Journal of Microscopy, Vol. 215, Pt 1 July 2004, pp. 13–23
Received 26 March 2003; accepted 5 January 2004

Quantitative analysis and classification of AFM images of human hair

S. P. GURDEN, V. F. MONTEIRO*, E. LONGO* &
M. M. C. FERREIRA
Department of Chemistry, State University of Campinas (UNICAMP), 13084-862, Campinas – SP, Brazil
**Interdisciplinary Laboratory of Electrochemistry and Ceramics, Federal University of São Carlos (UFSCar), 13565-905, São Carlos – SP, Brazil*

Key words. Atomic force microscopy, classification, cuticle, discriminant analysis, human hair, partial least squares.

Summary

The surface topography of human hair, as defined by the outer layer of cellular sheets, termed cuticles, largely determines the cosmetic properties of the hair. The condition of the cuticles is of great cosmetic importance, but also has the potential to aid diagnosis in the medical and forensic sciences. Atomic force microscopy (AFM) has been demonstrated to offer unique advantages for analysis of the hair surface, mainly due to the high image resolution and the ease of sample preparation. This article presents an algorithm for the automatic analysis of AFM images of human hair. The cuticular structure is characterized using a series of descriptors, such as step height, tilt angle and cuticle density, allowing quantitative analysis and comparison of different images. The usefulness of this approach is demonstrated by a classification study. Thirty-eight AFM images were measured, consisting of hair samples from (a) untreated and bleached hair samples, and (b) the root and distal ends of the hair fibre. The multivariate classification technique partial least squares discriminant analysis is used to test the ability of the algorithm to characterize the images according to the properties of the hair samples. Most of the images (86%) were found to be classified correctly.

Introduction

The analysis of the surface architecture of human hair has received a large amount of interest within various fields of science, including biology, forensics, cosmetics, medicine and dermatology (Orfanos & Happle, 1990; Robbins, 1994; Jollès *et al.*, 1997). The central cortex of a hair fibre is surrounded by thin cellular sheets, known as cuticles, which overlap each other from root to tip. Figure 1 gives a schematic representation. These cuticles fulfil a number of useful roles, including protection from physical and chemical insult and a tendency to maintain the hair in a clean and disentangled state (Swift, 1999). At the root-end of the hair there are approximately ten layers of cuticle, but physicochemical stresses, such as combing and sunlight, and cosmetic treatments, such as permanent waving and dying, can lead to the cuticle edge being gradually chipped away, revealing the top surface of the cuticle lying directly below. In some cases, the layer of cuticular sheets is removed entirely, leading to the exposure of the underlying central cortex of the hair at the distal end (Garcia *et al.*, 1978). The surface architecture of a hair fibre is responsible for its visual and tactile properties, and thus the effect of cosmetic treatments on the cuticle has been of great interest within the cosmetics industry (Robbins, 1994). The condition of the hair cuticles has the potential to assist in diagnosis of health disorders (Srivastava & Gupta, 1994; Pötsch, 1996; Pozebon *et al.*, 1999) and can also be used forensically to provide information on the identity and lifestyle of the hair's owner (Swift & Brown, 1972; Hoffman, 1991).

A number of techniques have been used to study the hair cuticle, including scanning electron microscopy (SEM) and transmission electron microscopy (TEM) (Swift & Brown, 1972; Swift & Bews, 1976; Hess *et al.*, 1990; Ruetsch *et al.*, 2000; Ahn & Lee, 2002), X-ray photoelectron spectroscopy (Goddard & Harris, 1987), confocal microscopy (Hadjur *et al.*, 2002), microdiffraction (Kreplak *et al.*, 2001), secondary ion mass spectrometry (Gillen *et al.*, 1999), goniometry (Feughelman & Willis, 2001) and lateral force microscopy (LFM) (McMullen *et al.*, 2000; McMullen & Kelty, 2001). One other technique that has been of particular interest recently is atomic force microscopy (AFM) (Binnig *et al.*, 1986; Rugar & Hansma, 1990). AFM offers a uniquely useful combination of advantages for analysis of the fine surface structure of the outer surface of hair (O'Conner *et al.*, 1995; Smith, 1997; You & Yu, 1997; Swift & Smith, 2000; Blach *et al.*, 2001). AFM is a non-invasive technique that requires no special sample preparation. By contrast, electron

Correspondence: Professor Márcia M. C. Ferreira. E-mail: marcia@iqm.unicamp.br

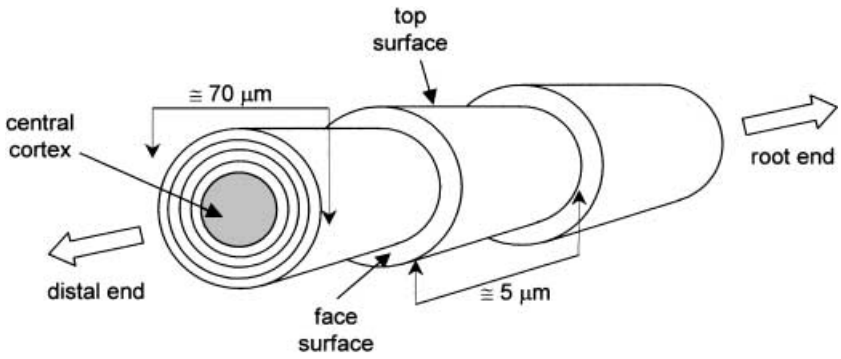

Fig. 1. Schematic representation of the overlapping cuticles which surround the central cortex.

microscopy techniques such as SEM and TEM require extensive sample preparation and high-vacuum conditions, which may introduce undesired artefacts to the fine detail of the hair surface. In comparison with optical microscopy, AFM offers ultrahigh resolution, which is essential in detecting fine surface properties of human hair. A major advantage of AFM is the ability to provide accurate topographic information about the sample surface, enabling height differences of less than 1 nm to be measured. AFM is, however, limited to measurement of the topographic morphology perpendicular to the sample plane, meaning that re-entrant surfaces (i.e. spaces obscured by the main surface) and subsurface information cannot be detected, in contrast to SEM or confocal microscopy using fluorescence.

One of the main problems with the application of imaging techniques to the analysis of physical properties of human hair is the difficulty of representative sampling. Head hairs, for example, can differ greatly depending upon the region of the head from which they originate. Variation is caused both by differences in the hair follicles that produce the hairs and differences in the environmental stresses (e.g. exposure to sunlight) and personal grooming patterns (e.g. brushing) to which the hairs are exposed from day to day. These latter factors are also reflected in large differences in cuticular structure according to the distance from the root and, thus, the age of that part of the hair fibre. Finally, biorhythmic variations also contribute to variation along the hair fibre (Sauermann *et al.*, 1988). This means that the use of imaging to perform a before-and-after study to determine the effect of a specific hair treatment should be carried out with great caution: (a) taking care to compare hair samples from the same regions and (b) by measuring as many images as is necessary to allow a statistically meaningful comparison.

