Analytical Methods

PAPER

Cite this: *Anal. Methods*, 2015, **7**, 1707

Amino acid composition of human scalp hair as a biometric classifier and investigative lead

Ayat H. B. Rashaid,[a] Peter B. Harrington[a] and Glen P. Jackson*[bc]

Hair shaft analysis is becoming increasingly important in several applications of forensic science. Keratin is the key component of human scalp hair and is composed of all 21 known amino acids, albeit in very different proportions. The method for the amino acid determination included protein acid hydrolysis followed by trimethylsilyl (TMS) derivatization of the amino acids, and the subsequent quantitation using gas chromatography/mass spectrometry (GC/MS). The amino acid composition of scalp hair of 64 Jordanian subjects (33 males and 31 females) with ages ranging from 1 to 77 years have been analyzed. Statistical comparisons between classification groups were based on the abundance of 14 abundant and acid-resistant amino acids, and included classification of hairs with a fuzzy rule building system (FuRES). Using leave-one-individual-out cross-validation, the FuRES classification rate was 94% for sex, 83% for age group, and 61% for the region of origin. For predicting sex from amino acid concentration in hair, the essential amino acids Phe and Thr gave the most significant differences with respect to their *F* statistic (*i.e.*, ratio of between-group to within-group variation), so they are the most discriminating for sex. Based on the same hair analyses, the non-essential amino acids Gly and Ala provide the largest loading scores classifying the subjects into two arbitrary age groups, <49 and >49 years. For region of origin classification, the amino acids Cys and Tyr had the highest loading scores in the classification rules and were therefore most discriminating. The techniques developed through this paper could complement the current methods of hair analysis, which include physical examinations and genomic or mitochondrial DNA analysis.

Received 30th October 2014
Accepted 25th November 2014

DOI: 10.1039/c4ay02588a

www.rsc.org/methods

Introduction

Human hair has been a subject of intensive investigation for a number of years. Study of the hair shaft is becoming increasingly important in several applications of forensic science.[1,2] In addition, hair is one of the more robust non-skeletal tissues in decomposed bodies and postmortem cases of forensic interest.[2] The growth rate of human scalp hair ranges from 0.8 to 1.3 cm per month.[3–5] Hair consists of the cuticle, cortex, and medulla which are formed *via* the keratinization of dying cells growing out of the root. Hair is mainly composed of alpha keratin protein (65–95%), and the remaining constituents are water, lipids, pigment, and trace elements.[4,6–8] Keratin contains strong disulfide bonds that link adjacent keratin chains and contains intra-molecular bonding of polar and non-polar amino acids, which lead to a stable structure that is highly resistant to chemical and biological degradation. The amino acid composition of human hair keratin may vary significantly among individuals without exhibiting obvious phenotypic differences.

Keratin, the protein which is the key component of hair, is composed of 21 known amino acids.[9] Changes in the amino acid composition of hair have long been a target of interest to nutritional scientists and health scientists. Profiling amino acids in human hair could be utilized for characterization of various types of hair derived from different individuals.[3,10] Therefore, the amino acid profile of human hair could be a useful tool with regard to classification of individuals into groups of biological or behavioral traits, in a complementary manner to work using isotope ratio analysis of amino acids in human hair.[11]

Hair follicles undergo a growth cycle comprising three stages: the anagen stage is the active growing phase where the cells in the bulb of the follicle are rapidly dividing; scalp hair stays in this phase from 4 to 6 years. The catagen stage, which is the transition phase, lasts for 2–3 weeks. During this phase cell division stops and the hair shaft becomes fully keratinized. The telogen stage is the resting phase wherein hair growth stops completely. The duration of this stage is from 4 to 6 months.[4,9,12] Nuclear DNA profiling results are typically only successful on hair roots in an appropriate growth cycle (anagen or catagen),

[a]*Center for Intelligent Chemical Instrumentation, Clippinger Laboratories, Department of Chemistry and Biochemistry, Ohio University, Athens, Ohio 45701, USA*

[b]*Department of Forensic & Investigative Science, West Virginia University, Morgantown, WV, 26506, USA. E-mail: glen.jackson@mail.wvu.edu; Fax: +1 304 293 2663; Tel: +1 304 293 9236*

[c]*C. Eugene Bennett Department of Chemistry, West Virginia University, Morgantown, WV, 26506, USA*

and when a subject's nuclear DNA is already in the reference database.[13] However, most hair found at crime scenes has roots that are in the telogen or resting phase,[14,15] wherein nuclear DNA is in very low abundance and severely degraded, and databases only contain DNA profiles from a small percentage of the population.

Mitochondrial DNA (mtDNA) profiling can be performed on the hair shaft, but the accompanying statistics are far less powerful than nuclear DNA because mtDNA is inherited maternally.[16] Accordingly, when DNA is not recoverable from hair samples, or when reference DNA samples are not available, analysis of the amino acid content of shaft hair could be a reliable method for linking questioned hair at the crime scene with suspects or victims, or for determining biometric traits and investigative leads about a donor.

The amino acid composition of hair is known to be influenced by a number of factors including age,[8,17–20] sex,[18,21] hair color,[18,22] and race,[22–24] as well as dietary habits and geographic origin.[9,17,25,26] In the context of investigating the relationship between age and amino acids of human hair, a decrease in tryptophan (Trp) concentration was noted between age groups 2–5 years and 13–19 years.[8] Furthermore, a gradual increase in Trp was observed between the ages of 13–19 and 60–80 years.[19] Statistically significant changes with increasing age showed increasing concentrations of serine (Ser), arginine (Arg), and proline (Pro), whereas a decrease in concentration of aspartic acid (Asp), glycine (Gly), alanine (Ala), valine (Val), isoleucine (Ile) and lysine (Lys) were shown with increasing age.[8] It has also been observed that cystine (Cyt) and cysteine (Cys) content decreased with age. For example, higher values of Cyt amino acid was found among infants (9 months) compared with children (8–9 years).[17] However, both Clay *et al.*[18] and Wilson and Lewis[21] concluded that no relation can be demonstrated between Cys and Cyt contents of the hair and age of the subjects.

Investigators have also examined the effects of ethnicity and sex on the amino acid composition in human hair. Regarding sex differences, Clay *et al.* found significantly more Cyt and Cys in the hair of the males compared to the females.[18] Rutherford and Hawk[22] also found that the Cyt and Cys content of Caucasian males was greater than that of females. However, in related work, Wilson and Lewis[21] analyzed twenty-three samples of hair from children and six samples from adults and found no consistent relationship between sex and the content of Cys and Cyt. Panayiotou[24] reported that females have a higher content of glutamic acid (Glu), Asp, and Trp amino acids than males.

A limited number of studies have investigated the amino acid composition of scalp hair under different dietary habits of different regions of origin. Takanohashi[17] conducted an experiment to determine whether the Cys content in the hair of infants and children could be changed under various circumstances *i.e.*, deficiency of vitamin A, animal protein, and other nutrients in their diet. Hair samples were obtained from three different districts; region A (people living on Japanese barnyard millet and very little fish), region B (people living on rice and moderate amount of fish), and region C (people living on a well-balanced diet). Results indicated that the Cys content of hair was low when the subject's diets were deficient in vitamin A or in high quality of protein, while it was observed that by supplementing their diets with liver oil or with plenty of milk, the Cys content increased. These results are in agreement with observations of experiments conducted on rats.[26] Heard and Lewis[26] reported that the addition of supplementary methionine (Met) to the basal diet known to be deficient in its content of sulfur containing amino acids resulted in the production of hair with a higher percentage of Cys than the hair of animals receiving the basal un-supplemented diet. Moreover, the experiments of Lightbody and Lewis[25] clearly show that the Cys content of hair can be changed by diet. They recognized a lower Cys content in the hair of rats than in those of rats fed on diets high in Cys content.

