

# Proteome Analysis of Human Hair Shaft

FROM PROTEIN IDENTIFICATION TO POSTTRANSLATIONAL MODIFICATION\*<sup>§</sup>

Young Jin Lee‡, Robert H. Rice§¶||, and Young Moo Lee‡¶|

**The human hair proteome was investigated using two-dimensional LC-MS/MS. Among the 343 identified proteins, 70 were detected in high relative abundance, including keratin intermediate filament proteins, largely extractable with denaturants. Over 300 proteins were found to constitute the insoluble complex formed by transglutaminase cross-linking. The intracellular distribution of identified proteins is wide from cytoplasm to nucleus, mitochondria, ribosome, and plasma membrane. These results help rationalize ultrastructural features visible in the mature hair. Keratins and several substrates for transglutaminase were found to be posttranslationally modified by methylation and dimethylation. Evidence for ubiquitination of hair proteins was also obtained. *Molecular & Cellular Proteomics* 5:789–800, 2006.**

Hair is synthesized in mammals by specialized follicles in the skin. In the follicles, cells undergo an intricate maturation program resulting in a hair shaft of terminally differentiated cells. In humans, a protective layer of cuticle cells surrounds the cortex, which in turn surrounds an intermittent column of medulla cells in the center of the shaft. In each of the three layers, the cells have distinctive shapes and contents (1, 2). The keratins and keratin-associated proteins (KAPs),<sup>1</sup> major constituents of the cortex, have been studied for many years after extraction with denaturants under reducing conditions. However, the contents of the medulla and cuticle cells have been much harder to examine because they present great difficulties in solubilization. A major advance in understanding hair structure occurred with the finding of transglutaminase activity in hair follicles (3, 4). The resulting extensive  $\epsilon$ -( $\gamma$ -glutamyl)lysine cross-linking accounted for the insolubility in the medulla of proteins, notably the major transglutaminase substrate, trichohyalin. A high content of isopeptide bonding has also been reported in the cuticle cells (5), rationalizing the low degree of protein extraction apparent ultramicroscopically even under harsh denaturing and reducing conditions (6).

From the ‡Molecular Structure Facility and §Department of Environmental Toxicology, University of California, Davis, California 95616  
Received, August 25, 2005, and in revised form, January 23, 2006  
Published, MCP Papers in Press, January 30, 2006, DOI 10.1074/mcp.M500278-MCP200

<sup>1</sup> The abbreviations used are: KAP, keratin-associated protein; cLC, capillary reverse phase LC; DTE, dithioerythritol; MudPIT, multidimensional protein identification technology; PTM, posttranslational modification; SCX, strong cation exchange chromatography; Xcorr, cross-correlation value.

In recent years, identification of proteins participating in cross-linked structures in epidermis and appendages has benefited from the approach of peptide generation with enzymatic fragmentation followed by separation and amino acid sequencing of the peptides. Applied to the corneocyte-cross-linked envelope, a number of constituent proteins have been identified, including some where the isolated peptides were cross-linked, evidence of their participation in isopeptide bonds (7). Advances in current proteomic technology prompt application of this approach to the hair shaft where much of the protein resists extraction, and hence few data are available on the participants. It is evident, however, that much of the resistance is transglutaminase-dependent because hair from individuals afflicted by TGM1-negative lamellar ichthyosis is subject to considerable extraction (8). Present efforts using multidimensional protein identification technology (MudPIT) (9) permitted identification of 343 proteins in the hair shaft.

Large scale identification of the protein components of hair is anticipated to be of importance in helping understand the biogenesis of this epidermal appendage. Much has been learned regarding the mRNA transcripts and translation products in the living cells of the follicle (1, 10) because they are amenable to study by standard molecular biological techniques, but how the protein products are utilized in the mature hair shaft remains uncertain due in part to the paucity of information on which components are retained and which are discarded during terminal differentiation. Moreover this information is anticipated to help understand the molecular basis of aberrant hair phenotypes or even skin afflictions because, as in the case of the skin fragility/woolly hair syndrome (11) and Netherton syndrome (12), the two may have the same or related origin.

## EXPERIMENTAL PROCEDURES

### *Preparation of Hair Samples*

*Extraction*—Scalp hair samples from three unrelated individuals (designated HR, SM, and JY) were examined. Typically 40 mg of hair were rinsed briefly in 5 ml of 2% SDS, 50 mM sodium phosphate (pH 7.8) and drained. The hair was then immersed in 5 ml of 2% SDS, 50 mM sodium phosphate (pH 7.8), 20 mM DTE and incubated overnight at 65 °C. The hair was pulverized by magnetic stirring for an hour at room temperature, and the soluble and insoluble materials were separated by centrifugation. The insoluble material was resuspended in 2% SDS, 50 mM sodium phosphate (pH 7.8), 20 mM DTE; incubated overnight at 65 °C; and extracted as before. After five such extractions, the protein content in each fraction (including buffer blanks) was estimated by reaction of aliquots with ninhydrin after digestion

with 10% sulfuric acid (13). The first two SDS-DTE treatments were found to remove  $\approx 80\%$  of the total protein extracted. In six experiments, the final insoluble material comprised  $13.3 \pm 3.9\%$  of the total protein.

**Digestion**—Aliquots of the soluble (first extract) and insoluble material were incubated for 0.5 h in 2% SDS, 20 mM DTE, 50 mM phosphate buffer (pH 7.8) and then incubated an additional 0.5 h after adding iodoacetamide to 40 mM, all at room temperature. Proteins were precipitated from the soluble extract by addition of 2.5 volumes of ethanol. The soluble and insoluble protein samples were rinsed with 70% ethanol and then freshly prepared 0.1 M ammonium bicarbonate and resuspended in fresh 0.1 M ammonium bicarbonate adjusted to 2 M in recrystallized urea. The urea, which at 2 M does not interfere with trypsin action, was added to assist in solubilizing proteins and released peptides. To each suspension was added bovine L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin (Worthington), which was stabilized by reductive methylation (14) to 1% by weight. After 6–8 h at room temperature with constant stirring, a second equal aliquot of the methylated trypsin was added, and the samples were stirred overnight at room temperature. An estimated 92% of the detergent-insoluble material was solubilized, whereas virtually all of the detergent-soluble material was solubilized in this way.