Despite the problem of representative sampling, some quantitative studies have been reported in the literature. Sauermann *et al.* (1988) used surface profilometry to measure roughness parameters and cuticle density, and used these to compare the effect of cosmetic treatments. O'Conner *et al.* (1995) used AFM to monitor the change in step height (the height of one cuticle relative to the adjacent one) under different hydration and pH conditions. You & Yu (1997) also used AFM to measure cuticle step height, using this parameter to quantify the effects of pH and heating on human hair. Smith (1998) developed a method for the rapid calculation of cuticle step height based upon taking the first-order derivative of the surface profiles in order to locate the position of the cuticle steps. McMullen & Kelty (2001) used AFM/LFM to measure friction coefficients capable of quantifying the effect of surface treatments such as bleaching and polymer coating. Feughelman & Willis (2001) used goniometry to quantify the effect of hair treatment on lustre by measuring the angle the cuticle top surface makes with the fibre axis.

This article aims to build upon some of the above work by presenting a computation algorithm capable of analysing an AFM image and calculating a set of parameters that describe as fully as possible the cuticular structure. It will be shown that these descriptors can be used for automatic classification of hair samples according to factors such as distance from the root-end and hair treatment.

The rest of the article is arranged as follows. First, the experimental aspects of the AFM hair measurements are described. Next, the computational analysis algorithm is described, with a full step-by-step description being given in an Appendix. The analysis method is applied to a set of images of human hairs with different properties (root end/distal end, untreated/treated). Finally, in order to demonstrate the usefulness of a multivariate characterization of the cuticle structure, the multivariate classification technique partial least squares discriminant analysis (PLS-DA) is applied to the resulting set of cuticle descriptors and used to classify the hair images.

Materials and methods

The hair used in this study was black, Caucasian human hair obtained from De Meo Brothers, New York, U.S.A. Each hair fibre had a thickness of approximately 65 μm and a length of approximately 20 cm. Hair tresses weighing approximately 1.0 g were used during the preparation steps.

Two types of hair samples were prepared: (a) untreated and (b) bleached. For the untreated samples, the hairs were first washed using 10% LESS (lauryl ether sodium sulphate) and then rinsed using distilled water. The hairs were then dried at

room temperature. For the bleached samples, a bleach solution was prepared by mixing a 6% hydrogen peroxide (H_2O_2) solution with a concentrated ammonium hydroxide solution and ammonium persulphate powder in a 2 : 1 : 0.5 weight ratio. The hair tresses were placed in the solution for 30 min and afterwards washed and dried using the same procedure as for the untreated samples.

The AFM images were measured using a Digital Instruments NanoScope IIIa instrument under atmospheric conditions at 25 °C using a loading force of 3.6 nN. Images were obtained in the contact mode using a silicon nitride pyramidal tip mounted on a 200-μm-long cantilever with a force constant of 0.06 N m^{-1}. Sections of single hair fibres of approximately 0.7 cm length were mounted on the sample holder using double-sided tape. The hairs were orientated horizontally with the root end to the right and the distal end to the left. The images were 30×30 μm (256×256 pixels) in size and were centred close to the central axis of the hair fibre so as to minimize the distortion in relative surface height found at the edges of the fibre where extreme curvature is present (Smith, 1998).

For both the untreated and the bleached tresses, hair sections were taken from two regions: (a) 0.5 cm from the root end and (b) 0.5 cm from the distal end. A total of 38 AFM images were available for analysis, as summarized in Table 1. All images were available as NanoScope version 4.43 files and these were read into the MATLAB mathematical computing environment (The MathWorks, Natick, MA, U.S.A.), where all subsequent image manipulation and analysis was performed.

Planification

Prior to further analysis, a planification step was performed, in which the background surface of the image is subtracted. Due to the cylindrical shape of the hair fibre, this background subtract was carried out by fitting a two-variable polynomial surface using a first-order polynomial in the *x*-axis direction and a third-order polynomial in the *y*-axis direction. An example surface is shown in Fig. 2. Note that in terms of the analysis algorithm described in this article, this planification step has the effect of correcting for lateral tilt of the hair fibre along the central axis, but does not correct for the influence of curvature towards the edges of the hair fibre. However, this latter effect was found to be minimal for the images analysed here.

Figure 3 shows two typical images: (a) an untreated fibre imaged 0.5 cm from the root end and (b) a bleached fibre imaged 0.5 cm from the distal end. The differences in the cuticular structure are clear. Whereas Fig. 3(a) exhibits the typical overlapping pattern of cuticles, Fig. 3(b) shows cuticles in a highly degraded state and it appears that the cuticular layer has been completely removed in the bottom half of the picture. One aim of the work presented in this article is to develop an algorithm capable of automatically recognizing and quantifying differences, due to both artificial and natural sources, between hair images such as these.

Table 1. AFM images available for analysis.

Treatment	Sampling position	Number of images
None	Root end	6
None	Distal end	14
Bleaching	Root end	10
Bleaching	Distal end	8

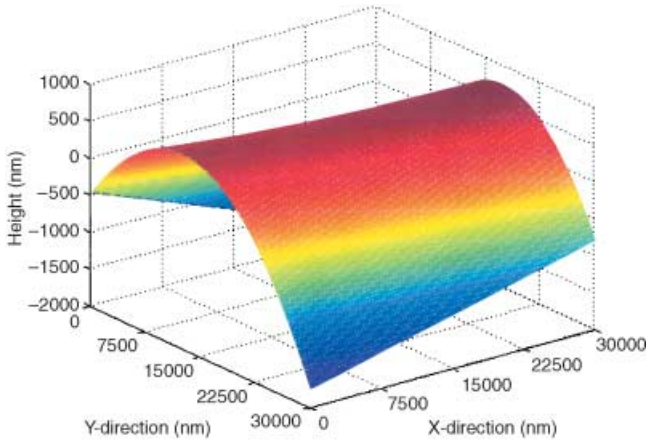

Fig. 2. Example of a background surface calculated during the planification step.

An algorithm for the quantitative analysis of cuticular structure

Quantitative analysis and classification of AFM hair images is carried out by characterizing the hair surface using descriptors that summarize the important characteristics of the cuticular structure. Some cuticular descriptors, such as the angle the cuticular sheet makes with the central axis or the cuticle density, are known to be related to overall properties such as glossiness or strength and, as such, can be used as indicators of hair quality. However, it is unlikely that one parameter alone is sufficient fully to describe differences in hair samples with different properties, and so a multivariate approach using an array of cuticular descriptors could be highly advantageous.