The essential amino acids of keratin that cannot be synthesized from other substances in the human body (*i.e.*, *de novo* synthesis) are leucine (Leu), Ile, Thr, Trp, Met, Lys, Phe, and Val. These amino acids are normally furnished in the diet in the form of protein (*e.g.*, meat, milk and eggs).[9] Therefore, dietary habits could be an important variable for the characterization of the essential amino acids in human hair.

The previous literature provides some evidence to suggest that sex, age, and region of origin may produce variation in certain amino acid contents of human hair. However, relatively few extensive studies of this nature have been reported. Profiling amino acids in human scalp hair has never been studied in the Arab World, and, to our knowledge, multivariate classification of individuals by their hair amino acid profiles has not been demonstrated before. Jordan has been selected for this purpose because it has three distinct geographical regions (mountain, valley, and desert) with distinct dietary habits for each region. This study is the first to show the feasibility of the amino acid profiles to classify subjects based on sex, age, and region of origin.

With reference to the previous studies, the general statement seems valid that relations between the amino acid distribution and each of sex, age, and region of origin were not established because of the number of samples was typically less than 20.[8,17,22,27] In the present study, the scalp hair from 64 subjects spanning different age groups from three different regions of Jordan (mountain, Jordan valley, and Bedouin region) with adequate information about the subjects (age, sex, height, weight, hair color, medical history, dietary habits, and chronic or acute diseases) were studied to establish the feasibility of using amino acid profiles of hair for classification of individuals. Moreover, the number of subsamples (3 trials) for each subject provided greater confidence for the method and the statistical tests.

Previous work has shown that Trp, Cyt, Met, Asn, Gln, His, and Arg amino acids are detrimentally affected by acid hydrolysis.[28] For this reason, this study focuses on the 14 most abundant amino acids in the protein of human hair, as obtained by acid hydrolysis using 6 M of hydrochloric acid (HCl).[28–31] The amino acids were quantified *via* their trimethylsilyl (TMS) derivatives[32–35] using gas chromatography/mass spectrometry (GC/MS).[29,36,37] The main objective of this study is to provide a discriminating technique for distinguishing

between hair samples from different human subjects based on factors related to sex, age, dietary habits, and region of origin that have been assumed to contribute to variations in the amino acid contents of human scalp hair.

Methods and instrumentation

Hair samples

The collection of hair samples for this study was approved by the Ohio University institution review board (IRB# 12X029) and King Abdullah University hospital institution review board (IRB# 10/215/2444). Scalp hair from 64 healthy Jordanian individuals was collected randomly from three different geographical regions of Jordan, each with a distinct geography, culture, and dietary habits. People who live in each region have their own nutritional habits. For example, the Bedouin, who live in deserts, typically depend on dairy products and lamb meat. Residents of the mountainous region depend more on fruits, vegetables, and rice. The primary food of people in the valleys is fish. Jordan therefore offers a unique opportunity to study people with quite distinct geographic and dietary lifestyles.

The names, addresses and phone numbers of potential subjects were taken from a database of the General Civil Status department. This department is a national bureau that stores all the personal information about all Jordanians. Pregnant women and subjects who were diagnosed with any chronic or acute disease were excluded from the study. The subjects were 33 men and 31 women between the ages of 1 and 77 years. They were also classified according to three geographical regions: 29 from the mountain, 23 from the valley, and 12 from the eastern desert regions.

In the scheduled home visits, consent forms from adults and assent forms from minors were signed prior to filling out interview questionnaires regarding age, sex, biometric information, health status, hair care, cosmetic treatment, and dietary habits. A pencil-diameter (1 cm) lock of hair adjacent to the scalp was cut using surgical scissors from the posterior vertex (back of the head) of each subject. All samples were coded and no specific individual was identifiable thereafter. All hair used in this study was reported to be rarely altered chemically, *i.e.*, bleaching, chemical straightening, dyes, and permanent waves. Because of Jordan's cultural factors, female hair was never directly exposed to the sun and none of the subjects smoked regularly or drank alcohol. These factors may not transcend other populations or study groups.

Chemicals and supplies

Any variation in conditions from one run to the next was controlled by referencing all data to an internal standard that was added to each acid hydrolysate.[38] L-Norvaline (Nor) (Sigma-Aldrich, St. Louis, MO, USA) was used as an internal standard because it is an *exogenous amino acid* that is not synthesized by the human body. The derivatizing agent was *N*,*O*-bis(trimethylsilyl) trifluoroacetamide (BSTFA) (Supelco Analytical, Bellefonte, PA, USA). A solution of 6 M HCl was used as a hydrolysis agent to liberate free amino acids from the hair shaft. Acetonitrile, methanol, acetone, and chloroform were purchased from GPS Chemicals (Columbus, OH, USA).

Glass vials of 4 mL volume with a phenolic rubber lined cap (Qorpak, Bridgeville, PA, USA) were used in this study. Solutions were filtered through a 13 mm × 0.45 μm, polyvinylidene difluoride (PVDF) filter (Bonna-Agela Technologies, Wilmington, DE, USA). The nitrogen generator was purchased from Parker Hannifin Corporation (Haverhill, MA, USA).

Hair analysis

Approximately 1 g from each hair specimen was rinsed twice in deionized water and vortexed in 1 : 1 : 1 (by volume) solution of chloroform, methanol, and acetone for 1 min to remove surface contamination, external oils, and lipids.[39] The hair was then dried under a constant stream of 4 L min^{-1} of nitrogen to remove the excess reagents. Afterwards, the hair was cut into 1 cm lengths and approximately 20 mg of hair was physically degraded and homogenized by placing the hair into a capped plastic vial with 8 stainless steel beads and pulverized for 1 min using a mini bead beater (BioSpec Products Inc., Bartlesville, OK, USA) at 4800 rpm. Because there is no known hair standard available with which to validate the method, horse heart myoglobin was used. Although myoglobin does not contain disulfide bonds, it does contain a high degree of alpha-helical structure; much like the keratins found in hair. A report on the method development and optimization is provided elsewhere.[28]

Determining the amino acid content of human hair involved the acid hydrolysis in 6 M HCl until its individual amino acids separate and become available for detection. The hair powder from each sample was separated into three subsamples and weighed to 3.00 mg ± 0.03 mg (1 s.d.) and transferred into a 4 mL glass vial with a phenolic rubber lined cap. Thereafter, 0.3 mg of Nor was added. Each subsample was hydrolyzed by adding 0.3 mL of 6 M HCl to the vial and was tightly capped. To heat the samples, an aluminum block with holes to accommodate the glass vials was mounted on a hot-plate for 24 h at 110 °C.[29–31,40] After cooling the solution to room temperature for 5 min, the solutions were filtered and dried under a constant 4 L min^{-1} nitrogen stream because the existence of excess moisture can result in poor reaction yield and instability of the derivatized products. The three subsamples of hair from the same individual were digested across spans of 3 days. In some cases, the amount of collected hair from an individual was not sufficient to prepare three digestions, so they were done in duplicate.

The dried amino acid residues were dissolved in a 100 μL aliquot of acetonitrile and derivatized with the addition of a 100 μL aliquot of BSTFA.[32,38,41] The glass vial was tightly capped, ultrasonicated for 1 min and mounted on a hot-plate for 30 min at 100 °C.[42,43] After cooling the solution at room temperature for 10 min, a 100 μL aliquot of the solution was transferred into a 3 mL glass auto-sampler vial containing a 200 μL plastic insert.

