Protein extraction from human anagen head hairs one millimeter or less in total length

Traci L Carlson¹, Mehdi Moini², Brian A Eckenrode³, Brent M Allred¹ & Joseph Donfack³

¹Visiting Scientist Program, Oak Ridge Institute for Science & Education, Oak Ridge, TN 37830, USA; ²Department of Forensic Sciences, George Washington University, 2100 Foxhall Road, NW Washington, DC 20007, USA ³Counterterrorism & Forensic Science Research Unit, Federal Bureau of Investigation Laboratory Division, 2501 Investigation Parkway, Quantico, VA 22135, USA

BioTechniques 64:170–176 (April 2018) 10.2144/btn-2018-2004

Keywords: hair; nano-flow liquid chromatography tandem-mass spectrometry; nLC-MS/MS; proteome; proteomics; protein extraction; protein/peptide profiling

A simple method for extracting protein from human anagen (i.e., actively growing hair stage) head hairs was developed in this study for cases of limited sample availability and/or studies of specific micro-features within a hair. The distinct feature segments of the hair from one donor were divided lengthwise (i.e., each of ~200–400 μm) and then pooled for three individual hairs to form a total of eight composite hair samples (i.e., each of ~1 mm or less in total length). The proteins were extracted, digested using trypsin, and characterized via nano-flow liquid chromatography tandem-mass spectrometry (nLC-MS/MS). A total of 63 proteins were identified from all eight protein samples analyzed of which 60% were keratin and keratin-associated proteins. The major hair keratins identified are consistent with previous studies using fluorescence *in situ* hybridization and nLC-MS/MS while requiring over 400–8000-fold less sample. The protein extraction method from micro-sized human head hairs described in this study will enable proteomic analysis of biological evidence for cases of limited sample availability and will complement hair research. For example, research seeking to develop alternative non-DNA based techniques for comparing questioned to known hairs, and understanding the biochemistry of hair decomposition.

Over the last two decades proteomic analyses have engendered considerable interest for their various applications within a variety of research fields including immunology, microbiology, and forensic science [1]. While protein analysis instrumentation has continued to advance in terms of sensitivity, accuracy, reproducibility and resolving power, sample preparation procedures need to follow suit. Analyses of small-to-micro-sized samples, such as those obtained during biopsies, would enable proteomic expansion into fields such as forensic science, where evidence sample size is often limited.

A human hair shaft is mainly composed of proteins (65–90%) [2]. The remaining constituents are water, lipids, sugars, pigment (melanins), nucleic acids and trace elements. As such, proteins can be exploited to further the probative value of hair evidence. The genetic information preserved in hair proteins has the potential to complement microscopic hair analysis when DNA is compromised or degraded. This genetic information in the form of genetically variant peptides is being explored for its potential to associate hair back to a single individual [3] or for its use in discriminating hairs based on ethnicity [4]. However, despite its strong potential and obvious advantages, the protein/peptide profiling technique is not devoid of limitations. A primary limitation is related to the adequate hair length required for analysis. In general, very large amounts of human hair samples (i.e., 2–40 mg) have traditionally been used for protein/peptide profiling via LC-MS/MS [3–6]. As per Szabo *et al.* [7], 200 mm of hair weighs on average about 1 mg. This means that between 40 and 800 cm of hair were used in previous method development research studies. In many forensic cases, usually only a single hair is recovered as evidence, therefore this represents a significant limitation of the application of protein/peptide profiling to forensic casework. Additionally, protein extraction methods used in the studies referenced [3–5,8,9] employed

METHOD SUMMARY

The protein extraction procedure from a single hair shaft sample described here represents an important and critical step towards implementation of protein profiling into forensic casework. Protein profiling may complement both microscopic and mitochondrial DNA analyses of hair evidence, or be an alternative if these analyses aren't possible.

protein extraction techniques such as bead beating and grinding by mortar and pestle, which result in sample loss, are more susceptible to contamination and are not suitable for small samples. Furthermore, the techniques described in certain studies require multiple extractions to obtain adequate levels of protein for analysis [10,11]. Finally, other methods published use on column sample clean-up which may result in sample loss and significant increase in instrument runtime. Consequently, the aim of this work was to produce a simple method that does not require grinding, multiple extractions, or on column sample clean-up and is efficient and sensitive enough to analyze 1 mm or shorter hairs to be applicable in cases of limited sample size such as forensic casework or for targeting micro-features in hairs (e.g., shaft and root region).