Some imaging instrument software packages allow the user to manipulate and examine measured images manually, enabling the calculation of certain cuticular descriptors such as step height and tilt angle for a specific cuticle at a specific location. However, for images in which a number of overlapping cuticles are shown, it is practically impossible to use a manual approach to produce reliable estimates of cuticle parameters due to the wide distribution of values, even within a given image, for a particular parameter. In this section, an algorithm capable of automatically scanning an AFM image and calculating descriptors at all possible locations is described. The overall cuticular structure is then characterized by a mean value and a measure of its statistical distribution (i.e. standard deviation) for each descriptor.

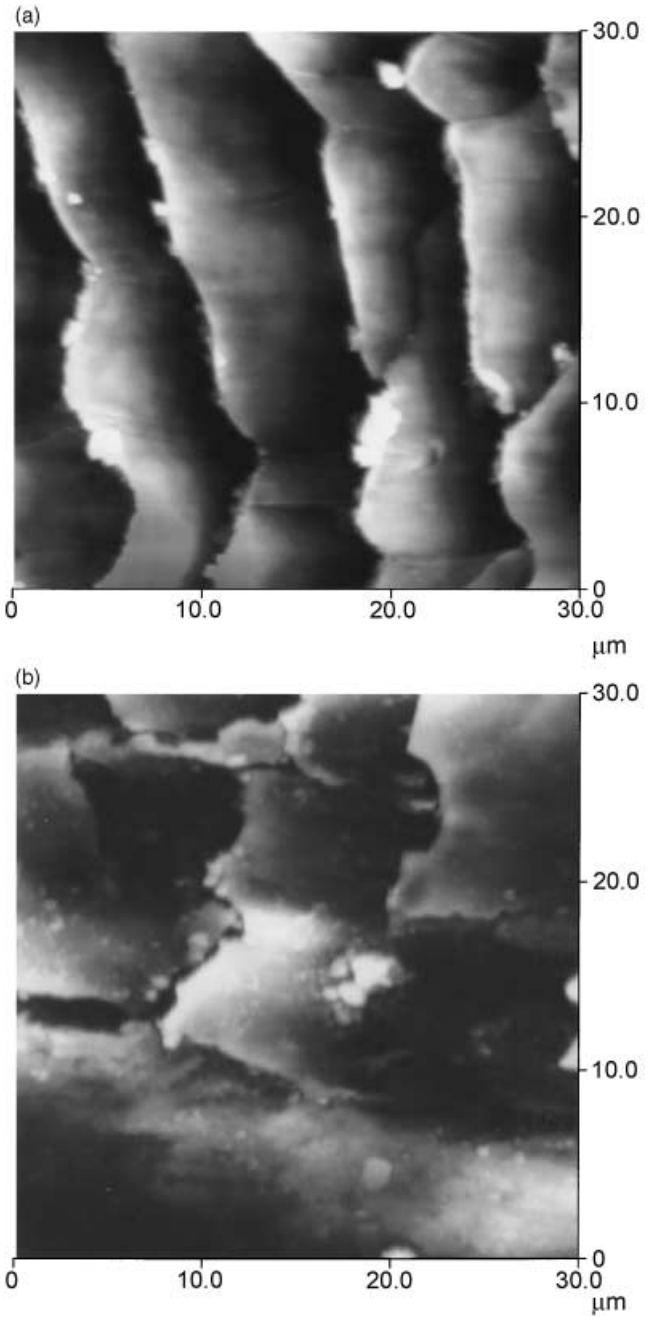

Fig. 3. AFM images of (a) an untreated Caucasian hair taken near the root end and (b) a bleached Caucasian hair taken near the distal end.

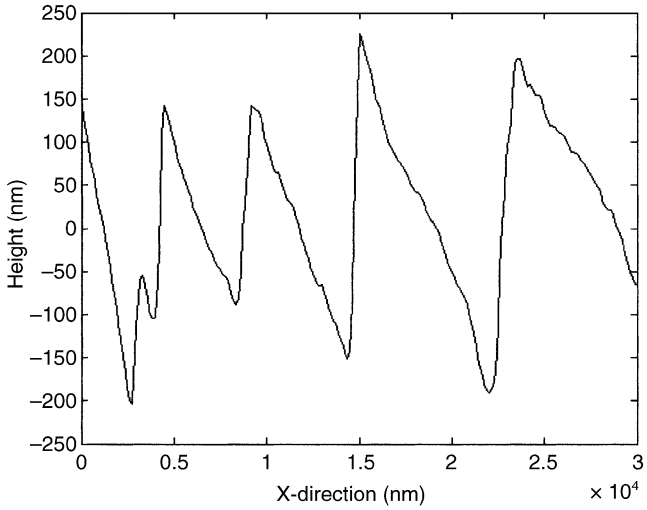

Fig. 4. Surface profile taken from one row of Fig. 3(a).

The algorithm works on the principle of surface profilometry. Changes in the height of the hair surface taken along the fibre axis, as described by the horizontal lines of an AFM image, are examined. An example of a surface profile is given in Fig. 4, in which the outline of the overlapping cuticles is clearly seen, with the steepest regions being at the face surfaces of the cuticles, i.e. where one cuticle ends and the cuticle beneath emerges. Smith (1998) developed a method for the rapid calculation of cuticle step height based upon using the first-order derivative of the surface profiles to locate the position of the cuticle steps. The algorithm presented in this article builds on that of Smith by (a) calculating a whole series of descriptors in order to characterize the cuticular structure fully, and (b) using an algorithm that is more robust against the false detection of cuticle steps due to irregularities such as score marks on the fibre surface.

The basis of the algorithm is the initial location of the face surfaces of the cuticles and the subsequent location of the low point, corresponding to the position at which one cuticle ends and the one below emerges (see point A in Fig. 5), and the high point, corresponding to the tops of the cuticles (point B in Fig. 5). Having located these points for each cuticle edge, a whole series of parameters may be derived, some of which are shown in Fig. 5. Although the algorithm is not particularly complex, it consists of a number of separate steps designed to ensure the correct location of the important profile points while avoiding fitting surface irregularities. Savitsky–Golay smoothing and derivate filters (Savitzky & Golay, 1964) are applied as one measure against fitting low-level variation in the surface profile. The algorithm is described fully in the Appendix.

The full array of cuticular descriptors calculated by the algorithm (with references to Fig. 5) is now given. Note that these descriptors are independent of image size and resolution.

1 *Step height*: the vertical distance between the top of the cuticle (point B) and the point where the top surface of the underlying cuticle emerges (point A).
2 *Tilt*: the angle the top surface of the cuticle makes with the axis of the fibre.
3 *Backtilt*: the angle the face surface of the cuticle makes with the axis of the fibre.