Instrumentation

The derivatized amino acids were analyzed using GC/MS and identified according to their retention time[44,45] and EI

fragmentation patterns using independently analyzed BSTFA derivatives of the amino acid standards. Amino acid derivatives were analyzed with a Shimadzu GC/MS instrument (QP-2010SE, Scientific Instrument, Inc. Columbia, MA, and USA). Amino acids were separated on 5% diphenyl-dipolymethylsiloxane capillary column (SHRX1-5MS, 30 m × 0.25 mm × 0.25 μm). Ultra high purity Helium was used as a carrier gas at a constant flow rate of 21.5 mL min^{-1}. The column flow was 1.0 mL min^{-1}. The temperature of the column was programmed to rise from 70 °C to 170 °C at a rate of 10 °C min^{-1}, and then was ramped to 280 °C at a rate of 30 °C min^{-1}, at which time it was held for 3 min. The total run time was 16.6 min. Sample injection was performed in split injection mode (1 : 20 ratio) at 280 °C using an injection volume of 1 μL. Triplicate injections of triplicate samples were run in random block design using acetonitrile as a solvent blank before each injection. Three solvent vials of acetonitrile, methanol, and acetone were used as cleaning solvents for the autosampler injection syringe.

The mass spectrometer was operated in Fast Automated Scan and SIM Type mode (FASST), which switches between full scan mode and selected ion monitoring during each GC analysis. The mass spectrometer was operated in full scan mode (TIC) from m/z of 50 to m/z of 500 with a scan time of 0.3 s, and selected ion monitoring mode (SIM) with a scan time of 0.2 s. For SIM, the appropriate ion set of one target ion and two reference ions as a characteristics mass fragment of the derivatized amino acid was used (Table 1). Selection of the characteristic ions for the quantification ion is very important. The target ions are the fragment ions of the highest intensity (base peak), while the reference ions are the two fragment ions of the next greatest intensities.[28,42] Whereas the TIC shows partial baseline overlap for Ile and Pro, the selected quant ions enabled sufficient discrimination between the two amino acids to enable unambiguous peak integration and quantitation.

Table 1 Target and reference ions for SIM mode as characteristics mass fragments of the 14 amino acids

Amino acid	Target ion (m/z)	Reference ions (m/z)
Ala	116	73, 147
Gly of di-TMS derivative	102	73, 147
Val	144	73, 218
Nor	144	73, 218
Leu	158	73, 102
Ile	158	73, 218
Pro	142	73, 216
Gly of tri-TMS derivative	174	73, 248
Ser	204	218, 73
Thr	73	218, 291
Asp	232	147, 100
Glu	246	73, 128
Phe	218	192, 73
Cys	220	156, 73
Lys	174	156, 73
Tyr	218	179, 280

Data analysis

Statistical analyses were performed using MATLAB R2012b (The MathWorks Inc., Natick, MA). All statistical tests were conducted at the 95% confidence level.

Amino acids in hair protein were determined by the measurement of their relative peak areas measured by GC/MS that were referenced to the peak area of norvaline internal standard. In previous work,[28] selected ion monitoring (SIM) yielded the best performance. The concentration (mg amino acid/mg hair) for each amino acid was calculated using external calibration curves of liquid standards. A random block design was used to collect triplicate measurements of each hair digestion across spans of 2–3 days.

A variety of statistical techniques were used to correlate or discriminate individuals or groups of individuals based on amino acid compositions. For example, to analyze the multivariate data, a multivariate fuzzy rule-building expert system (FuRES)[46] was used to build classifier and reveal the characteristic amino acids for the classes. FuRES constructs a classification tree that allows the visualization of the inductive structure of the classes. The inductive classification tree is a collection of membership functions for which each branch is a multivariate fuzzy rule. The entropy of the system decreases as the classification tree branches until all the objects are grouped into a single class at the leaves of the tree.[46] FuRES has been compared with more traditional classification methods in a variety of different analytical applications.[47–55]

Because FuRES is a supervised classifier, the results were validated using leave-one-out cross-validation. Several approaches were used for validation. Leave-one-out cross-validation involves leaving out a measurement (or profile), constructing the FuRES rules based on the remaining measurements (or profiles), and then classifying the omitted measurement (or profile). The classification prediction is repeated for every measurement (or profile) to obtain an overall estimate of the ruggedness of the classification system. The FuRES discriminant weights are stored for each model that is built during this process and the average weight with an arbitrary 95% confidence interval can be used as a diagnostic to determine which amino acids are most characteristic for a specific the classification. Note that because a random block design was used in acquiring the data, that replicate amino acid profiles were typically collected on different days.

Besides leaving each profile out, another mode of validation averaged 3 the amino acid profiles for each separate hair digestion. The averages were classified and each averaged amino acid profile corresponded to a separate digestion. Lastly, the averages were calculated for each triplicate and the three digestions, so each average amino acid quantity corresponded to a unique individual. This test is the most stringent because the model was built by leaving out profiles from an individual as opposed to leaving out one of several replicate measurements of an individual. Using averaged-within-person profiles is therefore the most honest representation because each profile is truly an independent measure of the classification factor.

The accuracy of the tests were evaluated by their sensitivity and specificity of the classifier predictions.[56] Sensitivity is

defined as the ratio of positive classifier responses for positive subjects. Specificity is defined as the ratio of negative classifier responses for negative subjects. Sensitivity plotted with respect to the false positive rate (1–specificity) can be displayed graphically using the receiver operating characteristic (ROC) curve.[56] The receiver operating characteristic curve (ROC) was generated to evaluate the sex, the age group, and the region of origin classifiers. The accuracy of a classifier depends on its ability to maintain the sensitivity and the specificity, so that the ideal result will have an area under the curve (AUC) of 1.[56] The area under the curve is calculated using the trapezoidal integration function of MATLAB.

Analysis of variance (ANOVA) is a statistical technique for determining the significance of the differences between two or more groups; it was used to compare the concentration (mg amino acid/mg hair) of each amino acid in different groups of subjects and to find which amino acids have the largest ratio between within-person variance and between-person variance.[57,58] Based on the ANOVA results of region of origin groups, *post-hoc* statistical tests such as Tukey's honest significant difference (HSD) was used to identify which groups were significantly different from another. Principal component analysis (PCA) was used to reduce the number of variables in a data set while retaining as much variation from the original data set. PCA was also used to visualize the clustering of the subjects scores.[59] ANOVA-PCA is a method that combined ANOVA with PCA.[60] Pooled ANOVA-PCA[61,62] was used to evaluate the differences among the class means for the three regions of origin, sex, and age after averaging the 9 profiles for the 64 individuals.

Results and discussion

The 14 amino acids of interest in this study are Ala, Gly, Val, Leu, Ile, Pro, Ser, threonine (Thr), Asp, Cys, Glu, Phe, Lys, and Tyr. Sharp single peaks for all amino acids were obtained except for Gly, which gave both di-trimethylsilyl derivative and tri-trimethylsilyl derivatives that were obtained after BSTFA derivatization reaction.[28,35] See Fig. 1. The observation of two products for certain amino acids, although not ideal, does not necessarily prevent quantitation. If both products are chromatographically resolved, the sum of the two peak heights or peak areas can be used for quantitation.

FuRES was used to evaluate the ability to classify the subjects according to the sex, age group, and region of origin based on multivariate analyses of the abundance of 14 amino acids in human hair. The age group was arbitrarily defined as two groups: one containing 1–44 year-olds and the other group containing 44–77 year-olds. FuRES was applied to the 528 amino acid profiles that for the most part were obtained from (3 replicate measurements of 3 separate digestions) for each individual. However, for some individuals the quantity of hair that was collected was not sufficient for 3 digestions, so 48 profiles were not collected. In a similar manner, FuRES was also applied to the 177 average amino acid profiles of the three injections for each digestion. In the third experiment, FuRES was applied to the average of 9 measurements of each individual (average across individuals). The classifier prediction rates of sex, age, and region of origin after leave-one-out cross-validation are presented in Table 2.