Materials & methods

Hair samples & hair cuttings

Twenty-five hair samples were collected at the Federal Bureau of Investigation (FBI) Laboratory with consent from a healthy 33-year-old male volunteer of Caucasian descent. These hairs were collected in accordance with the Trace Evidence Unit's (TEU) recommended protocol for sampling suspect scalp hairs [12]. Briefly, five individual hair strands were pulled from the front, top, back, left and right sides of the volunteer's scalp. Hairs were pulled to increase the chances of getting a higher frequency of actively growing hairs with stretched roots (i.e., anagen hairs). Only one human subject was enrolled in this research study because the goal was to develop a method to extract proteins from micro-sized hair samples and not to assess variation in the hair proteome in the human population or between biogeographic populations.

Following transmitted light microscopy analysis, of the 25 hairs, eight anagen hairs whose roots had no visible soft tissue (e.g., skin cells) attached to them were selected for protein/peptide profiling. The root bulb itself was removed using a sterile standard razor blade. Hair segments were cut in the pre-keratinized (i.e., hair root) and the keratinized (i.e., hair shaft) region of the hair (Figure 1). Each hair segment was between 200 and 400 μm in length and homologous segments from three hairs were combined for each sample (Table

1). For example, root or shaft segments from three hairs were combined to form one root region or shaft region sample. All sample processing steps that occurred prior to protein digestion (described below) were performed in a HEPA laminar flow hood. Additionally, laboratory space and tools were routinely cleaned with bleach and alcohol, especially between contact with different hair samples to minimize sample cross contamination.

Protein extraction, in-solution trypsin digestion and sample clean-up

Protein extraction, in-solution trypsin digestion and sample clean-up were adapted from a previous protocol for extracting wool proteins [13]. For protein extraction, 20 μL protein extraction buffer composed of 50 mM dithiothreitol (DTT), 8 M urea, and 100 mM ammonium bicarbonate in nano-pure water were added to each hair sample. The sample was shaken (American DADE Rotator V, USA) at ~50 rpm for ~24 h at room temperature (~21°C). The protein extraction buffer and all other solutions added to the sample prior to trypsin digestion were filtered before use to minimize external protein contamination, utilizing the Amicon® Ultra 3K Centrifugal filter (Millipore, USA). Proteins were reduced with the addition of 2 μL of 100 mM DTT for 1 to 1.5 h and subsequently alkylated in the dark for 1 h. with 20 μL of 50 mM iodoacetamide. After alkylation, 20 μL of 100 mM DTT were again added to quench the reaction. The sample was then brought up to a final volume of 200 μL with 138 μL of 100 mM ammonium bicarbonate to dilute the total urea concentration to less than 2 M.

A trypsin solution was prepared by adding 858 μL of 1 mM hydrochloric acid (HCl) to a single tube of 20 μg lyophilized trypsin (Sigma-Aldrich, USA). On average 11 μL of 3 pmol/ μL trypsin in 1 mM HCl was added into each composite hair sample (i.e., each of ~ 1000 μm or less in total length). Additionally, 20 μL (10% total sample volume) of acetonitrile (ACN) was added and the digest was maintained at 37°C for 15 to 16 h in a water bath. To quench the digestion process, the solution was acidified to a pH less than 4 by adding approximately 5 μL of glacial acetic acid.

The peptide samples were then purified (i.e., desalted) and concentrated with C18 (e.g., Pierce™ 100 μL C18 Tip,

ThermoFisher Scientific, USA) solid phase extraction tips. Three solutions were used to perform the Pierce™ C18 tip clean-up including the wetting solution (50% ACN in nano-pure water), the equilibration/washing solution (0.1% Trifluoroacetic acid [TFA] in nano-pure water), and the elution solution (0.1% TFA, 80% ACN in nano-pure water).