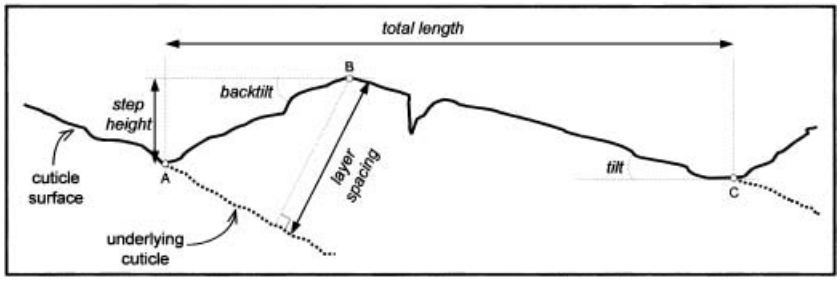

Fig. 5. A schematic representation of a surface profile showing some selected descriptors used to characterize the cuticular structure. The full list of descriptors is given in the text.

4 *Layer spacing*: the minimum distance between the top of the cuticle and the cuticle underlying it, measured at the cuticle end. Note that the cuticles of damaged and/or hydrated hair may become unattached at the end, allowing a layer of air or water to be present between cuticle layers. In this case, layer spacing should be interpreted as the distance between subsequent layers at the cuticle end, rather than the thickness of the cuticle as such.
5 *Face distance*: the distance between the top of the cuticle and the point where the top surface of the underlying cuticle emerges (i.e. distance A–B).
6 *Top distance*: the distance between the top of the cuticle and the point where the face-end of the overlying cuticle begins (i.e. distance B–C).
7 *Fit error*: a measure of how well the fitted profile (i.e. A–B–C) matches the actual surface profile, given as the root-mean-square of the difference between the actual and fitted profiles, i.e.

$$\text{fit error} = \sqrt{\frac{\sum (x_i - \hat{x}_i)^2}{N}}. \qquad (1)$$

A high fit error may imply a high degree of fine-detail irregularity on the cuticle surface due to, for example, surface striations (Swift & Smith, 2000).
8 *Cuticle density*: the number of cuticles per millimetre.
9 *Roughness*: a measure of the overall roughness of the surface profile, given as the total length of the profile divided by the length of the image. Thus, a completely flat profile will have a roughness of 1, but this will increase as the number and height of cuticles increases.
10 *Fitability*: this is a measure of the extent to which the surface profile consists of a recognizable cuticle pattern. Most images consisted of 3–6 overlapping cuticles, and only the regions to the extreme left and right of the surface profiles are not fitted by the algorithm. However, some images of damaged hair may contain regions with no – or only partial – cuticle coverage, in which case the majority of the profile remains unfitted by the algorithm. Fitability is given as the percentage of the surface profile length fitted by the algorithm.

Example analysis of two images

As an example of how the algorithm can be used to compare AFM images quantitatively, the two images shown in Fig. 3 were analysed. The calculated descriptor values are given in Tables 2 and 3.

Table 2. Cuticular descriptors calculated for the example image shown in Fig. 3(a).

	Mean	SD
Step height (nm)	386.27	103.89
Tilt (°)	3.73	1.18
Backtilt (°)	19.83	6.55
Layer spacing (nm)	473.43	147.40
Face distance (nm)	1451.66	708.69
Top distance (nm)	5955.05	1471.19
Fit error (nm)	31.78	10.12
Cuticle density (mm^{-1})	129.43	20.92
Roughness	1.0153	0.0046
Fitability (%)	75.17	13.05

Table 3. Cuticular descriptors calculated for the example image shown in Fig. 3(b).

	Mean	SD
Step height (nm)	437.59	166.07
Tilt (°)	3.02	1.50
Backtilt (°)	13.90	9.32
Layer spacing (nm)	505.19	158.77
Face distance (nm)	2225.84	1398.56
Top distance (nm)	7518.51	2997.80
Fit error (nm)	56.11	24.43
Cuticle density (mm^{-1})	69.73	23.43
Roughness	1.0206	0.0082
Fitability (%)	42.64	19.28

The calculated values for the untreated, root end image shown in Fig. 3(a) are given in Table 2. A total of 994 cuticle steps were found in all the 256 rows of the image. This means that the step height parameter is a mean of 994 measurements (although due to autocorrelation in the y-axis direction, these measurements are not fully independent). The mean step height value of 386 nm is in agreement with those found in the literature (You & Yu, 1997; Smith, 1998), as is the relatively

high standard deviation of 104 nm, demonstrating the large degree of variability in cuticular structure across the hair surface.

The mean value for cuticle layer spacing of 473 nm is also in agreement with the literature; Swift (1999) gives an approximate value for cuticle thickness of 500 nm. The difference between the step height and layer spacing values is due to the geometry of the cuticle surface, as indicated by the mean tilt and backtilt values of 3.73° and 19.83°, respectively. In Fig. 4, and in similar figures that have appeared in other publications, it misleadingly appears that the profiles have a saw-tooth shape, with the face surface of the cuticle being almost perpendicular to the fibre axis. However, this appearance is due to the difference between the *x*- and *y*-axis scales; in reality, the hair surface is fairly flat, although the difference between the tilt and backtilt angles is sufficient to produce the directional friction effect responsible for some important hair properties (Swift, 1999). In studies in which the swelling of the cuticular layers due to hydration is of interest, it may be that layer spacing is a more appropriate measure than step height, as it describes more accurately the distance between the cuticular layers. The large difference between layer spacing and face distance (distance A–B in Fig. 5) is a result of the face surface of the cuticle not being at 90° to the top surface; the cuticle edges are chisel-shaped, an effect increased by physical wear (and thus more pronounced at the distal end).

Table 3 shows the calculated descriptors for the bleached, distal-end images shown in Fig. 3(b). A total of 410 cuticle steps were found. The reduced number of cuticle measurements is one reason for the higher standard deviations of the descriptors, the other being the lack of uniformity in the cuticular pattern in comparison with the previous image. The cuticle density is 70 mm^{-1} (half that of the untreated, root-end image) due to the stripping away of the cuticular layers caused by the hair treatment and physical stress over time. The mean step height is 434 nm, higher than that of the previous image, probably due to the detachment of the very end of the cuticles from the layer beneath. However, the cuticle tilt is slightly lower: as the number of cuticles decreases, the cuticles lie flatter. The mean fit error is 56, twice that of the previous image, probably due to curvature in the normally flat cuticle top surface, again due to cuticle detachment, which may cause the hair to have a less glossy appearance. Finally, the fitability of Fig. 3(b) is much lower than that of Fig. 3(a), demonstrating that the cuticular structure is much less well defined, in this case due to the exposure of the underlying cortex.