Using leave-one-profile-out cross-validation, all three factors could be correctly predicted at better than 81% success rate when using all the objects (measurements) to build the classification models. When averaging within individuals, sex had the best classification rate of 94% compared to 83% for age and 61% for region of origin. The relatively poor performance for the classification by region of origin likely arises from the weakness of this factor relative to other compounding factors like age and sex, and the fact that origin was separated into three groups instead of two groups for the age and sex factors.

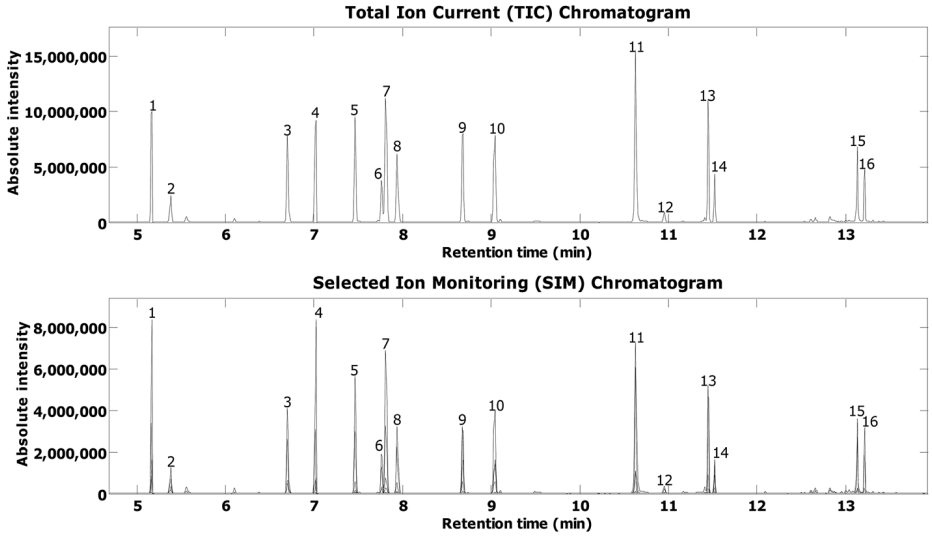

Fig. 1 A typical GC/MS profile for the analysis of human scalp hair keratin protein after derivatization with BSTFA at 100 °C for 30 min. SIM shows signals for target and reference ions that were described in GC/MS analysis step of the experimental section. Peaks: (1) Ala; (2) Gly2*; (3) Val; (4) Nor; (5) Leu; (6) Ile; (7) Pro; (8) Gly3**; (9) Ser; (10) Thr; (11) Asp; (12) Cys; (13) Glu; (14) Phe; (15) Lys; (16) Tyr, *di-trimethylsilyl derivative, ** tri-trimethylsilyl derivative.

Table 2 The FuRES leave-one-out cross-validation classification rates of sex, age group, and region of origin using each amino acid profile, averaged profiles within digestions, and averaged profiles within individuals

	Correct classification rate		
Factor	Profiles ($n = 528$)	Average within trials ($n = 177$)	Average within individuals ($n = 64$)
Sex	84%	85%	94%
Age group	88%	80%	83%
Region of origin	81%	68%	61%

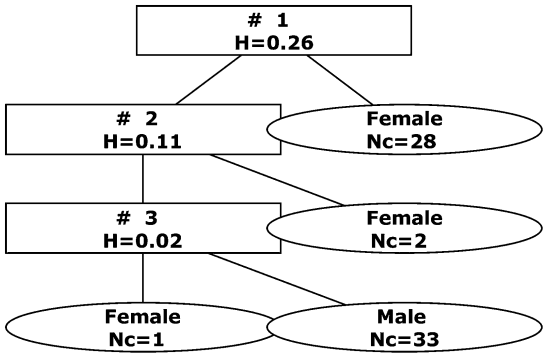

Fig. 2 FuRES sex classification tree from the average profiles of 14 amino acids in the hair for 64 individuals. This tree yielded a 94% prediction rate. N_c is number of individuals sorted to each node by the rules in the tree. H is the fuzzy entropy of classification for each rule and the rule number is given in each rectangle. The circular nodes are terminal and designate the class.

Sex classification

A FuRES classification tree is the collection of rules used to separate all the profiles. The rules are obtained by minimizing the entropy of classification (H). The root/trunk of the tree (typically shown with the root at the top) contains the most general rule. Moving down the tree from the root to the leaves at each branch the rules gets more specific (*i.e.*, they separate objects that are closer together in the data space). Fig. 2 is a FuRES classification for sex using the average profiles for individuals. This three-rule tree yielded a 94% classification rate, whereas Panayiotou[24] reported 80% success rate of fuzzy clustering for sex using FT-IR.

Fig. 3 is the ROC curve that was generated to identify the sex of the individuals. The true-positive rate or sensitivity is plotted on the ordinate as a function of the false positive rate on the abscissa. The classification threshold decreases as the curve progress from left to right. Accuracy of the classifier is measured by the area under the curve (AUC). The ideal diagnostic test is one that reaches the upper left corner of the graph with the AUC equal to 1.[56] The area under the curve of Fig. 3 is 0.95 indicating very good performance, especially for a classifier that is generalizing for individuals, which indicates that hair samples from new individuals are likely to be successfully classified.

The loadings of the FuRES discriminant for rule #1 for sex classification is given in Fig. 4, which separates the subjects into thirty-three males and three females on one side and twenty-eight females on the other side. Two more rules (loading scores not shown) are required to distinguish the three remaining females form the twenty-eight males. The positive peaks in Fig. 4 correspond to amino acids that are loaded higher for females than males (Val, Pro, Ser, Glu, and Phe), and therefore more abundant in females than males. Negative peaks correspond to amino acids that are loaded heavier for males than females (Ala, Gly, Leu, Ile, Thr, Asp, Cys, Lys, and Tyr), and therefore more abundant in males than females. If the 95% confidence intervals of the error bars cross the zero line, the amino acid is not significant for the classification. Phe is clearly

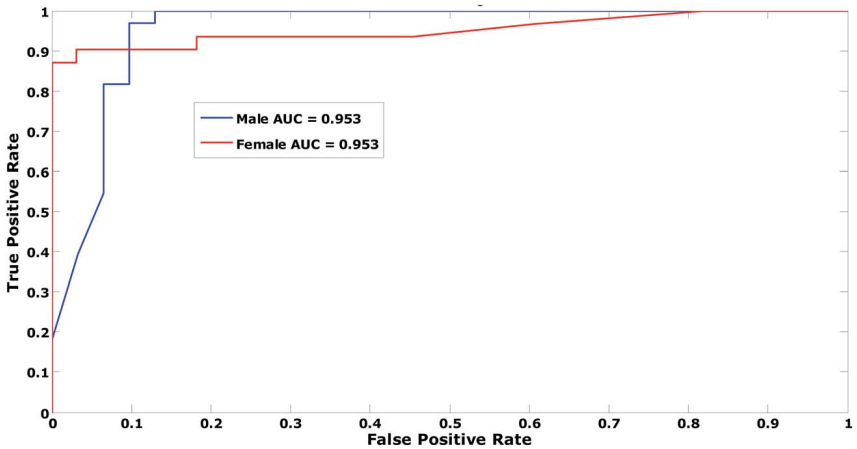

Fig. 3 Receiver operating characteristic (ROC) curve for identifying the sex of 64 subjects.