### Separation of Hair Peptides by Two Dimensional Chromatography

**Strong Cation Exchange Chromatography (SCX)**—Trypsin-digested hair samples were dried and redissolved in  $\sim 200 \mu\text{l}$  of Solvent A (see below, Mobile phase A) and then injected onto a polysulfoethyl A cation exchange column ( $100 \times 2.1 \text{ mm}$ ,  $5\text{-}\mu\text{m}$  diameter,  $300\text{-}\text{\AA}$  pore size) from PolyLC (Columbia, MD) with the flow rate of  $200 \mu\text{l}/\text{min}$  utilizing the following mobile phases as described ([www.proteomecenter.org/](http://www.proteomecenter.org/)): Mobile phase A, 5 mM potassium phosphate (pH 3.0), 25% acetonitrile; Mobile phase B, 5 mM potassium phosphate (pH 3.0), 25% acetonitrile, 350 mM potassium chloride. After the sample was loaded, the run was isocratic for 15 min at 100% mobile phase A, and peptides were eluted using a linear gradient of 0–25% B over 30 min followed by a linear gradient of 25–100% B in 20 min and then held for 5 min at 100% B. Fractions at 2-min intervals were collected and concentrated by vacuum centrifugation.

**Capillary Reverse Phase LC (cLC)**—The SCX fractions were loaded sequentially on an on-line trap column ( $0.25 \times 30 \text{ mm}$ , Magic C18AQ,  $5 \mu\text{m}$ ,  $100 \text{\AA}$ ) at a flow rate of  $10 \mu\text{l}/\text{min}$  with buffer A (see below). After application and removal of salt and urea, the flow rate was decreased to  $300 \text{ nl}/\text{min}$ , and the trap column effluent was switched to a home-built fritless reverse phase microcapillary column ( $0.1 \times 180 \text{ mm}$ , packed with Magic C18AQ,  $5 \mu\text{m}$ ,  $100 \text{\AA}$ , Michrom Bioresources, Auburn, CA) following a published procedure (15). The reverse phase separation of peptides was performed using a Paradigm MG4 system (Michrom Bioresources) with buffers of 5% acetonitrile, 0.1% formic acid (buffer A) and 80% acetonitrile, 0.1% formic acid (buffer B) using a 150-min gradient (0–10% B for 20 min, 10–45% for 110 min, and 45–100% B for 20 min).

### MS Analysis

**LC-MS and LC-MS/MS**—Peptide analysis was performed utilizing a Finnigan LCQ Deca XP Plus system (San Jose, CA) coupled directly to an LC column. An MS survey scan was obtained for the  $m/z$  range of 400–1,400, and MS/MS spectra were acquired for the three most intense ions from the survey scan. An isolation mass window of 3.0 Da was used for the precursor ion selection, and a normalized collision energy of 35% was used for the fragmentation. Dynamic exclusion for a 2-min duration was used to acquire MS/MS spectra from low intensity ions.

**Data Analysis**—The Sequest (16) analysis software (Bioworks version 3.1) was used to find the peptide sequences in a human protein database that best match the observed MS/MS spectra. DTA files (Bioworks version 3.1) in ASCII format for each MS/MS spectrum were generated from the raw data for the peptide mass range of 500–3,500, minimum ion count of 10, and minimum signal of  $10^5$ . The International Protein Index human database (ipi.HUMAN.v3.01, December 3, 2004, downloaded from [ftp.ebi.ac.uk/pub/databases/IPI](http://ftp.ebi.ac.uk/pub/databases/IPI)) containing 46,941 entries was used for Sequest searching. Peptide (parent ion) tolerance of 2.5 Da and fragment ion tolerance of 1 Da were allowed, and fixed modification of carbamidomethylation on Cys (+57 Da) and differential modification of oxidation on Met (+16 Da) were used. DTASelect software (17) was used to filter out low score matching. Criteria used for filtering were: cross-correlation values (Xcorr) larger than 1.9, 2.2, and 3.75 were used for singly, doubly, and triply charged ions, respectively, for both individual half- or fully tryptic peptides;  $\Delta\text{Cn}$  values (the difference in Xcorr with the next highest value) less than 0.08 were removed from the matched sequences (18–20). Proteins with at least two peptide ions satisfying the above criteria were considered correctly identified. Manual examination was carried out for all proteins identified with less than five peptides. Criteria used for manual validation (18) included (a) good quality MS/MS spectra with fragment ions clearly above base-line noise, (b) minimum of four continuous y or b series ions observable, (c) intense y and b ions corresponding to the N-terminal cleavage of Pro but weak ones for C-terminal Pro except where Pro is the second residue from the peptide N terminus, and (d) all the major intense ions interpretable with ProteinProspector ([prospector.ucsf.edu/](http://prospector.ucsf.edu/)). For the proteins matched with four or five peptides, at least one MS/MS spectrum was validated manually to meet the identification criteria. Proteins covered with two to three peptides were accepted only when a minimum of two spectra passed the manual validation criteria. Further description of the criteria used for protein identification is given under “Results.”

## RESULTS

The soluble and insoluble fractions were digested separately with trypsin, and the peptides generated were first separated off line using strong cation exchange chromatography, typically obtaining 34 SCX fractions. On-line reverse phase chromatography coupled to mass spectrometry was then performed for each fraction. The resulting experimental MS/MS spectra were searched with Sequest software against theoretical MS/MS spectra obtained from the human database.

### Identification of Proteins by Sequest Search

A typical cLC-MS/MS analysis of the 34 fractions of the insoluble hair proteome generated  $>200,000$  MS/MS spectra. DTASelect filtering of the Sequest results specifying both “fully tryptic” (both cleavages specific for trypsin) and “half-tryptic” (one cleavage specific for trypsin and the other non-specific) cleavage provided over 1,700 non-redundant peptides that were mapped to 250–300 proteins. As illustrated in Table I, 310 potential proteins were identified in one individual (JY) when the criterion of accepting half-tryptic peptides was applied.