Classification using PLS-DA

To demonstrate further the usefulness of an algorithm capable of automatically calculating a multivariate set of cuticular descriptors, a study is now presented in which cuticular descriptors are used to classify the hair samples according to one of the four following groups:

1 untreated, root end;
2 untreated, distal end;
3 bleached, root end;
4 bleached, distal end.

Although a number of sophisticated classification methods exist, some allowing estimates of the probability of class membership and the possibility of dual group membership (Frank & Friedman, 1989; Frank & Lanteri, 1989; Lavine, 2000a), a simple implementation of discriminant analysis (DA) is applied here, based upon partial least-squares regression (PLS), a multivariate regression technique widely used in the chemical sciences (Kemsley, 1996; Lavine, 2000b; Wold & Josephson, 2000). In PLS, a set of descriptor variables, **X**, is related to a set of response variables, **Y**, using a series of least-squares fitting steps. Collinearity in **X** and **Y** is handled by using a projection on to a reduced-dimension subspace, which for **X** is given by:

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{t}_r \mathbf{p}_r^{\mathrm{T}} + \mathbf{E} \qquad (2)$$

where R is the number of PLS model components, $\mathbf{t}_r$ is the scores vector for the rth component and $\mathbf{p}_r$ is the loadings vector for the rth component. The scores vectors describe the relationship between the samples in the model subspace and the loadings vectors describe the importance of each descriptor within the model. **E** is a matrix of residuals that can be used to determine how well each sample fits the model. PLS can be used for data exploration, i.e. uncovering the relationships both within and between the **X** and **Y** data sets. PLS can also be used to make response predictions for new samples, i.e. predicting **Y** for new **X** data. For more information on this technique, the reader is referred to the literature (Geladi & Kowalski, 1986; Hoskuldsson, 1988; Martens & Næs, 1989).

PLS is used as a discriminant analysis technique by constructing a **Y** matrix consisting of dummy variables used to indicate class membership. A PLS model is then built between the set of **X** descriptors and **Y**. The set of responses predicted by the model is used to produce a set of predicted class memberships. The true and predicted class memberships are then compared to evaluate how successful the model is at classifying the given samples. Given new **X** data, the PLS-DA model can be used to make class membership predictions for new samples. A schematic representation of PLS-DA is given in Fig. 6.

In the case here, there are a total of 38 samples (images), as summarized in Table 1. The cuticular descriptors (means and standard deviations) calculated for each hair image make up a 38×20 data matrix **X** (samples × descriptors) and so, each sample corresponds to a point (vector) in this 20-dimensional space. The **Y** data matrix consists of two columns, the first describing hair treatment (untreated = 0 / bleached = 1) and the second describing sampling position (root end = 0 / distal end = 1), as shown in Table 4. Thus, **Y** is a 38×2 data matrix (samples × dummy variables). Note that as the class membership for the hair samples is dependent upon two binary factors (treatment type and sampling position), two dummy variables can be used to define the four classes of hair sample. In another

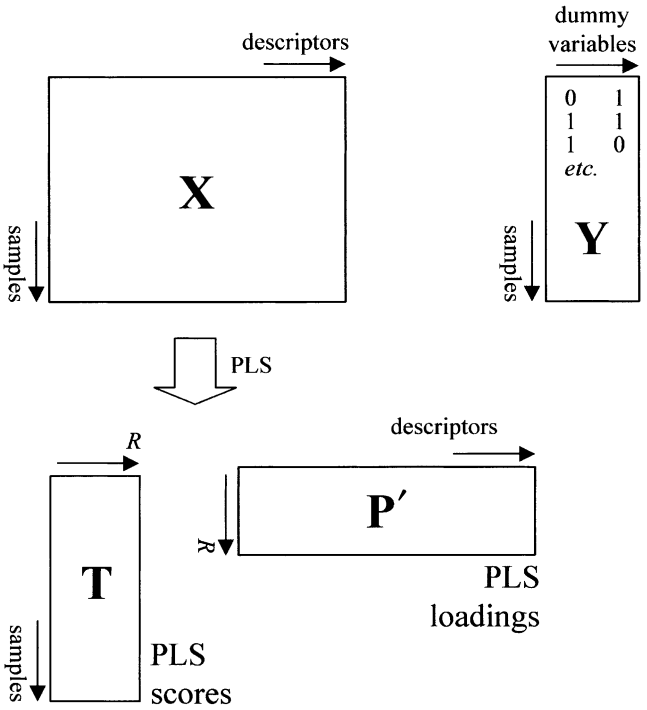

Fig. 6. Schematic of PLS-DA. The matrix of descriptors, **X**, is regressed against a matrix, **Y**, defining class membership for each sample. Collinearity in **X** is handled by decomposing the matrix into a set of scores, **T**, and loadings, **P**.

Table 4. Dummy variables for **Y** matrix used for PLS-DA.

		Y	
Sample number	Sample type	$\mathbf{y}_1$	$\mathbf{y}_2$
1		0	0
⋮	untreated, root end	⋮	⋮
6		0	0
7		0	1
⋮	untreated, distal end	⋮	⋮
20		0	1
21		1	0
⋮	bleached, root end	⋮	⋮
30		1	0
31		1	1
⋮	bleached, distal end	⋮	⋮
38		1	1

classification study, such as where hair samples were classified according to colour, it would be necessary to use a separate dummy variable for each class, e.g. brown (yes = 1/no = 0), blonde (yes = 1/no = 0), red (yes = 1/no = 0) and so on.

Before building the PLS model, the data were autoscaled (Martens & Næs, 1989), in which each column is mean-centred and then divided by its standard deviation. This preprocessing

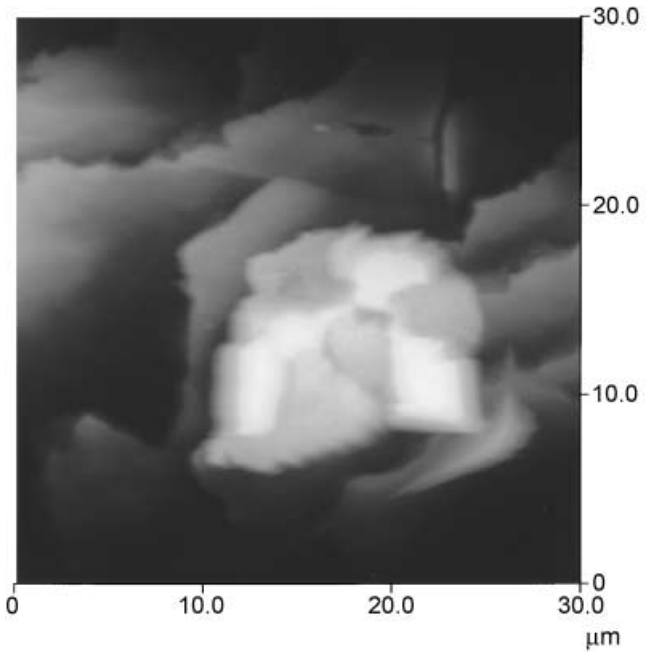

Fig. 7. Sample detected as an outlier by PLS.

ensures that each cuticular descriptor is given an equal chance to contribute to the model, regardless of the differences in physical units. A PLS model was then built, using leave-one-out cross-validation (Martens & Næs, 1989; Shao, 1993) to determine the optimum number of model components, found to be three (i.e. $R = 3$). Note that were more samples available (e.g. more than 1000), an alternative model validation strategy using separate training and test sets would be possible.