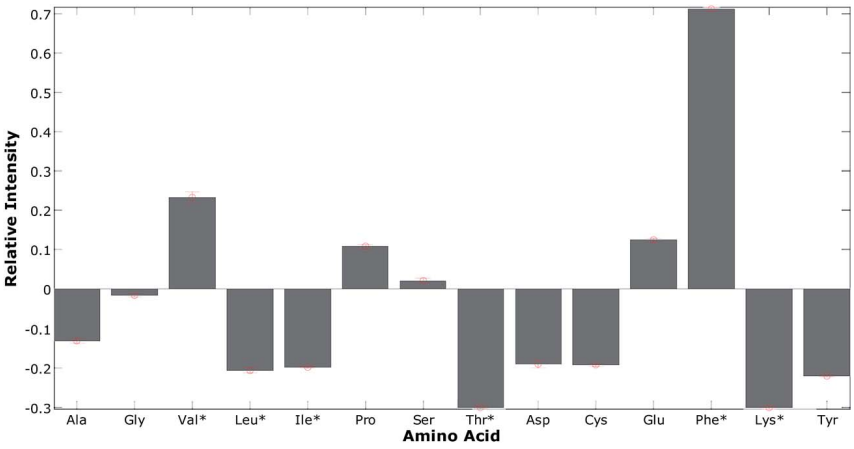

Fig. 4 Variable loadings scores of the 14 amino acids of sex classification. *Essential amino acids.

the most prominent feature and is found to be present in higher abundance in females relative to males. In contrast, males typically had higher concentrations of Thr and Lys. Although less prominent, males were found to have higher Cys content than females, in agreement with the results of others.[18,22]

As expected from the FuRES results, a significant difference ($p < 0.01$) was found between within-person variances and between-person variances for the two sex groups after ANOVA-PCA analysis. The impact of sex and source of variation between the groups was evaluated by comparing the concentration (mg amino acid/mg hair) of each amino acid in both male and female groups (Table 3). In Table 3, the essential amino acids (Phe and Val) and the non-essential amino acid, Pro, have significant differences for the ratio of their between-group to within-group variance to variance (*i.e.* F statistic). Therefore, these amino acids should be highly discriminating for sex. The loading scores in FuRES also indicate that these amino acids are heavily loaded and therefore very important for classifying sex. Consequently, the multivariate FuRES and the univariate ANOVA results agree.

The results of the ANOVA and FuRES agree with the results obtained by Panayiotou,[24] who reported that Phe and Glu amino acids correlate with the proportion of α-helical structure relative to β-sheet structure in hair. Panayiotou found that females have more α-helix structure in their hair than males and therefore have more Glu and Phe. In this present study, Phe has the most significant difference for the within-group variance to between-group variance and the heaviest loaded peak in the female group. In our population, this result could also be explained by ultraviolet radiation (UV) that has the ability to decrease the Phe amino acid of hair *via* photo-degradation upon exposure to sunlight.[9] A response in the survey indicated that all of the females subjects in the present study covered their hair with a Hijab that presumably blocks UV from the sun and prevents photodegradation of Phe. This result has also been supported by Nogueira *et al.*,[24] who reported that the aromatic amino acids such as Phe were degraded upon exposure to sunlight.

Table 3 One-way ANOVA results of the effect of sex group on the amino acid concentration in hair between and within two sex groups. Values reported as mg amino acid/mg hair and were averaged within each person before ANOVA

Amino acid	Male mean ± 95% CI ($n = 33$)	Female mean ± 95% CI ($n = 31$)	Mean squares of variance (between groups)	Mean squares of variance (within groups)	F-value	p-value
Ala	12.5 ± 2.0	12.4 ± 2.9	0.6	46	0.01	0.91
Gly	3.6 ± 0.6	4.1 ± 1	3.7	4.8	0.8	0.38
Val[a]	6.8 ± 0.8	9.8 ± 2.1	140	18	7.7	0.007
Leu[a]	15.2 ± 1.6	15.8 ± 3.2	5.5	47	0.1	0.73
Ile[a]	4.4 ± 0.6	4.8 ± 0.9	3.5	4.4	0.8	0.38
Pro	9.8 ± 1.0	12.7 ± 2.5	133	26	5.2	0.03
Ser	1.8 ± 0.7	1.8 ± 0.6	0.1	3.8	0.01	0.91
Thr[a]	2.8 ± 0.7	2.3 ± 0.5	4.5	2.8	1.6	0.21
Asp	9 ± 1.4	10.3 ± 2.2	30	25	1.2	0.28
Cys	3.1 ± 1.0	2.8 ± 1.0	2.0	8.2	0.2	0.62
Glu	8.5 ± 1.6	9.3 ± 1.6	10	19	0.5	0.46
Phe[a]	2.3 ± 0.2	3.4 ± 0.6	18	1.3	13.7	0.0005
Lys[a]	7.7 ± 1.1	7.8 ± 1.3	0.3	11	0.03	0.87
Tyr	6.0 ± 0.6	5.7 ± 0.7	1.8	3.4	0.5	0.47

[a] Essential amino acid.

Conflicting results were found with relation to the impact of sex on the contents of Cys in human hair. Clay *et al.*[18] and Rutherford and Hawk[22] found significantly more Cys in the hair of males compared to females. In contrast, Wilson and Lewis[21] found no consistent relation between the Cys and sex. In the present study, no such difference was found in Cys concentration between males and females.

Age classification

To examine the effect of age on the amino acid content, the subjects were split into two arbitrary age groups: group A comprised subjects with ages from 1 to 49 years and group B of elder people with ages ranging from 50 to 77 years. Classification of subject groups was modeled by FuRES. The two groups were separated using a four-rule tree. For example, the second rule separates 47 subjects of group A and one subject of group B. The fourth rule separates 9 subjects of group B and only one subject of group A. Fig. 5 is a FuRES classification tree demonstrating the classification model. The correct classification rate for the model was 83%, as assessed using the leave-one-individual-out cross-validation.

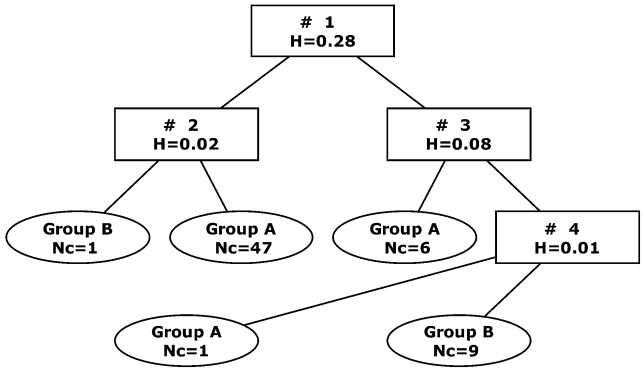

Fig. 5 FuRES age group classification tree from the average amino acid profiles of 14 amino acids in the hair from 64 subjects (*n* of 9 for each average) with an 83% classification rate. *H* is the entropy of the classification rule and the number of rules required to build the tree are given in the rectangles. The number of subjects and name of the group are written in each leaf node.

The area under the curve in Fig. 6 is 0.85, representing a very good test that could effectively distinguish age group A from age group B. The ROC is also supported by the similar FuRES classification rate of 83%.

A significant difference of ($p = 0.03$) was found between within-person variances and between-person variances for the two groups after ANOVA-PCA analysis. The sources of variation among the groups were evaluated by comparing the average concentrations (mg amino acid/mg hair) of each amino acid from the three regions are presented in Table 4. ANOVA revealed that the non-essential amino acids (Gly and Ala) and essential amino acid Ile appear to be the most discriminating between age groups; *F* values were 5.4, 5.6 and 7.9, respectively.