The C18 bed was initially prepared by pipetting 20 μL of the wetting solution up and down 20 times to thoroughly wet the packing material. Next, equilibration of the packing material occurred by aspirating then dispensing to waste, 20 μL of the equilibration/wash solution, five times. Once the tip was equilibrated it was ready to receive the sample. Each sample was aspirated and dispensed approximately 50 times. To ensure that the entire sample (~240 μL) passed over the C18 packing material, dedicated transfer tubes were set up and 50 μL at a time of the sample were aspirated and then dispensed into the new tube. Four to five aspirations were needed to completely transfer the sample to the new tube.

The peptides, now bound to the C18 packing material in the tip, were washed by aspirating and dispensing to waste 20 μL of the equilibration/washing solution, five times. Finally the samples were eluted using the elution solution by aspirating 10 μL of the solution and dispensing it into a new PCR tube; this occurred three times.

The entire peptide purification procedure, excluding the wetting step, was repeated again utilizing the same tip resulting in a final eluant of 60 μL to which 1 μL of 0.5 pmol Angiotensin II (Sigma-Aldrich, USA) was added. The Angiotensin II was added to account for LC injection variability (if any) and mass spectrometer performance. After clean-up and elution, the samples were dried down in a Speedvac (ThermoSavant SPD1010, USA) and then reconstituted in 12 μL of the liquid chromatography (LC) starting mobile phase.

Sample blanks were prepared in parallel for every sample set and were taken through every step (i.e., protein extraction without a hair sample); including miming the cutting of a non-existent hair, to assess the level of contamination to which all samples may have been exposed.

Figure 1. Hair segments. Image of a hair showing the keratinized and pre-keratinized regions of the hair segmented for protein analysis. The shaft segments were obtained closer to the distal end of the pre-keratinized region.

Liquid chromatography and mass spectrometry

nLC-MS and proteomics analyses were performed at the Department of Forensic Sciences, at George Washington University (GWU). Tryptic peptides were separated by nanoflow liquid chromatography (nLC) using an Easy nLC 1000™ (Thermo Scientific, USA). Analytical separation was performed by first injecting 5 μ L of tryptic peptides onto a 0.75 μ m id x 2 cm long Acclaim PepMap (Thermo Scientific, USA) precolumn. The precolumn was connected to a 50 cm long, 0.75 μ m id., PEPMAP C18 LC column (Thermo Scientific, USA) using a zero dead-volume Tee. The other opening of the Tee was connected to the waste through a multiport valve within the nLC system. The precolumn was then washed with 20 μ L of solvent A (0.1% formic acid in water) while keeping the waste stream open. After the precolumn washing step, the multiport valve was switched to provide flow through the Tee and into the analytical column. Both the precolumn and the analytical column were at room temperature. The nLC mobile phase

composition was ramped from 95/5 to 50/50 (%water/% ACN in 0.1% Formic acid) in 25 min at a flow rate of 250 nL/min. The Easy nLC was interfaced to an LTQ Velos Orbitrap Elite™ mass spectrometer using Excalibur software V 3.0.3 and utilizing a nano electrospray ionization source (Thermo Scientific, USA). The spray voltage was set to +2.5 kV. Mass spectral analysis was performed using the data-dependent mode using one full scan (m/z 350–1600 at a resolution of 60,000 in profile mode) followed by 15 MS/MS scans using a 2 m/z isolation width and normalized collision energy of 25 excluding singly charged ions.

Mass spectrometry data analysis for protein identification

The resulting MS/MS spectra of peptides were processed for protein identification using Proteome Discoverer™ (PD, vs. 2.1) (Thermo Scientific, USA). All resulting spectra were searched in human protein databases (UniProt and National Center for Biotechnology Information's RefSeq non-redundant protein databases) using SEQUEST [14]. Search param-

eters included: full trypsin enzymatic activity, two missed cleavages, and peptides between the molecular weight of 350–5,000 mass tolerance at 20 ppm for precursor ions and 0.8 Da for fragment ions. Fixed modifications included: Asparagine deamidation, cysteine carbamidomethylation and methionine oxidation.

The list of proteins/peptides identified and associated statistics were exported from PD as a Microsoft Excel file into “R” statistics [15] for further statistical analysis. Peptides were filtered by retaining only those exhibiting a “high” confidence peptide-spectrum matching (PSM) as determined by PD. The “two-peptide” rule was applied for the identification of a protein [16]. In other words, proteins were identified if two or more peptides were available with at least one unique peptide not shared by other proteins.