The PLS model built using the 38 samples detected two samples that had very different properties to the other samples, so-called 'outliers'. These outliers were detected by their very high residuals, indicating they did not fit the general pattern defined by the model. Outliers can have a negative effect during the model training stage and should be removed. One of the outlying samples, that of an untreated, distal-end sample, is shown in Fig. 7. This image has an unusual cuticle formation in the centre and it also appears that the fibre is not aligned horizontally. This shows how, whereas it is possible to use visual inspection to detect unusual samples, PLS-DA provides diagnostics for outlier detection that can be used for automatic outlier detection and removal. The other outlier, a bleached, distal-end sample, was also found to have an incorrectly aligned fibre and was removed. Although some further samples were found to fit the model quite poorly, these were retained in order to maintain a realistic level of sample variability.

The final PLS model was built on 36 samples and used three model components. Seventy-three per cent of the variation in the **X** data was explained by the model. The predicted **Y** responses were rounded to either 0 or 1, and used to produce a set of predicted class memberships that were compared with

Table 5. Summary of PLS-DA classification results.

Sample type	No. of samples	Predicted class: Untreated, root end	Untreated, distal end	Bleached, root end	Bleached, distal end	Outlier
Untreated, root end	6	6	0	0	0	0
Untreated, distal end	14	0	11	2	0	1
Bleached, root end	10	1	2	7	0	0
Bleached, distal end	8	0	0	0	7	1

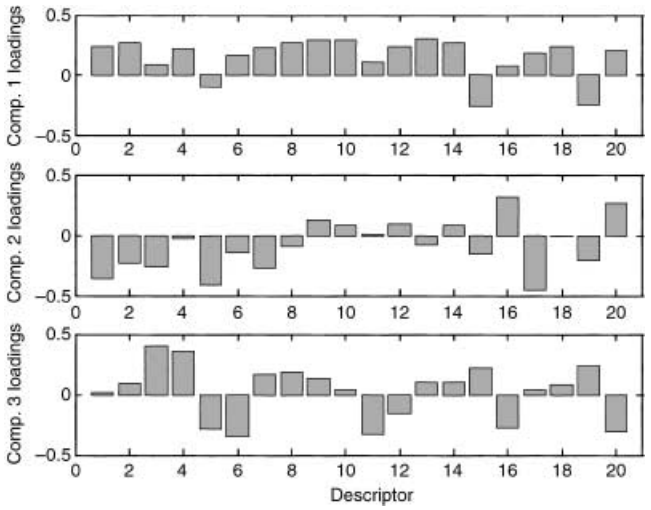

Fig. 8. PLS loadings for the three model components. The descriptors are: 1–2, step height (mean and standard deviation); 3–4, tilt; 5–6, backtilt; 7–8, layer spacing; 9–10, face distance; 11–12, top distance; 13–14, fit error; 15–16, cuticle density; 17–18, roughness; 19–20, fitability.

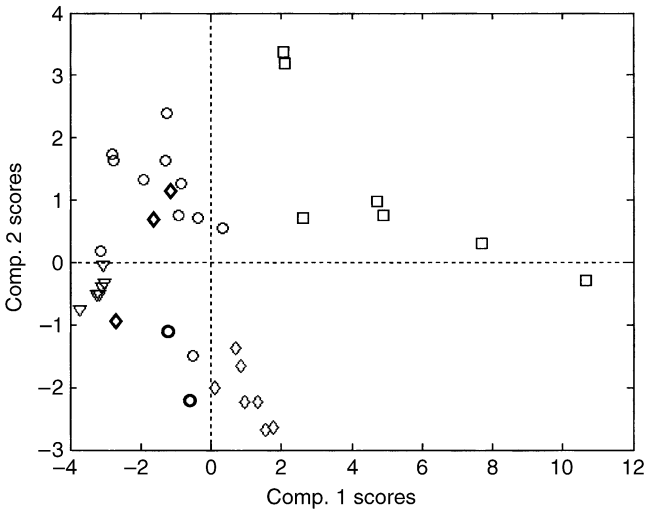

Fig. 9. Scores plot of component 1 vs. component 2. Symbols indicate the sample class: (▽) untreated/root end; (○) untreated/distal end; (◇) bleached/root end; (□) bleached/distal end. Misclassified samples are shown in bold, where the symbols refer to the true, not the predicted, class.

the true class memberships. Of the 36 samples, 31 were classified correctly, giving a success rate of 86%. The results are summarized in Table 5.

The model loadings are plotted in Fig. 8. These loadings describe the influence of each descriptor on each model component. Component 1 has a large negative contribution from the mean values of cuticle density (descriptor 15) and fitability (19), with most of the other descriptors having a positive contribution. This component can be interpreted as describing hair ageing and/or damage, both of which can lead to low cuticle density and poorly defined cuticular structure. Component 2 describes a large negative contribution from the mean values of step height (1), tilt (3), backtilt (5), layer spacing (7) and roughness (17). A combined increase in these parameters can indicate detachment of the cuticle layer ends from the cell beneath, another sign of hair damage. Component 2 also has a positive contribution from the standard deviations of cuticle density and fitability, variables that describe the degree of uniformity in the horizontal spacing between the overlapping cuticle layers. Finally, component 3 describes a positive contribution from the tilt descriptors (3,4) and a negative contribution from the backtilt descriptors (5,6) and the mean top distance (11). This component was found to discriminate between root and distal ends for the untreated samples, and describes how the angles the cuticles make with the central axis change with respect to cuticle density and sample age.

Figure 9 shows the scores for component 1 plotted against component 2. (Two further scores plots, 1 vs. 3 and 2 vs. 3, are also possible, but are not shown here.) This plot shows how the samples are clustered in the subspace defined by the first two components of the PLS model. Five samples were misclassified, these being shown in bold in Fig. 9, and a possible reason for the misclassification is given below.

Clustering

The six untreated/root end samples (∇) are found to the left of the plot (i.e. component 1 negative), and are closely clustered.

These samples are the most homogeneous in terms of cuticle structure, as the hair surface has not been greatly damaged by physical or chemical stress, and the algorithm finds similar descriptor values regardless of which particular image is analysed. The untreated/distal end (○) and bleached/root end (◇) samples also form fairly well-defined groups (with the exception of two untreated/distal-end samples and the misclassified samples).