The findings indicated that the concentration of Ala, Gly and Ile are significantly higher in the younger group (A) than the older group (B). Asp was significantly lower in the older group (B) at the confidence level of 87%, which is supported by the finding by Rieck.[8] Wilson and Lewis[21] and Clay *et al.*[18] found no relation between Cys content of hair and age, which is also supported here. Sex did not correlate significantly with age.

No statistically significant difference was found between age groups for Ser. Rieck[8] studied the age-dependent hair of three sisters between the ages 1 to 24 years and reported increasing Ser concentration with increasing age. However, our age groups reported here span a much larger range of years, so the within-group variation for the 'young' group A might obscure potential differences between the older group B.

Whereas dietary habits could be an important variable that affect the characterization of the essential amino acids in human hair, genetic or phenotypic factors are more likely to influence the non-essential amino acids. Relatively significant variation in Thr and Ile between the two age groups could be attributed to their dietary habits.

Region of origin classification

Fig. 7 is a FuRES classification tree demonstrating the classification model of the region of origin. The classification rate for

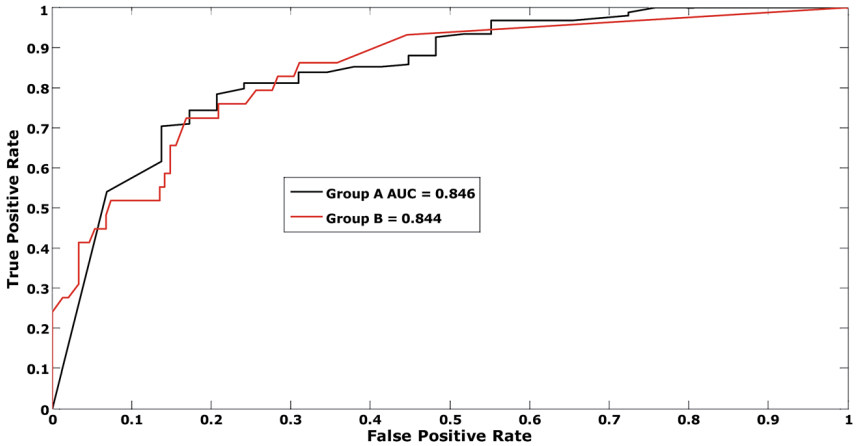

Fig. 6 Receiver operating characteristic (ROC) curve for identifying age group (1–49 years or 50–77 years) of the 64 subjects.

Table 4 One-way ANOVA results of the effect of age group on the amino acid concentration in hair between and within two age groups. Values reported as mg amino acid/mg hair and were averaged within each person before ANOVA

Amino acid	Group A mean ± 95% CI ($n = 54$)	Group B mean ± 95% CI ($n = 10$)	Mean squares of variance (between groups)	Mean squares of variance (within groups)	F-value	p-value
Ala	13.3 ± 1.9	7.6 ± 1.6	236	42	5.6	0.02
Gly	4.2 ± 0.6	2.3 ± 0.6	24	4.4	5.4	0.02
Val[a]	8.3 ± 1.1	7.6 ± 4.3	9.1	20	0.4	0.51
Leu[a]	15.8 ± 1.9	13.9 ± 4.5	48	47	1.0	0.31
Ile[a]	4.9 ± 0.6	2.9 ± 0.6	32	4.0	7.9	0.01
Pro	11.3 ± 1.5	10.5 ± 3	24	27	0.9	0.35
Ser	1.8 ± 0.5	1.9 ± 1.4	0.1	3.8	0.03	0.86
Thr[a]	2.7 ± 0.5	2.0 ± 0.6	3.0	2.9	1.0	0.31
Asp	10.0 ± 1.5	7.6 ± 1.2	58	25	2.4	0.13
Cys	3.1 ± 0.8	2.3 ± 1.6	4.0	8.1	0.5	0.49
Glu	9.1 ± 1.3	8.1 ± 1.5	12	19	0.7	0.42
Phe[a]	2.9 ± 0.4	2.6 ± 0.5	1.5	1.6	0.9	0.34
Lys[a]	7.8 ± 1.0	7.6 ± 1.4	1.1	11	0.1	0.75
Tyr	5.9 ± 0.5	5.7 ± 0.6	0.2	3.4	0.1	0.80

[a] Essential amino acid.

the model was 61% after the leave-one-individual-out cross-validation. The three regions of origin were separated using a nine-rule tree. For example, the second rule separates 22 subjects of the mountain region to the right and one subject from the desert to the left side of the tree.

In the ROC curves, the AUCs are 0.72, 0.71, and 0.83 for the mountain, the valley, and the desert, respectively (Fig. 8). The classification according to region of origin is not nearly as accurate as by sex or by age group, but it also is much better than random which would be indicated by an AUC of 0.5. Region of origin is less effective than age and sex in discriminating the hairs of individuals. A larger and more-equal sample size of each region could improve the classifier.

ANOVA-PCA of the 14 amino acids gave a significant difference among the means of the three groups of amino acid profiles for region of origin (p of 0.03). The sources of variation among the groups were evaluated by comparing the average concentrations (mg amino acid/mg hair) of each amino acid from the three regions (Table 5). The amino acids Cys, Ile, Glu and Gly have the most significant differences for between-group variance to within-group variance. Therefore, they may be important for the classification of subjects according to their region of origin.

Cys is a semi-essential amino acid, which means that it can be biosynthesized in humans and furnished by diets.[9] Cys is found in most high-protein foods, including meat, eggs, dairy products.[17] Higher levels of Cys in the hair samples from desert region compared to mountain and valley region is an expected observation, which may be attributed to the preference of meat, egg, and dairy product by the subjects living in the deserts of

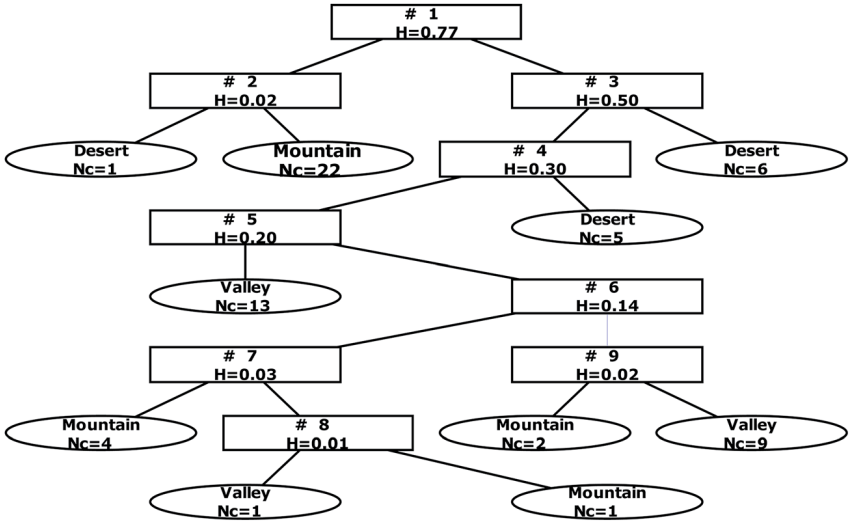

Fig. 7 FuRES region of origin classification tree from the relative abundance of 14 amino acids in the hair of 64 subjects (n of 9 for each subject) with a 61% classification rate.