Results & discussion

The main objective of this study was to develop a simple and efficient method suitable for hair protein extraction from limited starting material and to identify the extracted proteins. The mass spectra of peptides from all eight samples processed for protein identification resulted in 3409 peptides that were matched with high confidence PSMs including those redundantly identified. Only *Homo sapiens* peptides exhibiting high confidence PSMs were used in downstream statistical analyses. Because members of the same protein family may exhibit a certain percentage of amino acid sequence homology, of the 3,409 peptides identified, 1,507 (44%) were

Table 1. Number of proteins/peptides identified in each protein sample.

Index	Sample identification	Segment type	Total length (μ m)	Number of peptides	Number of unique peptides	Number of proteins
1	Sple53b	Root	972.8	506	211	51
2	Sple54b	Root	1100.8	360	169	44
3	Sple55b	Root	972.8	373	182	45
4	Sple56b	Shaft	947.2	438	181	43
5	Sple80b	Root	1075.2	391	182	43
6	Sple81b	Root	844.8	488	203	45
7	Sple82b	Root	921.6	402	173	40
8	Sple83b	Shaft	947.2	451	206	43
9	PBLK072016	Process Blank	0	3	6	3
10	PBLK072616	Process Blank	0	3	8	3

A protein sample is a composite of 3 segments of anagen head hairs from the two distinct regions shown in Figure 1, each varying between 200 and 400 μ m in length.

Table 2. List of keratins and keratin-associated proteins detected by nLC-MS/MS.

Index	Protein Symbol	Full protein name	Frequency (%)			Average number of unique peptides		Class	Tissue expression	Ref.
			Shaft	Root	Shaft&Root	Shaft	Root			
1	KRT1	Keratin, type II cytoskeletal 1	100	100	100	26.0	18.3	Keratin	Skin	[17]
2	KRT10	Keratin, type I cytoskeletal 10	100	100	100	19.5	16.2	Keratin	Skin	[17]
3	KRT2	Keratin, type II cytoskeletal 2 epidermal	100	100	100	18.5	14.5	Keratin	Skin	[17]
4	KRT32	Keratin, type I cuticular Ha2	100	100	100	3.5	2.2	Keratin	Hair	[17]
5	KRT33A	Keratin, type I cuticular Ha3-I	100	100	100	12.0	8.8	Keratin	Hair	[17]
6	KRT33B	Keratin, type I cuticular Ha3-II	100	100	100	7.0	6.7	Keratin	Hair	[17]
7	KRT34	Keratin, type I cuticular Ha4	100	100	100	14.0	18.3	Keratin	Hair	[17]
8	KRT35	Keratin, type I cuticular Ha5	100	100	100	8.5	8.0	Keratin	Hair	[17]
9	KRT5	Keratin, type II cytoskeletal 5	100	100	100	3.0	3.5	Keratin	Skin	[17]