By contrast, the bleached/distal end samples (□) are very scattered, indicating that representative sampling of this group is more difficult. These samples are found to the right of the plot: as component 1 describes negative contributions from cuticle density and fitability, it is logical that the samples that have been damaged both by cosmetic treatment and by long-term physical stress (being far from the root end) have the highest component 1 scores and are therefore characterized by low cuticle density and poor fitability.

These results suggest that the number of samples required to ensure representative sampling differs according to distance from the root end. For samples taken from near the root end, around 5–10 images may be sufficient to carry out before-and-after comparison studies of the effect of hair treatment on cuticular structure. Studies on distal-end samples may require many more images to provide statistically meaningful results.

Root- and distal-end samples

In general, the root-end samples (∇ and ◇) have negative component 2 scores and the distal end samples (○ and □) have positive component 2 scores. This shows that at the root end of the hair, where the cuticles are more abundant, the degree of tilt and step height of the cuticles is higher, as may be expected. At the distal end, where the cuticles are less abundant, the cuticles lie flatter and have a lower degree of backtilt due to physical wear. Distal-end cuticles also have a less uniform pattern, also due to the effect of prolonged physical stress, which chips away at the cuticle ends, leaving an irregular cuticle edge.

Effect of bleaching

For the samples measured at the root end, the bleached samples (◇) lie further to the right than the untreated samples (∇). This indicates that one effect of bleaching is to remove the cuticles that protect the central hair cortex, thus making the hair less resistant to breakage or splitting. The bleached samples also have lower component 2 scores, symptomatic of cuticle detachment, which also leaves the hair in a weaker condition, more vulnerable to subsequent damage.

For the samples measured at the distal end (untreated, ○; bleached, □), the removal of cuticle layers is even more pronounced, showing that bleaching of already vulnerable distal-end hair can lead to complete removal of the cuticular layer in some cases, exposing the underlying central cortex.

Although not carried out here, it is possible to use the PLS-DA classification algorithm to define the 'centre' of each class (after the removal of misclassified samples) and thereby to calculate a characteristic set of descriptor values for each class. In this way, the effect of hair treatments on cuticle density, tilt angle, etc., can be compared quantitatively.

Sample misclassification

Five samples were misclassified, these being shown in bold in Fig. 9. In four cases, the error is due to an untreated/distal-end sample being misclassified as a bleached/root-end sample or vice versa. As the effect of bleaching is to degrade the cuticle layer – something that also occurs naturally as the hair ages (i.e. moves further from the root end) – then it is logical that these two hair groups are the most difficult to separate. One option, not pursued here, would be to build a new PLS-DA model using only these two groups. This would enable a hierarchical classification scheme and lead to improved classification results. This would be especially useful for the classification of new hair samples where the class membership is unknown in advance.

Conclusions

An algorithm for the automatic, quantitative characterization of AFM images of human hair samples has been presented. The cuticular descriptors provide a range of information about the hair surface, and it is thought that this will enable faster and more accurate analysis of the effect of physicochemical conditions on the hair surface. The multivariate characterization of cuticular structure has been shown to be useful for classification of hair samples in an example study.

There are still some areas for improvement in the algorithm used to analyse the AFM image. False location of cuticle edges due to surface irregularities, although minimal, still occurs on occasion. Optimization of the size of the smoothing window may help, although for the images used here, a one-size-fits-all approach using a window 2000 nm in width seemed to work quite well. The characterization of hair images with only partial cuticle coverage, such as that shown in Fig. 3(b), is also problematic, as demonstrated by the importance of the fitability descriptor. An image mapping technique capable of distinguishing between cuticle and cortex surfaces could be useful here. Finally, the algorithm only considers horizontal changes in the surface profile, and thus does not explicitly consider the shape of the cuticle edges running vertically. Some experimentation with techniques capable of finding strong edges in an image (Gonzalez & Woods, 1992), followed by quantifying the degree of 'serratedness' of the cuticle edge, was carried out, but this has not yet been perfected.

Acknowledgements

S.P.G. and M.M.C.F. gratefully acknowledge financial support from the State of São Paulo Research Foundation (FAPESP).

V.F.M. and E.L. gratefully acknowledge financial support from FAPESP and the Centers for Research, Innovation and Dissemination (CEPID) program, and the Brazilian National Research Council (CNPq).

References

Ahn, H.J. & Lee, W.-S. (2002) An ultrastructural study of hair fiber damage and restoration following treatment with permanent hair dye. *Int. J. Dermatol.* **41**, 88–92.

Binnig, G., Quate, C. & Gerber, C. (1986) Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933.

Blach, J., Loughlin, W., Watson, G. & Myhra, S. (2001) Surface characterization of human hair by atomic force microscopy in the imaging and F-d modes. *Int. J. Cosmet. Sci.* **23**, 165–174.

Feughelman, M. & Willis, B.K. (2001) Mechanical extension of human hair and the movement of the cuticle. *J. Cosmet. Sci.* **52**, 185–193.

Frank, I.E. & Friedman, J.H. (1989) Classification: oldtimers and newcomers. *J. Chemom.* **3**, 463–475.

Frank, I.E. & Lanteri, S. (1989) Classification models: discriminant analysis, SIMCA, CART. *Chemom. Intell. Laboratory Syst.* **5**, 247–256.

Garcia, M.L., Epps, J.A., Yare, R.S. & Hunter, L.D. (1978) Normal cuticle-wear patterns in human hair. *J. Soc. Cosmet. Chem.* **29**, 155–175.

Geladi, P. & Kowalski, B.R. (1986) Partial least squares regression: a tutorial. *Anal. Chim. Acta*, **185**, 1–17.

Gillen, G., Roberson, S., Ng, C. & Stranick, M. (1999) Elemental and molecular imaging of human hair using secondary ion mass spectrometry. *Scanning*, **21**, 173–181.

Goddard, E. & Harris, C. (1987) An ESCA study of the substantivity of conditioning polymers on hair substrates. *J. Soc. Cosmet. Chem.* **38**, 233–246.

Gonzalez, R. & Woods, R. (1992) *Digital Image Processing*, pp. 414–428. Addison-Wesley, Boston, MA.

Hadjur, C., Daty, G., Madry, G. & Corcuff, P. (2002) Cosmetic assessment of the human hair by confocal microscopy. *Scanning*, **24**, 59–64.

Hess, W., Seegmiller, R., Gardner, J., Allen, J. & Barendregt, S. (1990) Human hair morphology: a scanning electron microscopic study on a male caucasoid and a computerized classification of regional differences. *Scanning Microsc.* **4**, 375–386.