To increase the confidence level of the protein identification, we validated the MS/MS spectra manually as described under “Experimental Procedures.” We also compared the

TABLE I  
Total number of human hair proteins detected from the insoluble fraction of three individuals and false positive rates at the given identification criteria

The total number of validated non-redundant proteins from the three samples is 343. False positive rates given in parentheses were calculated from the proteins that passed Sequest criteria but failed to pass manual validation.

Sample	Tight criteria without manual validation <sup>a</sup>	Loose criteria <sup>b</sup>		
		Before manual validation	After manual validation <sup>c</sup>	After further removal of redundant proteins <sup>d</sup>
SM	215 (3.2%)	248 (6.9%)	231	221
JY	258 (8.1%)	310 (15.2%)	263	247
HR	214 (2.8%)	253 (11.4%)	224	211

<sup>a</sup> Tight criteria. Accepted fully tryptic peptides only. Xcorr at least 1.9, 2.2, and 3.75 for singly, doubly, and triply charged peptide ions, respectively.  $\Delta Cn$  at least 0.08. Identified with at least two peptides for each protein.

<sup>b</sup> Loose criteria. Same as tight criteria except half-tryptic peptides were also accepted.

<sup>c</sup> MS/MS spectra of the proteins that passed loose criteria were subjected to manual validation.

<sup>d</sup> Redundant proteins were further removed from manually validated proteins. Namely proteins with no unique peptide were removed. For proteins exhibiting unique peptides, at least one was manually examined for final verification.

number of proteins identified by two different criteria, assuming fully tryptic (tight criterion) and half-tryptic cleavage (loose criterion). As can be seen in Table I, the false positive rates, calculated from the proteins that passed Sequest criteria but failed to pass manual validation, were low (3–8%) when the tight criterion accepting only fully tryptic peptides was used. When half-tryptic peptides were also allowed, false positive rates were significantly increased (7–15%). However, most of the false positives must have been removed by manual validation. We also performed Sequest searches against the reversed sequence database to have another measure of the false positive rate (data not shown). Estimated false positive rates obtained by searching against the reversed sequence database were a bit lower, 6–11% for half-tryptic peptides compared with manual validation. With one exception, none of the false positive proteins could have passed the manual validation.

The sole false positive protein in the reversed sequence database search that could potentially escape the manual examination is noteworthy in view of the structural similarity of the peptide ions between the real and reversed databases. The two peptides share very similar amino acid sequence and composition, PAM<sup>ox</sup>DLFQDR (where superscript ox indicates oxidation (+16 Da)) (real) derived from the sequence of Hsp89- $\alpha\delta N$  versus APFDLFENR, a reversed sequence. Note that Met<sup>ox</sup> and Phe have the same residual mass of 147 Da; internal fragment masses of QD and EN are the same, 243 Da; and mass differences between Gln and Glu and Asp and Asn are only 1 Da, which may fall into the mass error range of the three-dimensional ion trap mass spectrometer. Yet the real sequence of PAM<sup>ox</sup>DLFQDR yielded higher Xcorr than the analogue (2.67 versus 2.25 for doubly charged ion and 2.01 versus 1.93 for singly charged ion).

After manual examination of the initial filtered spectra, a second manual filtering process was performed to remove redundant proteins to the extent possible. If a set of peptides matches the sequence of two or more separate proteins, the current version (version 1.9) of DTASelect does not eliminate

this redundancy. In addition to removal of such redundancy, at least one unique peptide was manually examined for final validation when a protein exhibited unique peptides so that it could be distinguished from others. When we used the criteria accepting both full and half-tryptic peptides combined with the described manual validation procedure and removal of redundant proteins, the total numbers of proteins identified with high confidence were 211, 247, and 221 for the insoluble fraction obtained from HR, JY, and SM, respectively (Table I).

#### Proteins Identified from Insoluble Proteome

**Keratin and KAP Families**—Table II shows the major human hair proteins detected in the insoluble fraction from all three individuals. The proteins in Table II each exhibited at least four unique peptide sequences and are sorted by the sum of unique peptides. As can be seen, seven of the top 20 proteins are types I and II keratins of cuticular origin. Non-redundant MS/MS spectra matched to keratins are almost 50% of the total number of spectra (non-redundant) acquired in two-dimensional LC-MS/MS: 49% in HR (3,470 of 7,040), 43% in JY (4,135 of 9,597), and 51% in SM (3,051 of 5,946).

A total of five KAPs belong to the category of major proteins as illustrated in Table II. These were not as dominant as keratins in abundance assuming the number of MS/MS spectra is proportional to protein amount (data not shown) (21). The complete list of KAPs identified can be found in Supplemental Table 3. However, KAPs are likely under-represented because they are relatively small proteins (10–15 kDa) containing limited numbers of trypsin cleavage sites that could result in poor sequence coverage and prevent identification by the MudPIT approach. We have attempted using various endopeptidases including subtilisin A and/or chemical cleavage (CNBr) to improve the sequence coverage without much success.

**Non-keratin Proteins**—Desmosomal proteins are among the most abundant non-keratin proteins found in the insoluble hair proteomes. The number of unique peptides observed in

TABLE II  
Major human hair proteins identified in the insoluble fractions in all three individuals

At least four unique peptide sequences were identified from each individual. Only one protein entry is listed if several proteins were indistinguishable. The list is sorted in decreasing order of the sum of unique peptides in all three samples (not shown). IPI, International Protein Index.