10	KRT83	Keratin, type II cuticular Hb3	100	100	100	3.0	2.5	Keratin	Hair	[17]
11	KRT85	Keratin, type II cuticular Hb5	100	100	100	16.0	17.0	Keratin	Hair	[17]
12	KRT86	Keratin, type II cuticular Hb6	100	100	100	5.0	6.2	Keratin	Hair	[17]
13	KRT9	Keratin, type I cytoskeletal 9	100	100	100	15.5	14.2	Keratin	Skin	[17]
14	KRTAP11-1	Keratin-associated protein 11-1	100	100	100	3.5	3.7	KRTAP	Hair	[18]
15	KRTAP2-3	Keratin-associated protein 2-3	100	100	100	2.0	1.6	KRTAP	Hair	[18]
16	KRTAP3-1	Keratin-associated protein 3-1	100	100	100	2.5	2.7	KRTAP	Hair	[18]
17	KRTAP3-2	Keratin-associated protein 3-2	100	100	100	1.5	1.7	KRTAP	Hair	[18]
18	KRTAP3-3	Keratin-associated protein 3-3	100	100	100	1.0	1.3	KRTAP	Hair	[18]
19	KRTAP4-3	Keratin-associated protein 4-3	100	100	100	2.0	2.0	KRTAP	Hair	[18]
20	KRTAP4-4	Keratin-associated protein 4-4	100	100	100	2.5	2.3	KRTAP	Hair	[18]
21	KRTAP9-3	Keratin-associated protein 9-3	100	100	100	1.0	1.2	KRTAP	Hair	[18]
22	KRT39	Keratin, type I cytoskeletal 39	100	83.3	87.5	4.0	3.2	Keratin	Hair	[17]
23	KRT72	Keratin, type II cytoskeletal 72	100	83.3	87.5	0.0	2.0	Keratin	Hair	[17]
24	KRT82	Keratin, type II cuticular Hb2	100	83.3	87.5	3.0	2.6	Keratin	Hair	[17]
25	KRTAP4-11	Keratin-associated protein 4-11	100	83.3	87.5	0.0	2.0	KRTAP	Hair	[18]
26	KRT14	Keratin, type I cytoskeletal 14	100	66.7	75.0	1.5	2.4	Keratin	Skin	[17]
27	KRTAP24-1	Keratin-associated protein 24-1	100	66.7	75.0	1.0	1.0	KRTAP	Hair	[18]
28	KRT81	Keratin, type II cuticular Hb1	100	50.0	62.5	1.0	1.0	Keratin	Hair	[17]
29	KRTAP13-2	Keratin-associated protein 13-2	100	50.0	62.5	1.5	1.0	KRTAP	Hair	[18]
30	KRTAP19-5	Keratin-associated protein 19-5	100	50.0	62.5	2.0	2.0	KRTAP	Hair	[18]
31	KRT16	Keratin, type I cytoskeletal 16	50	50.0	50.0	1.0	2.7	Keratin	Skin	[17]
32	KRT40	Keratin, type I cytoskeletal 40	100	33.3	50.0	1.0	1.0	Keratin	Hair	[18]
33	KRTAP13-1	Keratin-associated protein 13-1	100	33.3	50.0	1.0	1.5	KRTAP	Hair	[18]
34	KRT6C	Keratin, type II cytoskeletal 6C	0	50.0	37.5	0.0	3.5	Keratin	Skin	[17]
35	KRTAP9-7	Keratin-associated protein 9-7	50.0	33.3	37.5	1.0	1.0	KRTAP	Hair	[18]
36	KPRP	Keratinocyte proline-rich protein	50.0	16.7	25.0	1.0	1.0	Keratin	Hair	[18]
37	KRT38	Keratin, type I cuticular Ha8	0	33.3	25.0	0.0	1.0	Keratin	Hair	[17]
38	KRT13	Keratin, type I cytoskeletal 13	0	16.7	12.5	0.0	1.0	Keratin	Skin	[17]