Hoffmann, K. (1991) Statistical evaluation of the evidental value of human hairs possibly coming from multiple sources. *J. Forensic Sci.* **36**, 1053–1098.

Hoskuldsson, A. (1988) PLS regression methods. *J. Chemom.* **2**, 211–228.

Jollès, P., Zahn, H. & Höcker, H. (eds) (1997) *Formation and Structure of Human Hair*. Birkhäuser, Basel, Switzerland.

Kemsley, E.K. (1996) Discriminanat analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods. *Chemom. Intell. Laboratory Syst.* **33**, 47–61.

Kreplak, L., Merigoux, C., Briki, F., Flot, D. & Doucet, J. (2001) Investigation of human hair structure by microdiffraction: direct observation of cell membrane complex swelling. *Biochim. Biophys. Acta*, **1547**, 268–274.

Lavine, B.K. (2000a) Clustering and classification of analytical data. *Encyclopedia of Analytical Chemistry* (ed. by R. A. Meyers), pp. 9689–9710. John Wiley & Sons, Chichester, U.K.

Lavine, B.K. (2000b) Chemometrics. *Anal. Chem.* **72**, 91R–97R.

Martens, H. & Næs, T. (1989) *Multivariate Calibration*. John Wiley & Sons, Chichester, U.K.

McMullen, R.L., Jachowicz, J. & Kelty, S.P. (2000) Correlation of AFM/LFM with combing forces of human hair. *IFSCC Mag.* **3**, 39–45.

McMullen, R.L. & Kelty, S.P. (2001) Investigation of human hair fibers using lateral force microscopy. *Scanning*, **23**, 337–345.

O'Conner, S.D., Komisarek, K.L. & Baldeschwieler, J.D. (1995) Atomic force microscopy of human hair cuticles: a microscopic study of environmental effects on hair morphology. *J. Invest. Dermatol.* **105**, 96–99.

Orfanos, C.E. & Happle, R. (eds) (1990) *Hair and Hair Diseases*. Springer, Berlin.

Pötsch, L. (1996) A discourse on human hair fibres and reflections on the conservation of drug molecules. *Int. J. Legal Med.* **108**, 285–293.

Pozebon, D., Dressler, V.L. & Curtius, A.J. (1999) Análise de caeblo: uma revisão dos procedimentos para determinação de elementos traço e aplicações. *Química Nova*, **22**, 838–846.

Robbins, C.R. (1994) *Chemical and Physical Behavior of Human Hair*, 3rd edn. Springer, Berlin.

Ruetsch, S.B., Kamath, Y. & Weigmann, H.D. (2000) Photodegradation of human hair: an SEM study. *J. Cosmet. Sci.* **51**, 103–125.

Rugar, D. & Hansma, P. (1990) Atomic force microscopy. *Phys. Today*, **43**, 23–30.

Sauermann, G., Hoppe, U., Lunderstädt, R. & Schubert, B. (1988) Measurement of the surface profile of human hair by surface profilometry. *J. Soc. Cosmet. Chem.* **39**, 27–42.

Savitzky, A. & Golay, M.J.E. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639.

Shao, J. (1993) Linear model selection by cross validation. *J. Am. Stat. Assoc.* **88**, 486–494.

Smith, J.R. (1997) Use of atomic force microscopy for high-resolution non-invasive structural studies of human hair. *J. Soc. Cosmet. Chem.* **48**, 199–208.

Smith, J.R. (1998) A quantitative method for analysing AFM images of the outer surfaces of human hair. *J. Microsc.* **191**, 223–228.

Srivastava, A.K. & Gupta, B.N. (1994) The role of human hairs in health and disease with special reference to environmental exposure. *Vet. Human Toxicol.* **36**, 556–560.

Swift, J.A. (1999) Human hair cuticle: biologically conspired to the owner's advantage. *J. Cosmet. Sci.* **50**, 23–47.

Swift, J.A. & Bews, B.W. (1976) The chemistry of the human hair cuticle. The isolation and amino acid analysis of various subfractions of the cuticle obtained by pronase and trypsin digestion. *J. Soc. Cosmet. Chem.* **27**, 289–300.

Swift, J.A. & Brown, A.C. (1972) The critical determination of fine changes in the surface architecture of human hair due to cosmetic treatment. *J. Soc. Cosmet. Chem.* **23**, 695–702.

Swift, J.A. & Smith, J.R. (2000) Atomic force microscopy of human hair. *Scanning*, **22**, 310–318.

Wold, S. & Josephson, M. (2000) Multivariate calibration of analytical data. *Encyclopedia of Analytical Chemistry* (ed. by R. A. Meyers), pp. 9710–9736. John Wiley & Sons, Chichester, U.K.

You, H. & Yu, L. (1997) Atomic force microscopy as a tool for study of human hair. *Scanning*, **19**, 431–437.

Appendix

An algorithm for the automatic characterization of the cuticular structure of hair surfaces using AFM imaging

In the study described in this paper, the AFM images, depicting horizontally aligned hair fibres with the distal end to the left and the root end to the right, showed an area of $30\,000 \times 30\,000$ nm. This image is stored in a matrix, $\mathbf{Z}$, with dimensions 256×256 data points, giving a data resolution of ≈ 118 nm per data point. Each row of $\mathbf{Z}$, denoted $\mathbf{z}_n$ and with dimensions 1×256, describes a surface height profile.

The steps of the algorithm for characterization of the cuticular structure are as follows.

1 Take the first horizontal surface profile, $\mathbf{z}_1$, and calculate the first-order derivative spectra using a second-order polynomial Savitzky–Golay filter with a window size of 2000 nm.
2 Search the first-order derivative spectrum for the highest peak: this peak represents the position of a cuticle face end. Note the position of this peak, find its beginning and end, and then eliminate it from the first-order derivative spectra. Continue locating subsequent peaks and stop when all peaks above a threshold of $0.07 \times$ data resolution have been found.
3 Locate the beginnings and ends of the cuticle face ends (i.e. points A and B, respectively, in Fig. 5). The beginning is found by performing a line search for a minimum point in the surface profile, starting 2000 nm to the left of the position of the face end. The end is found by performing a line search for a maximum point in the surface profile, starting 2000 nm to the right of the position of the face end. Repeat this step for all cuticle face ends found in Step 2. In cases where, due to surface irregularities, (a) the step height of the face end is less than 200 nm or (b) the positions of two apparent face ends overlap, remove the specious points.
4 Reconstruct a cuticular profile using the points found in Step 3. Use this profile to calculate the descriptors listed in the main text for each cuticle. Note that some descriptors (i.e. tilt, thickness) require that at least two cuticle face ends are present.
5 Repeat steps 1–4 for the subsequent rows of the image, keeping a record of all calculated descriptors. Once all rows have been analysed, calculate the mean and standard deviation of each descriptor.