IPI accession number	Number of unique peptides			Description
	SM	JY	HR	
IPI00013933	72	119	58	Splice isoform 1 of desmoplakin
IPI00300053	29	72	64	Keratin, type II cuticular HB2
IPI00291540	44	37	46	Keratin, type I cuticular HA2
IPI00305719	39	41	24	Selenium-binding protein 1
IPI00376274	30	25	28	Predicted: similar to hypothetical protein A030011M19 (Ig-like domain <sup>a</sup> )
IPI00071509	21	39	18	Splice isoform 1 of plakophilin 1
IPI00420079	25	29	20	JUP (plakoglobin)
IPI00182655	30	26	17	Keratin, type II cuticular HB6
IPI00428691	28	26	15	Desmoglein 4
IPI00013890	17	24	20	14-3-3 protein $\sigma$
IPI00003935 etc.	31	12	14	H2A family (O/Q, X, Y, Z) <sup>b</sup>
IPI00216711	22	25	9	Keratin-associated protein 11-1
IPI00032541	14	20	20	Keratin, type II cuticular HB5
IPI00465330	16	23	15	Type I hair keratin KA35
IPI00294649	12	17	19	Keratin, type I cuticular HA5
IPI00021439	19	16	12	Actin, cytoplasmic 1
IPI00397845	13	10	23	Similar to RIKEN cDNA 2310002B14
IPI00292715	13	19	12	Type I hair keratin 4
IPI00261031	13	18	11	Predicted: similar to hephaestin
IPI00303476	17	14	10	ATP synthase $\beta$ chain, mitochondrial precursor
IPI00186711	11	17	13	Plectin 6
IPI00008524	10	20	10	Splice isoform 1 of polyadenylate-binding protein 1
IPI00020101 etc.	13	12	14	H2B family (A, Q) <sup>c</sup>
IPI00455315	13	16	7	Annexin A2 isoform 2
IPI00152871	16	12	7	Leucine-rich repeat-containing protein 15 precursor
IPI00453473	14	12	9	HIST1H4F protein
IPI00419237	7	16	12	LAP3 protein
IPI00016752	12	8	14	S100 calcium-binding protein A3
IPI00013475	5	10	18	$\beta$ -Tubulin
IPI00456625	11	12	10	Ka36 protein
IPI00025491	11	12	9	Eukaryotic initiation factor 4A-I
IPI00300376	10	8	13	Protein-glutamine glutamyltransferase E precursor
IPI00394683	8	10	12	Keratin-associated protein 10-12
IPI00219018	10	8	11	Glyceraldehyde-3-phosphate dehydrogenase
IPI00025447	10	10	9	Elongation factor 1- $\alpha$ 1
IPI00328398	7	10	12	Lysozyme g-like protein 2 precursor
IPI00291006	10	11	8	Malate dehydrogenase, mitochondrial precursor
IPI00019912	7	15	7	Peroxisomal multifunctional enzyme type 2
IPI00024067	7	7	15	Clathrin heavy chain 1
IPI00011229	9	12	7	Cathepsin D precursor
IPI00031423	9	7	11	Keratin, type I cuticular HA3-II
IPI00216984	9	9	9	Calmodulin-like 3
IPI00220301	10	8	9	Peroxiredoxin 6
IPI00294840	7	15	5	Predicted: absent in melanoma 1 (crystalline $\beta$ $\gamma^a$ )
IPI00021405	9	10	5	Splice isoform 1 of lamin A/C
IPI00007702	8	8	7	Heat shock-related 70-kDa protein 2
IPI00186290	6	10	7	Eukaryotic translation elongation factor 2
IPI00017292	5	7	10	Splice isoform 1 of $\beta$ -catenin
IPI00022482	6	7	9	Sialidase 2
IPI00013265	10	8	4	Keratin-associated protein 2
IPI00003865	5	6	9	Splice isoform 1 of heat shock cognate 71-kDa protein
IPI00008486	5	6	9	Protein-arginine deiminase type III
IPI00180675	8	4	8	Tubulin $\alpha$ -3 chain <sup>d</sup>
IPI00003362	4	7	8	78-kDa glucose-regulated protein precursor
IPI00465028	5	7	7	Triose-phosphate isomerase 1
IPI00465248	6	6	6	$\alpha$ -Enolase (IPI database version 3.05)
IPI00375323	7	5	6	Keratin-associated protein 12-2
IPI00394674	8	6	4	Keratin-associated protein KAP10-11
IPI00219575	5	5	8	Bleomycin hydrolase
IPI00016461	5	7	6	1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase $\delta$ 1
IPI00472082	4	8	5	Maspin precursor

TABLE II—continued

IPI accession number	Number of unique peptides			Description
	SM	JY	HR	
IPI00465431	4	7	6	LGALS3 protein
IPI00219757	5	6	6	Glutathione S-transferase P
IPI00008529	8	5	4	60 S acidic ribosomal protein P2
IPI00006663	5	7	5	Aldehyde dehydrogenase, mitochondrial precursor
IPI00031461	4	5	7	Rab GDP dissociation inhibitor $\beta$
IPI00000875	4	5	6	Eukaryotic translation elongation factor 1 $\gamma$
IPI00414676	5	4	4	Heat shock 90-kDa protein 1, $\beta$
IPI00385244	5	4	4	Phosphoglycerate mutase 1 (brain)
IPI00479186	5	4	4	Pyruvate kinase 3 isoform 1

<sup>a</sup> ProSite predictions of protein families and domains for uncharacterized proteins translated from genomic or cDNA sequences (us.expasy.org/prosite/).

<sup>b</sup> Histone 2A families O (or Q), X, Y, and Z (IPI00216457, IPI00059366, IPI00219037, and IPI00249267) are highly homologous and taken as a single entry. Otherwise with only one or two unique peptides, they would not be included as major proteins.

<sup>c</sup> Histone 2B families A and Q (IPI00020101 and IPI00003935) are highly homologous, sharing 10–12 peptides, and hence are combined as a single entry.