Proteins are sorted by their respective shaft and root frequencies (%) derived from a sample size of 8. Individual shaft and root frequencies (%) were derived from a sample size of 4. The column Ref. indicates the research that links the tissue expression to the proteins identified in this study.

found to be unique and predicted to have originated from 63 proteins (Tables 2 & 3).

Two box plots showing the distribution of percentage of sequence coverage for proteins identified in shaft and root segments are depicted in Figure 2. The variability in percentage

of sequence coverage in each sample group is very similar. For example, looking at the spread (i.e., length of the box or distance between the 1st quartile and the 3rd quartile), the spread in root segments is slightly greater (25.49%) than the spread in shaft segments (24.38%), suggesting a slightly greater variability

in percentage of sequence coverage in the latter group. The difference in spread of the sample groups was not statistically significant (the p-value from the Two-sample Kolmogorov–Smirnov test of sequence coverage distribution is 0.21). This conclusion is also supported by the approximately equal values of the

medians (i.e., 13.30% for the root and 13.78% for the shaft segments). There is no statistically significant difference in the two sample groups when comparing the median values (the p-value from the Mann–Whitney test for the difference in medians is 0.62).

Proteins identified are further organized into two main groups: keratins and non-keratins. As expected, keratins and keratin-associated proteins (KRTAPs) are the major human hair proteins detected in all hair segments (n = 24), accounting for 60% of the total number of proteins identified (Table 2). Earlier work seeking to assess the *in situ* expression and relative quantification levels of specific keratins in distinct layers/sections of the hair have traditionally used immunoassay techniques [19]. For example, Moll *et al.* [17] used immunofluorescence to characterize the expression of keratins in human hair roots. Recently, other techniques that do not require prior knowledge of the proteins expressed in

the target tissue and do not need labeled antibodies, such as mRNA profiling or protein/peptide profiling via LC-MS/MS, have been utilized [3,8,10–11]. In this study, 11 keratins known to be highly expressed in hairs that were observed at high frequency (i.e., detected in more than 70% of the samples analyzed) were: KRT32, KRT33A, KRT33B, KRT34, KRT35, KRT39, KRT72, KRT82, KRT83, KRT85, and KRT86. The high frequency hair keratins identified are in agreement with major hair proteins identified in previous research [3,4,11,17]. Of particular note, Rice *et al.* [11], did not detect KRT38 in hair shaft or nail samples; similar results were observed in this study using a relatively small sample set. KRT38 was not detected in shaft segments, but was present in root segments (Table 2). Even though the frequency of KRT72 was high (87.5%), its relative quantity was low in this study and absent in Rice *et al.*'s work [11]. In addition, the high frequency (i.e., ≥70%) of hair follicle-specific and/

or skin-specific keratins such as KRT1, KRT2, KRT9 and KRT14 suggests that hairs selected for protein extraction in this study were not completely free of follicular/scalp tissue. Furthermore, shaft segments were cut at the distal end of the pre-keratinized region, which may have contributed skin cells. Comparison of the protein profiles of shaft and root segments revealed extensive proteins overlap. In total 63 proteins were identified. Of these, 47 (~73%) were detected in both shaft and root segments and 16 (~27%) were only detected in root segments. Of the 16 proteins only detected in root segments, three were keratins and 13 were non-keratins (e.g., cytoplasmic, histones, membrane and junctional proteins) likely indicating a difference in physiological state between both sections of the hair – the hair root being metabolically active and the shaft metabolically inactive.

KRTAPs constitute another major component of the hair shaft known to provide additional mechanical strength

Table 3. List of non-keratin proteins detected by nLC-MS/MS.

Index	Protein Symbol	Full protein name	Frequency (%)			Average number of unique peptides	
			Shaft	Root	Shaft&Root	Shaft	Root
1	DSG4	Desmoglein-4	100	100	100	2.5	3.2
2	HRNR	Hornerin	100	100	100	3.0	2.7
3	JUP	Junction plakoglobin	100	100	100	2.0	2.7
4	SFN	14–3–3 protein sigma	100	100	100	2.0	4.3
5	VSIG8	V-set and immunoglobulin domain-containing protein 8	100	100	100	2.5	2.8
6	BLMH	Bleomycin hydrolase	100	83.3	87.5	1.0	1.0
7	DSP	Desmoplakin	50.0	100	87.5	7.0	5.8
8	TRIM29	Tripartite motif-containing protein 29	100	83.3	87.5	1.0	1.5
9	HSPB1	Heat shock protein beta-1	0	83.3	62.5	0.0	1.4
10	PKP1	Plakophilin-1	50.0	66.7	62.5	4.0	1.7
11	LGALS7	Galectin-7	50.0	50	50.0	1.0	1.3
12	PRR9	Proline-rich protein 9	0	66.7	50.0	0.0	1.3
13	DSG1	Desmoglein-1	50.0	33.3	37.5	1.0	1.0
14	EFHD1	EF-hand domain-containing protein D1	0	50	37.5	0.0	1.0
15	HIST1H1B	Histone H1.5	0	50.0	37.5	0.0	2.0
16	HIST1H2BK	Histone H2B type 1-K	0	50.0	37.5	0.0	1.0
17	LAP3	Cytosol aminopeptidase	0	50.0	37.5	0.0	3.0
18	CALML3	Calmodulin-like protein 3	0	33.3	25.0	0.0	1.0
19	FLG2	Filaggrin-2	50.0	16.7	25.0	1.0	1.0
20	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	0	33.3	25.0	0.0	1.0
21	RPL6	60S ribosomal protein L6	0	33.3	25.0	0.0	2.0
22	SELENBP1	Selenium-binding protein 1	0	33.3	25.0	0.0	1.0
23	TGM3	Protein-glutamine gamma-glutamyltransferase E	0	33.3	25.0	0.0	1.0
24	GSTP1	Glutathione S-transferase P	0	16.7	12.5	0.0	1.0
25	RPL24	60S ribosomal protein L24	0	16.7	12.5	0.0	1.0

Proteins are sorted by their respective shaft and root frequencies (%) derived from a sample size of 8. Individual shaft and root frequencies (%) were derived from a sample size of 4. All of the proteins are expressed in hair shafts as described in [8,10–11].