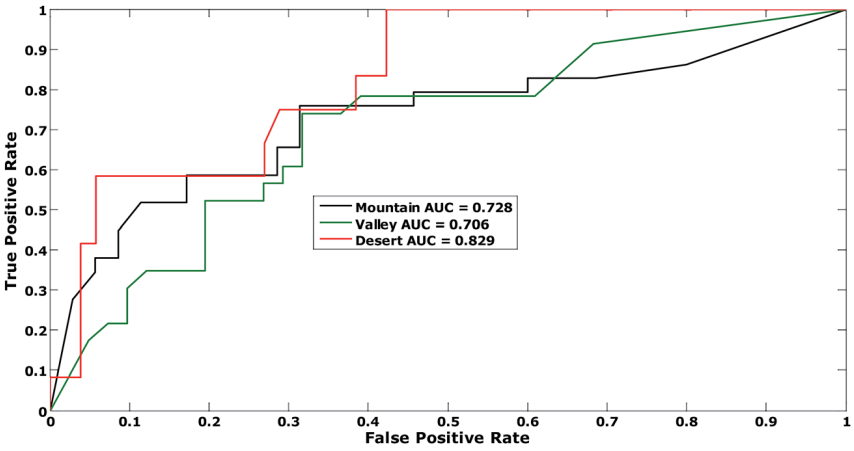

Fig. 8 Receiver operating characteristic (ROC) curve for identifying the region of origin of 64 subjects.

Table 5 One-way ANOVA results of the effect of origin on the amino acid concentration in hair between and within three regions of origin. Values reported as mg amino acid/mg hair and were averaged within each person before ANOVA

Amino acid	Mountain mean ± 95% CI ($n = 29$)	Valley mean ± 95% CI ($n = 23$)	Desert mean ± 95% CI ($n = 12$)	Mean squares of variance (between groups)	Mean squares of variance (within groups)	F-value	p-value
Ala	11.1 ± 0.4	14.4 ± 2.7	12.0 ± 3.0	68	45	1.5	0.22
Gly	3.0 ± 0.2	5.0 ± 0.7	3.6 ± 1.2	25	4.1	6.1	0.02
Val[a]	8.8 ± 0.5	8.2 ± 1.3	6.8 ± 1.6	16	20	0.8	0.37
Leu[a]	16.2 ± 0.7	15.7 ± 2.1	13.6 ± 2.6	29	47	0.6	0.44
Ile[a]	4.0 ± 0.2	5.5 ± 0.7	4.3 ± 1.3	14	4.1	3.5	0.06
Pro	12.1 ± 0.5	10.6 ± 1.7	9.9 ± 2.5	26	27	1.0	0.33
Ser	2.2 ± 0.3	1.6 ± 0.4	1.1 ± 1.7	5.4	3.7	1.5	0.23
Thr[a]	2.5 ± 0.2	2.9 ± 0.4	2.0 ± 0.8	3.4	2.8	1.2	0.28
Asp	10.7 ± 0.6	9.4 ± 1.5	7.4 ± 1.3	46	25	1.9	0.17
Cys	4.1 ± 0.4	2.5 ± 0.4	1.2 ± 0.6	39	7.0	5.6	0.02
Glu	7.9 ± 0.4	10.8 ± 1.1	7.8 ± 3.4	61	17	3.5	0.07
Phe[a]	2.9 ± 0.1	3.0 ± 0.4	2.3 ± 0.5	2.2	1.6	1.4	0.25
Lys[a]	8.2 ± 0.4	7.5 ± 1.0	7.1 ± 1.2	5.6	11	0.5	0.48
Tyr	6.4 ± 0.2	5.4 ± 0.4	5.3 ± 0.4	8.1	3.2	2.5	0.12

[a] Essential amino acids.

Jordan. This outcome is compatible with the findings of Takanohashi,[17] who observed that subjects who drink plenty of milk have higher Cys content in their hair than those subjects who consume less milk. In addition, the Bedouin, who reside in the desert region, are exposed to the sun more than other groups because they live in tents and tend sheep. They reported in the study that their scalp hair was exposed to the sunlight more than 4 h per day. Ultraviolet radiation (UV) has the ability to alter the amino composition of hair in photo-degradation upon exposure to sunlight.[9] Cyt contains two Cys amino acids that are linked by a disulfide bond.[3] Breaking the disulfide bond through photochemical degradation of Cyt produces more Cys residues.[9,63]

The abundance of Phe and Tyr amino acids are significantly lower in the desert group and this result is similar to the finding of Nogueira *et al.*,[63] who reported that the aromatic amino acids such as Tyr and Phe were degraded upon exposure to sunlight. The significant variation between the three groups in essential amino acids (Ile and Phe) plays an important role in the classification of region of origin. It may be attributed to a combination of the different dietary habits in the three different regions and sunlight exposure related to lifestyle habits.

Conclusions

The ability to classify the hair of subjects into age, sex, and region of origin using profiles of 14 amino acids obtained from hair samples has been demonstrated for a large-scale and rigorous evaluation. The classification rules extracted from FuRES were consistent with the results of ANOVA, and for the most part other reports in the literature. The findings of this study provide good evidence for the feasibility of amino acids to distinguish between subjects or groups of subjects. The correct classification rate was 94% for sex, 83% for age, and 61% for

region of origin after the leave-one-individual-out cross-validation, which is a general and rigorous evaluation. The present study is expected to help guide future research to assist forensic professionals to discrimination or link questioned hair and known samples from suspects or victims. Future analyses could provide investigative leads to law enforcement by predicting the sex, age group, and region of origin of the donor of an unknown hair sample found at a crime scene.

Profiling amino acids in human hair is just one way of examining hair. It should be emphasized that this technique alone is unlikely to lead to individualization to the extent that nuclear DNA can. The techniques developed through this work could complement the current methods of hair analysis, which include physical examinations and genomic or mitochondrial DNA analysis. The relative abundance of amino acids in hair may provide investigative leads when a suspect or victim's DNA profile is not in a database. Additional studies could be carried out, firstly to decrease the acid hydrolysis time and secondly to develop methods to enable analysis of more than 14 amino acids, perhaps through HPLC and without the need for derivatization.

Acknowledgements

The Ohio University Center for Intelligent Chemical Instrumentation at Department of Chemistry and Biochemistry are thanked for travel support. AHBR acknowledges financial support from the Jordan University of Science and Technology. Xinyi Wang and Dr Zhengfang Wang are acknowledged for their help. GPJ acknowledges support by Grant no. 2013-DN-BX-K007 awarded by the National Institute of Justice (NIJ), Office of Justice Programs (OJP), US Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect the views of the Department of Justice.