<sup>d</sup> IPI00387144 ( $\alpha$ -tubulin, ubiquitous), found only in sample HR with a single unique peptide, is highly homologous to tubulin  $\alpha$ -3, sharing six peptides. The peptides from this protein have been treated as if derived from tubulin  $\alpha$ -3.

desmoplakin exceeds those of keratins, as can be seen in Table II, indicative of their high relative abundances in the hair shaft. Desmoplakin is a large molecular mass (~332 kDa) structural constituent of the cytoskeleton that plays a vital role in keratinocyte adhesion in linking the transmembranous desmosomal cadherins to the cytoplasmic keratin filament network (22). Other desmosomal proteins identified include plakoglobin, plakophilin, and desmoglein 4. These three cell junction proteins provide essential adhesion structures in most epithelia linking the intermediate filament networks of neighboring cells (22). Additional desmosomal proteins not listed in the major protein category in Table II are plakophilin 3, desmocollin 3 (seen in all three individuals), and desmocollin 1 (seen at a low level in one individual) (Supplemental Table 1). Other notable structural proteins include histones: H2A family, H2B family, and HIST1H4F protein. Their relative abundances, particularly histone families H2A and H2B, are high as indicated by the total numbers of MS/MS spectra (548 for H2B-A and 199 for H2A-O, see Supplemental Table 1). Other structural components classified as major proteins (Table II) are tubulin ( $\alpha$  and  $\beta$  chains), actin,  $\beta$ -catenin, plectin 6, and lamin A/C.

Relatively abundant proteins of unknown physiological function include a hypothetical protein, A030011M19, and four metal (selenium and calcium)-binding proteins. The functions of the latter are not fully understood, but they are found to be expressed widely including in the liver, lung, kidney, and mammary gland (23). Localization of S100 calcium-binding protein family members exhibits specificity in the hair follicle. For example, S100A2 is found in the outer root sheath, S100A3 is found in the cortex and cuticle, and S100A6 is found in the inner root sheath (24). Accordingly we found S100 calcium-binding protein A3 to be highly abundant (the number of MS/MS spectra ranges from 52 to 160 among three samples), but S100 A2 and A6 were not detected. It is also worth noting that S100 A14 was found in all three samples with lower abundance (Supplemental Table 1); this is a

form previously known to be highly expressed in colon and moderately expressed in thymus, kidney, liver, small intestine, and lung (25) but not known in hair. Along with the S100 calcium-binding protein family, the calcium-binding proteins calmodulin-like protein 3 and annexin A2 were also detected, consistent with the need for calcium ions in the proper function of desmosomal cadherin (e.g. desmogleins) in the living cells of the hair follicle.

Detection of transglutaminase E (TGM3) as a major protein found in all three individuals attests to its likely high importance in formation of  $N^{\epsilon}$ -( $\gamma$ -glutamyl)lysine cross-links of protein substrates in hair. This enzyme was relatively abundant judging from the number of assigned MS/MS spectra (ranging from 25 to 39) in samples from all three individuals. Another enzyme involved in protein cross-linking, transglutaminase K (TGM1), was detected in two samples (Supplemental Table 1) with lower abundance.

Other major proteins identified can be assigned to several categories. 1) Protein stabilization: all heat shock proteins including heat shock-related 70-kDa protein 2, heat shock cognate 71-kDa protein, heat shock 90-kDa protein 1, and 78-kDa glucose-regulated precursor. 2) Protein synthesis: ribosomal proteins including 60 S acidic ribosomal protein P2. 3) Translation regulation: elongation factor 1- $\alpha$  and  $\gamma$ 1, eukaryotic translation elongation factor 2, and initiation factor 4A-1. 4) Signaling: the 14-3-3 proteins, leucine-rich repeat-containing protein 15, 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase, and annexin A2. The 14-3-3 proteins, reportedly having an important role in keratin filament organization (26), have received considerable attention recently for their regulatory action through their protein binding motifs, including to phosphoserine and phosphothreonine (27). 5) Glycolysis: glyceraldehyde-3-phosphate dehydrogenase and triose-phosphate isomerase. 6) Other metabolism: malate dehydrogenase, peroxiredoxin 6, pyruvate kinase 3,  $\alpha$ -enolase, phosphoglycerate mutase, and aldehyde dehydro-

genase. 7) Enzymes involved in protein or amino acid metabolism: protein-arginine deiminase type III responsible for the conversion of arginine residues to citrullines in trichohyalin. 8) Proteolytic enzymes: lysozyme g-like protein, bleomycin hydrolase, and cathepsin D. 9) Excision of carbohydrate: sialidase 2. 10) Unknown function: LAP3 protein, LGALS3, and several hypothetical proteins just to mention a few.

*Proteins Identified from Soluble Fraction*

Two-dimensional LC-MS/MS analysis of soluble fraction proteins was also performed for one of the hair samples (HR) as shown in Supplemental Table 2. Despite searching ~140,000 MS/MS spectra with Sequest, only 37 proteins were identified, and 23 of them (62%) were keratins and KAP family proteins. Of 3,751 identified non-redundant MS/MS spectra, 3,678 (98%) were from keratins and KAPs. Hence it is likely that the soluble fraction contains mostly disulfide-cross-linked keratins and KAPs extractable with SDS-DTE. All the minor non-keratin proteins observed in the soluble fraction were detected in the insoluble fraction as well. These non-keratin proteins may have formed disulfide bridges with keratins and KAPs before extraction or were trapped in the keratin matrix without participating in cross-links.