Figure 2. Protein coverage. Box plots illustrating the sequence coverage of proteins identified. Values on the y-axis represent the percentage of sequence coverage and the two segment types analyzed are listed on the x-axis. Box plots depict the median (center bold line) and the range (dashed bars) of the percentage of sequence coverage for all eight hair samples tested. Open circles represent outliers. The spread is defined as the length of the box or distance between the 1st quartile and the 3rd quartile.

by forming a network of disulfide cross-linking among keratin filaments [18]. KRTAPs belonging to eight KRTAP families (i.e., KRTAP-1 to 4, KRTAP-9, KRTAP-11, KRTAP-13, and KRTAP-19) were detected in root and shaft samples. Members of KRTAPs detected have also been observed using *in situ* hybridization techniques [18] and LC-MS/MS techniques [3,11].

Approximately 40% of the proteins identified in all hair segments were grouped as non-keratins or non-KRTAPs (Table 3). According to previous studies [10,20–21], there seems to be a strong correlation between the number of signature peptides (e.g., unique peptides) and the relative quantity of the protein. The higher the number of peptides, the higher the relative quantity of protein from which they are derived. Although peptide counting can be regarded as an approach to estimating relative protein abundance, it is by no means an accurate measure of protein absolute quantities. A comparison of relative quantities of shaft and root proteins is depicted in Tables 2 & 3. As can be seen, while the relative quantity of prominent hair proteins (e.g., KRT32, KRT33A and KRT33B, KRT34, KRT35, KRT85, and KRT86) is very high, the relative quantity of non-keratin proteins (e.g., SELENBP1, TGM3, GSTP1 and RPL24) is very low in this study (Tables 2 & 3). This is probably a reflection of the hair length used (i.e., ~1 mm or less). Of particular note, and consistent

with previous research using LC-MS/MS, cytoplasmic, membrane and junctional proteins such as DSG4, HRNR, JUP, SFN, VSIG8, DSP, JUP, PKP1 exhibited the highest quantities among the non-keratin proteins [8,11].

Conclusion

The protein extraction and sample clean-up procedures discussed in this manuscript for proteomics analysis of a small section of hair proved effective for identification of major proteins in hair. This procedure may now be used for various forensically relevant hair proteomics. Proteins are a valuable source of biological information that are commonly procured as evidence in the form of hair. Although the field of proteomics remains underutilized by the forensic science community, proteomics has the potential to play a critical role in our ability to support and advance hair examination. The method presented in this article has proven to be successful at extracting proteins from hairs ~1 mm or less in total length. Additionally the extraction and clean-up procedures presented here could be adapted for other protein sources and utilized by researchers from other disciplines whose limiting factor is small sample size and the inability to obtain meaningful protein profiles from those samples.

Acknowledgments

The authors would like to thank Eshwar Jagerdeo and Cynthia L. Morris-Kukoski

for their comments. This research was supported in part by an appointment to the Visiting Scientist Program at the Federal Bureau of Investigation Laboratory Division, administered by the Oak Ridge Institute for Science and Education, through an interagency agreement between the US Department of Energy and the FBI. This is publication number 16–38 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for information only and inclusion does not imply endorsement by the FBI or the U.S. Government. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the U.S. Government.

Author contributions

TL Carlson and BM Allred participated in the study as visiting scientists at the FBI Laboratory and were primarily involved in wet laboratory work. Carlson and Allred performed hair cuttings and optimized protein extraction, in-solution trypsin digestion and sample clean-up protocols. Additionally, TL Carlson led manuscript writing. M Moini and BA Eckenrode were primarily involved in wet laboratory work at The George Washington University. On several occasions, Eckenrode transported proteins samples, prepared by TL Carlson and BM Allred, from the FBI Laboratory to Moini's Laboratory at The George Washington University where all nano-flow liquid chromatography tandem-mass spectrometry (LC-MS/MS) works took place. M Moini and BA Eckenrode performed optimization and data acquisition of all LC-MS/MS experiments. J Donfack conceived the original idea and developed the research plan. J Donfack worked closely with TL Carlson and BM Allred during protein extraction protocol optimization at the FBI laboratory. J Donfack wrote R-Scripts which were used to performed LC-MS/MS data mining and interpretation.