References

1 H. Sachs, *Forensic Sci. Int.*, 1997, **84**, 7–16.
2 A. S. Wilson, H. I. Dodson, R. C. Janaway, A. M. Pollard and D. J. Tobin, *Br. J. Dermatol.*, 2007, **157**(3), 450–457.
3 I. Kempson and E. Lombi, *Chem. Soc. Rev.*, 2011, **40**, 3915–3940.
4 R. Wennig, *Forensic Sci. Int.*, 2000, **107**(1–3), 5–12.
5 G. Loussouarn, C. El Rawadi and G. Genain, *Int. J. Dermatol.*, 2005, **44**, 6–9.
6 K. Morioka, *Exp. Dermatol.*, 2009, **18**(7), 577–582.
7 B. Bhushan, *Biophysics of Human Hair: Biological and Medical Physics*, Springer, New York, 2010.
8 W. Rieck, *Arch. Gerontol. Geriatr.*, 1997, **25**(1), 59–71.
9 C. Robbins, *Chemical and Physical Behavior of Human Hair*, Springer, New York, 2002.
10 M. O. Han, J. A. Chun, W. H. Lee, J. W. Lee and C. Chung, *J. Cosmet. Sci.*, 2007, **58**(5), 527–534.
11 G. P. Jackson, Y. An, K. I. Konstantynova and A. H. B. Rashaid, *Sci. Justice*, 2014, DOI: 10.1016/j.scijus.2014.07.002.
12 V. A. Boumba, K. S. Ziavrou and T. Vougiouklakis, *Int. J. Toxicol.*, 2006, **25**(3), 143–163.
13 M. R. Wilson, D. Polansky, J. Butler, J. A. DiZinno, J. Replogle and B. Budowle, *Biotechniques*, 1995, **18**(4), 662–669.
14 C. A. Linch, S. L. Smith and J. A. Prahlow, *J. Forensic Sci.*, 1998, **43**, 305–314.
15 A. Hellmann, U. Rohleder, H. Schmitter and M. Wittig, *J. Leg. Med.*, 2001, **114**(4–5), 269–273.
16 J. M. Taupin, *Sci. Justice*, 2004, **44**, 95–100.
17 T. Takanohashi, *Pediatr. Int.*, 1961, **4**(1), 7–13.
18 R. C. Clay, K. Cook and J. I. Routh, *J. Am. Chem. Soc.*, 1940, **62**(10), 2709–2710.
19 A. Bertazzo, M. Biasiolo, C. V. Costa, E. Cardin de Stefani and G. Allegri, *Farmaco*, 2000, **55**(8), 521–525.
20 G. Allegri, C. Costa, M. Biasiolo and R. Arban, *Adv. Exp. Med. Biol.*, 1991, **294**, 467–470.
21 R. H. Wilson and H. N. Lewis, *J. Biol. Chem.*, 1927, **73**, 543–553.
22 T. A. Rutherford and P. B. Hawk, *J. Biol. Chem.*, 1907, **3**, 459–489.
23 J. Menkarr, L. Wolfram and I. Mao, *J. Soc. Cosmet. Chem.*, 1966, **17**, 769–787.
24 H. Panayiotou, Ph.D. dissertation, Queensland University of Technology., 2004.
25 D. H. Lightbody and H. B. Lewis, *J. Biol. Chem.*, 1929, **82**, 485–497.
26 E. V. Heard and H. B. Lewis, *J. Biol. Chem.*, 1938, **123**, 203–210.
27 C. R. Robbins and C. H. Kelly, *Text. Res. J.*, 1970, **40**, 891–896.
28 A. H. B. Rashaid, G. P. Jackson and P. B. Harrington, *Enliven: Bio Anal Techniques*, 2014, **1**(1), 1–12.
29 R. W. Zumwalt, J. S. Absheer, F. E. Kaiser and C. W. Cherke, *J.–Assoc. Off. Anal. Chem.*, 1987, **70**(1), 147–151.
30 A. J. Darragh and P. J. Moughan, *J.–Assoc. Off. Anal. Chem.*, 2005, **88**(3), 888–893.
31 R. D. Phillips, *J. Food Sci.*, 1983, **48**, 284–285.
32 D. L. Stalling, C. W. Gehrke and R. W. Zumwalt, *Biochem. Biophys. Res. Commun.*, 1968, **31**(4), 616–622.
33 I. Molnár-Perl and Z. F. Katona, *Chromatographia*, 2000, **51**, S228–S236.
34 C. W. Gehrke, H. Nakamato and R. W. Zumwait, *J. Chromatogr.*, 1969, **45**, 24–51.
35 C. W. Gehrke and K. Leimer, *J. Chromatogr.*, 1970, **53**(2), 201–208.
36 M. Friedman, *J. Agric. Food Chem.*, 2004, **52**, 385–406.
37 K. Robards, P. R. Haddad and P. E. Jackson, *Principles and Practice of Modern Chromatographic Methods*, Academic Press, San Diego, 1994.
38 J. Miller, *Chromatography: Concepts and Contrasts*, John Wiley & Sons Inc., Hoboken, 2005.
39 Y. An, Z. Schwartz and G. P. Jackson, *Rapid Commun. Mass Spectrom.*, 2013, **27**, 1481–1489.
40 M. P. Bartolomeo and F. Maisano, *J. Biomol. Tech.*, 2006, **17**(2), 131–137.

41 J. Gross, *Mass Spectrometry*, Springer, New York, 2004.
42 X. Shen, C. Deng, B. Wang and L. Dong, *Anal. Bioanal. Chem.*, 2006, **384**(4), 931–938.
43 C. H. Deng, X. Y. Yin, L. J. Zhang and X. M. Zhang, *Rapid Commun. Mass Spectrom.*, 2005, **19**, 2227–2234.
44 T. G. Sobolevsky, A. I. Revelsky, B. Miller, V. Oriedo, E. S. Chernetsova and I. A. Revelsky, *J. Sep. Sci.*, 2003, **26**, 1474–1478.
45 C. Vidal-madjar and G. Guiochon, *Sep. Purif. Rev.*, 1973, **2**, 1–125.
46 P. B. Harrington, *J. Chemom.*, 1991, **5**, 467–486.
47 P. B. Harrington, N. E. Vieira, P. Chen, J. Espinoza, J. K. Nien, R. Romero and A. L. Yergey, *Chemom. Intell. Lab. Syst.*, 2006, **82**(1–2), 283–293.
48 P. B. Harrington, C. Laurent, D. F. Levinson, P. Levitt and S. P. Markey, *Anal. Chim. Acta*, 2007, **599**(2), 219–231.
49 P. Rearden, P. B. Harrington, J. J. Karnes and C. E. Bunker, *Anal. Chem.*, 2007, **79**(4), 1485–1491.
50 Z. F. Xu, C. E. Bunker and P. D. Harrington, *Appl. Spectrosc.*, 2010, **64**(11), 1251–1258.
51 X. B. Sun, C. M. Zimmermann, G. P. Jackson, C. E. Bunker and P. B. Harrington, *Talanta*, 2011, **83**(4), 1260–1268.
52 W. Y. Lu, J. G. Rankin, A. Bondra, C. Trader, A. Heeren and P. B. Harrington, *Forensic Sci. Int.*, 2012, **220**(1–3), 210–218.
53 F. Yang, J. Tian, Y. H. Xiang, Z. Y. Zhang and P. B. Harrington, *Cancer Epidemiol.*, 2012, **36**(3), 317–323.
54 Z. F. Wang and P. D. Harrington, *Anal. Bioanal. Chem.*, 2013, **405**(28), 9219–9234.
55 J. R. Wang, Z. Y. Zhang, Z. W. Zhang, Y. H. Xiang and P. B. Harrington, *Anal. Methods*, 2014, **6**(19), 7695–7702.
56 J. A. Hanley and B. J. McNeil, *Radiology*, 1982, **143**, 29–36.
57 A. Gelman, *Ann. Stat.*, 2005, **33**, 1–53.
58 J. Pallant, *SPSS Survival Manual*, Open University Press, Maidenhead, 3rd edn, 2007.
59 E. R. Malinowski, *Factor Analysis in Chemistry*, John Wiley & Sons, Inc., New York, 3rd edn, 2002.
60 P. B. Harrington, N. E. Vieira, P. Chen, J. Espinoza, J. K. Nien, R. Romero and A. L. Yergey, *Chemom. Intell. Lab. Syst.*, 2005, **82**, 283–293.
61 D. L. Luthria, L. Z. Lin, R. J. Robbins, J. W. Finley, G. S. Banuelos and J. M. Harnly, *J. Agric. Food Chem.*, 2008, **56**(21), 9819–9827.
62 D. L. Luthria, S. Mukhopadhyay, R. J. Robbins, J. W. Finley, G. S. Banuelos and J. M. Harnly, *J. Agric. Food Chem.*, 2008, **56**(14), 5457–5462.
63 A. C. S. Nogueira, L. E. Dicelio and I. Joekes, *Photochem. Photobiol. Sci.*, 2006, **5**, 165–169.