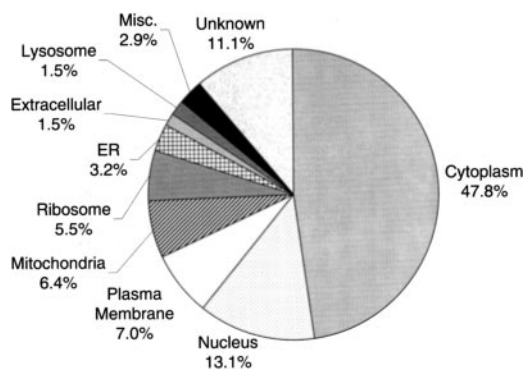
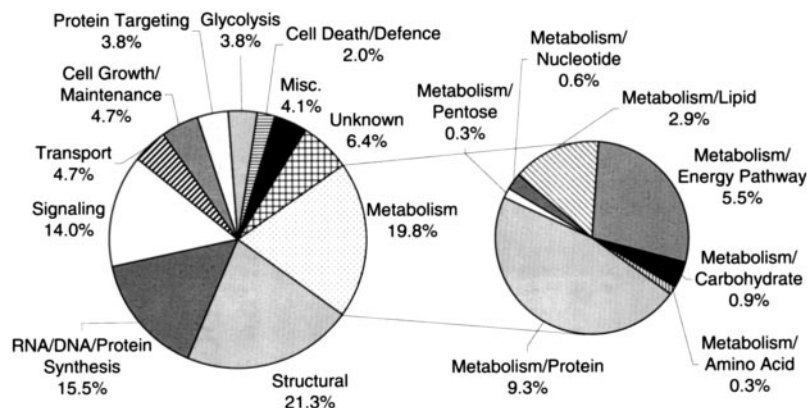


FIG. 1. Cellular distribution of the 343 identified proteins. Cellular locations were assigned with assistance from the Human Protein Reference Database ([www.hprd.org/protein](http://www.hprd.org/protein)). ER, endoplasmic reticulum.

FIG. 2. Functional distribution of the 343 identified proteins. Assignment of biological function was carried out with assistance from the Human Protein Reference Database ([www.hprd.org/protein](http://www.hprd.org/protein)) and in some cases the human heart mitochondrial proteome (49). In the latter, subclasses of DNA synthesis, protein modification, transcription, and translation are all classified as DNA/RNA/protein synthesis.



*Cellular Location and Functional Classification of Proteins Cross-linked by Transglutaminases*

The intracellular locations of the identified proteins are shown in Fig. 1. As illustrated, the majority of proteins identified (47.8%) originated from the cytoplasm. The next three significant locations were the nucleus (13.1%), plasma membrane (7%), and mitochondria (6.4%). The original location of ~11% of the detected proteins is unknown.

As can be seen in Fig. 2 the proteins have a diversity of functions from structural (21.3%) to DNA/RNA/protein synthesis (15.5%), metabolism (19.8%), and signaling (14%). Minor but still significant functions include transport (4.7%), protein targeting (3.8%), glycolysis (3.8%), cell growth/maintenance (4.7%), and cell death/defense (2%). Proteins of unknown function correspond to about 6.4%. The observed wide cellular functional distribution of proteins reflects the wide range of proteins that can serve as substrates for transglutaminases.

*Posttranslational Modification (PTM) of Human Hair Proteins*

To find PTMs, we constructed a subprotein database comprised of all the identified proteins from the three hair samples. All MS/MS spectra were searched again several times with Sequest against this human hair protein database of much smaller size. As a start, four PTMs that were analyzed including ubiquitination (delta mass of +114 Da on lysine), methylation (+14 Da on lysine, arginine, and histidine), trimethylation/acetylation and dimethylation (+42 Da on lysine, +28 Da on lysine), and phosphorylation (+80 Da on serine, threonine, and tyrosine). Because inclusion of these four PTMs in searches would increase the risk of false positives, very tight criteria were used to ensure the high level of confidence in identification: only fully tryptic peptides without missed cleavages (except modified lysine and/or arginine) and with Xcorr >1.9 for singly charged, >2.2 for doubly charged, and >3.75 for triply charged peptide ions. Furthermore all MS/MS spectra meeting these criteria were manually examined with higher standards than used for protein identification: no unassignable major peaks were allowed, and most y and b series