Financial & competing interests disclosure

Work in Moini's lab is supported by the National Science Foundation under Grant Number CHE 1440849. The authors have no other relevant affiliations or financial involvement with any organi-

zation or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

References

- Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nat. Genet.* 33 Suppl. 311–323 (2003).
- Robbins C. The cell membrane complex: three related but different cellular cohesion components of mammalian hair fibers. *J. Cosmet. Sci.* 60, 437–465 (2009) and Robbins CR Chemical and Physical Behavior Of Human Hair. Springer Science + Business Media, LLC 3rd Edition (1994).
- Parker GJ, Leppert TDS, Anex JK. Demonstration of protein-based human identification using the hair shaft proteome. *PLoS ONE* 11, e0160653 (2016).
- Laatsch CN, Durbin-Johnson BP, Rocke DM Human hair shaft proteomic profiling: individual differences, site specificity and cuticle analysis. *PeerJ* 2, e506 (2014).
- Huntzinger M, Fairbanks D, Parker G. Use of proteins to obtain measure of identity in the absence of usable DNA. Proceedings of International Symposium on Human Identification, 22nd Annual Scientific Meeting, Washington, DC, October 3–6, Poster No 108 (2011).
- Parker G. Methods for conducting genetic analysis using protein polymorphisms. United States Patent Number US8877455 B2 (2014).
- Szabo S, Jaeger K, Fischer H, Tschachler E, Parson W, Eckhart L. *In situ* labeling of DNA reveals interindividual variation in nuclear DNA breakdown in hair and may be useful to predict success of forensic genotyping of hair. *Int. J. Legal Med.* 126, 63–70 (2012).
- Lee YJ, Rice RH, Lee YM. Proteome analysis of human hair shaft: from protein identification to posttranslational modification. *Mol. Cell Proteomics* 5, 789–800 (2006).
- Parker GJ, Anex DS, Leppert MF, Baird L, Matsunami N, Tepper L. Polymorphic Peptide Hair Project. Lawrence Livermore National Laboratory Report. Lawrence Livermore National Laboratory, CA, USA (2014).
- Rice RH. Proteomic analysis of hair shaft and nail plate. *J. Cosmet. Sci.* 62, 229–236 (2011).
- Rice RH, Xia Y, Alvarado RJ, Phinney BS. Proteomic analysis of human nail plate. *J. Proteome Res.* 9, 6752–6758 (2010).
- Deedrick WD, Koch S. Microscopy of Hair Part 1: A Practical Guide and Manual for Human Hairs Forensic Science Communications. US Federal Bureau of Investigation (2004).
- Araki N, Moini M. Age estimation of museum wool textiles from Ovis aries using deamidation rates utilizing matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 25, 3396–3400 (2011).
- Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5, 976–989 (1994).
- R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria ISBN 3–900051–07–0 (2009).
- Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii AP. Working Group on Publication Guidelines for, and D. Protein Identification. The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol. Cell Proteomics* 3, 531–533 (2004).
- Moll R, Divo M, Langbein L. The human keratins: biology and pathology. *Histochem. Cell Biol.* 129, 705–733 (2008).
- Shimomura Y, Ito M. Human hair keratin-associated proteins. *J. Invest. Dermatol. Symp. Proc.* 10, 230–233 (2005).
- Ling MM, Ricks C, Lea P. Multiplexing molecular diagnostics and immunoassays using emerging microarray technologies. *Expert Rev. Mol. Diagn.* 7, 87–98 (2007).
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell Proteomics* 4, 1265–1272 (2005).
- Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12, 1231–1245 (2002).

First draft submitted: 29 December 2017;

Accepted for publication 19 March 2018.

Address correspondence to Joseph Donfack, Counterterrorism & Forensic Science Research Unit, Federal Bureau of Investigation Laboratory Division, 2501 Investigation Parkway, Quantico, VA 22135, USA. Tel.: +1 703 632 4593; Fax: +1 703 632 4500; Joseph.Donfack@ic.fbi.gov

To purchase reprints of this article, contact: s.cavana@future-science.com