, while the sensitivity and specificity were reversed in value between the two levels. The LIBS data had greater consistency along the 61 tuning parameter windows (Figure S5, SI section), and WDXRF still had 3 gaps before stabilizing at the maximum value of 99.2% (Figure S6, SI section). The information from the two datasets was complementary considering that the chemical information was evident after the fusion. Thus, to take advantage of the easy combination of the instrumental data, as shown in Figure 5, and to verify the
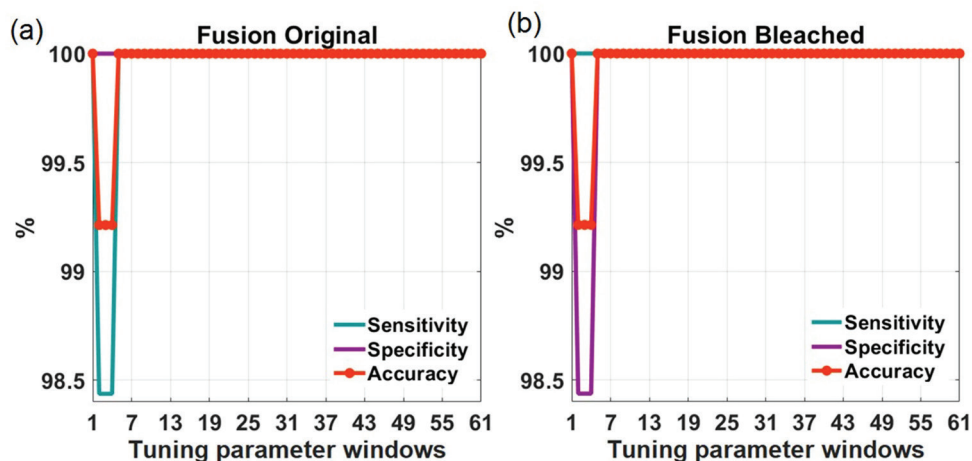
Figure 6. Classification results and fusion of all 17 classifiers using the fusion input matrix from two instruments (LIBS and WDXRF) for (a) range 1 and (b) range 2.

performance of these data in a robust fashion, it is possible to verify that the fusion of the data improved the figures of merit for the classification. As shown in Figure 6, the interval for accuracy was between 99.2 and 100% for both of them, in comparison of 94.5-99.2% for WDXRF (Figure S6, SI section) and 99.2% for LIBS (Figure S5, SI section) in both ranges. Thus, the fusion of data is a feasible process with the method used in this study, considering the increase in accuracy.

Conclusions

Analytical data obtained from WDXRF and LIBS can be used to characterize human hair samples. However, there is a lack of organic detection capability with WDXRF. LIBS is able to provide a wide array of analytical information on inorganic and organic constituents. The sum fusion classification revealed that using both sources of data (WDXRF and LIBS) were achieved high indexes for accuracy and specificity and sensitivity above 99.2%, mainly for the cases where a clear discrimination between samples cannot be visualized. The analytical method presented here opens the possibility of a forensic application in addition to routine analysis.

Supplementary Information

Supplementary data are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

References

1. Birngruber, C.; Ramsthaler, F.; Verhoff, M. A.; *Forensic Sci. Int.* **2009**, *185*, e19.

2. Yu, Y.; Yang, W.; Wang, B.; Meyers, M. A.; *Mater. Sci. Eng., C* **2017**, *73*, 152.

3. Morrison, J.; Watts, G.; Hobbs, G.; Dawnay, N.; *Forensic Sci. Int.* **2018**, *285*, 147.

4. Goodwin, W. H.; *Forensic Sci. Int.* **2017**, *278*, 221.

5. Kurouski, D.; Van Duyne, R. P.; *Anal. Chem.* **2015**, *87*, 2901.

6. Boll, M. S.; Doty, K. C.; Wickenheiser, R.; Lednev, I. K.; *Forensic Chem.* **2017**, *6*, 1.

7. Cooper, G. A. A.; Kronstrand, R.; Kintz, P.; *Forensic Sci. Int.* **2012**, *218*, 20.

8. Kempson, I. M.; Lombi, E.; *Chem. Soc. Rev.* **2011**, *40*, 3915.

9. Harkey, M. R.; *Forensic Sci. Int.* **1993**, *63*, 9.

10. Robbins, C. R.; *Chemical and Physical Behavior of Human Hair*; Springer: New York, USA, 1998, ch. 2.

11. Wang, Q.; Zhong, H.; Zheng, Y.; Zhang, S.; Liu, X.; Zhang, X.; Zhang, X.; Zhang, Z.; *Analyst* **2014**, *139*, 5678.

12. Nomura, C. S.; Oliveira, P. V.; *Anal. Methods* **2010**, *2*, 49.

13. Santos, M. C.; Sperança, M. A.; Pereira, F. M. V.; *X-Ray Spectrom.* **2018**, *47*, 252.

14. Popescu, C.; Höcker, H.; *Chem. Soc. Rev.* **2007**, *36*, 1282.

15. Brownfield, B.; Lemos, T.; Kalivas, J. H.; *Anal. Chem.* **2018**, *90*, 4429.

16. Geurts, B. P.; Engel, J.; Rafii, B.; Blanchet, L.; Suppers, A.; Szymańska, E.; Jansen, J. J.; Buydens, L. M. C.; *Chemom. Intell. Lab. Syst.* **2016**, *156*, 231.

17. Khajehzadeh, N.; Haavisto, O.; Koresaar, L.; *Miner. Eng.* **2017**, *113*, 83.

18. Bevilacqua, M.; Bro, R.; Marini, F.; Rinnan, A.; Rasmussen, M. A.; Skov, T.; *TrAC, Trends Anal. Chem.* **2017**, *96*, 42.

19. Emara, E. M.; Imam, H.; Hassan, M. A.; Elnaby, S. H.; *Talanta* **2013**, *117*, 176.

20. Society of Hair Testing; *Forensic Sci. Int.* **2004**, *145*, 83.

21. https://fflm.ac.uk/publications/recommendations-collecting-hair-samples-for-toxicology, accessed in May 2019.

22. National Institute of Standards and Technology (NIST); *Atomic Spectra Database*, version 4, online; NIST, Gaithersburg, MD, USA, 2010. Available at http://physics.nist.gov/asd3, accessed in May 2019.

23. *Matlab R2017a*; MathWorks, Natick, MA, USA, 2017.

24. Castro, J. P.; Pereira-Filho, E. R.; *J. Anal. At. Spectrom.* **2016**, *31*, 2005.

25. Kumar, R.; Vishal, S.; *TrAC, Trends Anal. Chem.* **2018**, *105*, 191.

26. Panchuk, V.; Yaroshenko, I.; Legin, A.; Semenov, V.; Kirsanov, D.; *Anal. Chim. Acta* **2018**, *1040*, 19.