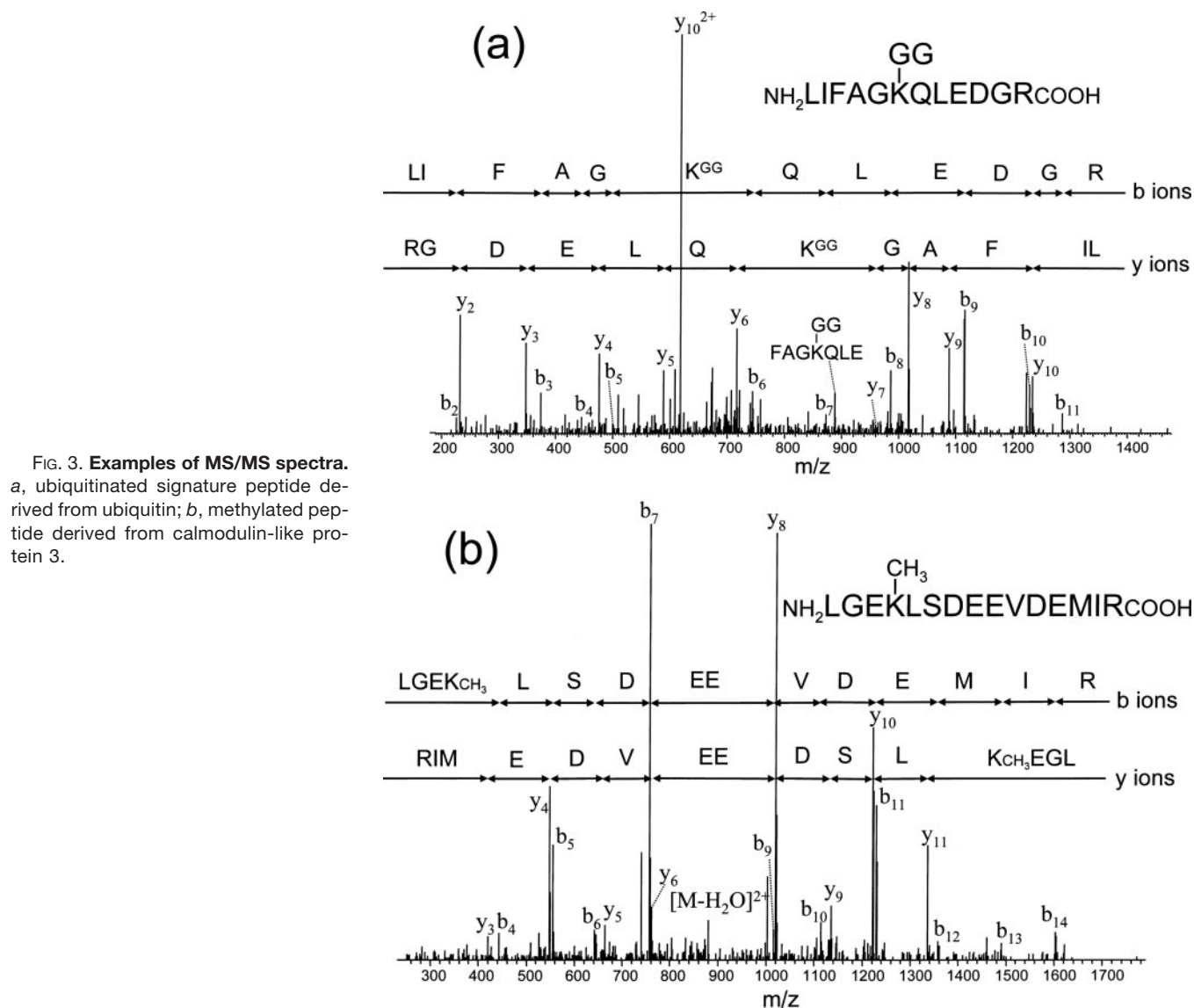


FIG. 3. Examples of MS/MS spectra. a, ubiquitinated signature peptide derived from ubiquitin; b, methylated peptide derived from calmodulin-like protein 3.

ions had to be observed. In this way, we were able to detect examples of ubiquitination, methylation, and dimethylation.

Fig. 3a shows the MS/MS spectrum of a peptide derived from ubiquitinated ubiquitin. The ubiquitinated peptide, LI-FAK<sup>GG</sup>QLEDGR (where superscript GG represents the C-terminal Gly-Gly residues of ubiquitin) is derived by trypsin digestion of ubiquitin itself. Detection of this signature peptide (28) arises from the oligomerization of ubiquitin to initiate protein degradation by the proteasome. Detected in hair samples from all three individuals, its high abundance (thus readily detectable) should be indicative of the prevalence of ubiquitin-tagged proteins in the proteome.

Using stringent criteria, we detected methylation of the 12 proteins shown in Table III. Besides histones, well known for this modification (29, 30), conclusive evidence was obtained for methylation/dimethylation of eight other hair shaft proteins: calmodulin-like protein 3, desmoplakin, sialidase 2, ac-

tin (cytoplasmic 1), unknown protein (similar to RIKEN cDNA 4732495G21 gene), keratin I-HA2, keratin II-HB2, and keratin II-HB6. Methylation of mouse and bovine actin (cytoplasmic 1) on His-73 is well known ([www.expasy.org/](http://www.expasy.org/)). The residue in calmodulin homologous to Lys-115 in calmodulin-like protein 3 is trimethylated (31), but in the current study a single methylation (delta mass of +14) was observed in the latter (Fig. 3b). In the case of keratin II-HB2, a lysine residue in the sequence LQQETNNVKAQR seems to be modified both by dimethylation and trimethylation (or acetylation). To the best of our knowledge, methylation/dimethylation/trimethylation of keratins, desmoplakins, and sialidase 2 have not been reported previously. Even in the histone molecules, none of the methylation sites were known except the two recently reported in histone 3, VTIMPK<sup>m</sup>DIQLAR (where superscript m indicates methylation (+14 Da)) and EIAQDFK<sup>dm</sup>TDLR (where superscript dm indicates dimethylation (+28 Da)) (32).

TABLE III

Posttranslational modification of mono-, di-, and trimethylation of human hair proteins identified from the insoluble fraction

Three cLC-MS/MS data sets from insoluble fractions were re-searched against 343 identified proteins with possible ubiquitination (+114 on Lys), methylation (+14 on Lys, Arg, His), trimethylation/acetylation and dimethylation (+42 on Lys, +28 on Lys, Arg), and phosphorylation (+80 on Ser, Thr, Tyr). Very tight criteria were used (fully tryptic peptides, no missed cleavage except modified Lys or Arg, Xcorr  $\geq$  1.9, 2.2, and 3.75 for +1, +2, and +3 ions, respectively) and manually validated with high standards (No unassignable significant intensity peaks were allowed, and most y and b series ions should be identified.). Observed modifications on amino acids are denoted with the following superscript symbols: m, methylation (+14 Da); dm, dimethylation (+28 Da); tm, trimethylation/acetylation (+42 Da) (This assignment is tentative because it cannot be distinguished from acetylation (+42 Da), and also the possibility of carbamylation cannot be totally eliminated.); ox, oxidation (+16 Da). Bold indicates posttranslationally modified amino acid residues.

Protein	Sample <sup>a</sup>	MS/MS <sup>b</sup>	Sequence <sup>c</sup>	Xcorr	z <sup>d</sup>
Calmodulin-like 3	1	2	R.LGEK <sup>m</sup> LSDEEVDEMIR.A	3.6939	2
	1	1	R.LGEK <sup>m</sup> LSDEEVDEM <sup>ox</sup> IR.A	2.8793	2
Desmoplakin	1	1	R.IQESK <sup>dm</sup> NQCTQVQER.E	4.3951	2
Sialidase 2	1	1	R.ADLGAYLNPR <sup>m</sup> PPAPEAWSEPVLAK.A	3.9975	3
Actin, cytoplasmic 1	3	4	K.YPIEH <sup>m</sup> GIVTNWDDMEK.I	4.5829	2
	1	1	K.YPIEH <sup>m</sup> GIVTNWDDM <sup>ox</sup> EK.I	4.004	2
Similar to RIKEN cDNA 4732495G21 (actin <sup>e</sup> )	2	2	K.YPIEH <sup>m</sup> GVVTNWDDMEK.I	4.2996	2
H1, 5	1	1	R.K <sup>dm</sup> ATGPPVSELITK.A	3.0329	2
H2B, Q	2	2	K.VLK <sup>dm</sup> QVHPDTGISSK.A	3.6588	2
	2	2	R.LLLPGELAK <sup>dm</sup> HAVSEGTK.A	3.8778	2
H3	1	1	R.VTIPK <sup>m</sup> DIQLAR.R	2.2606	2
	2	6	R.EIAQDFK <sup>dm</sup> TDLR.F	2.8644	2
	1	1	R.VTIM <sup>ox</sup> PK <sup>dm</sup> DIQLAR.R	2.5116	2
HIST1H4F	1	4	R.DNIQGITK <sup>m</sup> PAIR.R	3.0434	2
	1	6	R.GVLK <sup>m</sup> VFLENVIR.D	3.8519	2
	1	6	R.DNIQGITK <sup>dm</sup> PAIR.R	3.5092	2
	2	11	R.GVLK <sup>dm</sup> VFLENVIR.D	4.365	2
Keratin, I, HA2	1	1	R.LASYLTR <sup>dm</sup> VR.Q	2.236	2
Keratin, II, HB2	1	2	R.FASFINK <sup>dm</sup> VR.F	2.9124	2
	1	1	K.K <sup>dm</sup> YEEELSLRPCVENEVVALK.K	3.9168	3
	1	2	R.LQQETENVK <sup>dm</sup> AQR.	3.5013	2
	2	4	R.LQQETENVK <sup>tm</sup> AQR.C	3.6563	2
Keratin, II, HB6	2	2	K.K <sup>dm</sup> YEEEVSLR.A	3.1015	2
	1	1	R.K <sup>dm</sup> SDLEANVEALIQEIDFLR.R	5.3306	2
	1	1	K.LAELEGALQK <sup>dm</sup> AK.Q	3.6607	2

<sup>a</sup> Number of samples where the peptides were found.

<sup>b</sup> Number of MS/MS spectra observed for the sequences.

<sup>c</sup> Amino acids at N-terminal and C-terminal side of the peptide sequences are also shown before and after dots.

<sup>d</sup> Charge state of peptide ions.

<sup>e</sup> ProSite prediction of protein families and domains for uncharacterized proteins translated from genomic or cDNA sequences (us.expasy.org/prosite/).

## DISCUSSION

In this extensive MudPIT examination of the human hair shaft proteome, we identified a total of 343 proteins from three hair samples. However, the possibility must be considered that some proteins present in the hair escaped detection. The huge dynamic range of protein abundance in cells and the insolubility of many membrane proteins in general are limiting factors. In addition, the unknown digestibility of proteins extensively cross-linked by N<sup>ε</sup>-( $\gamma$ -glutamyl)lysine and the solubility of the peptides contribute to the uncertainty. These factors help rationalize the finding that of the total 343 proteins identified only 143 were seen in samples from all three individuals. As an example, trichohyalin, a major constituent of cross-linked material in the hair shaft medulla (1), was detected in only one individual proteome (JY) in this study (Supplemental Table 1). Its under-representation is likely due

to its high degree of cross-linking, the extensive conversion of constituent arginine to citrulline residues, resulting in fewer trypsin cleavage sites, and the low proportion of protein in human hair arising from the single intermittent column of medulla cells at the center of the shaft.

The hair shaft is comprised of a surprisingly large number of intracellular proteins. The SDS-DTE-soluble proteins consist primarily of a moderately complex but well studied mixture of keratins and KAPs. "Hard" keratins are the most abundant structural proteins in hair. Encoded by a large multigene family in two major sequence groups ("acidic" or type I and "basic" or type II), analogous to the "soft" keratins found in the epidermis, they form a cytoplasmic network of 8–10-nm-diameter intermediate filaments. Their expression is coordinated with that of smaller matrix components such as the associated cysteine-rich and glycine/tyrosine-rich KAP fami-



lies (1) that fill the spaces between keratin intermediate filaments. KAPs are envisioned to interact with keratin filaments and stabilize the complexes with disulfide linkages.

A novel finding in the present work is the methylation, dimethylation, and trimethylation observed in the keratins. These modifications are all in the central domain dominated by  $\alpha$ -helical subsegments in contrast to other modifications (glycosylation, phosphorylation, and isopeptide bonding) located in the N- or C-terminal domains (33). Some of the observed protein methylation sites show strong sequence homologies, including **LASYLTRVR** (Arg-113 in keratin H2A) versus **FASFINKVR** (Lys-137 in keratin HB2) and **KY-EEEVSLR** (Lys-193 in keratin HB6) versus **KYEEELSLR** (Lys-207 in keratin HB2), where matching residues are shown in bold. Lys-115 in calmodulin is known as a site of trimethylation, and this lysine residue is found within a highly conserved six-amino acid loop (LGEKLT) that forms a turn between helix-6 and -7 (31). A very similar six-amino acid loop was also found in calmodulin-like protein 3 (Table III), LGEKLS, showing conservative replacement of Thr with Ser. Hence methylation sites seem to have sequence motifs in keratins and calmodulins. As shown in Table III, methylation was observed not only in keratins and calmodulin-like protein 3 but also in desmoplakin, sialidase 2, actin, and an unknown protein. We also found extensive methylation on histones (Table III). It is worth noting that methylation but not dimethylation on residue Lys-122 (VTIMPK<sup>m</sup>DIQLAR and VTIM<sup>ox</sup>PK<sup>dm</sup>DIQLAR) was observed in a recent extensive investigation of histone 3 methylation (32). The biological significance of the newly discovered methylation/dimethylation/trimethylation sites on keratins, calmodulin-like protein 3, desmoplakin, sialidase, actin, and histones remains to be clarified.

Very careful interpretation of MS/MS spectra was required in characterization of methylation and/or dimethylation sites because the small mass changes make it difficult to distinguish these modifications from possible point mutations (or isoforms). With the lower resolution ion trap mass spectrometer used for the present study, for example, it is not possible to differentiate a dimethylated Lys residue from replacement of Lys by Arg; with nominal masses of 156 Da, the two choices differ by only 0.025 Da. Other possible amino acid replacements in the surrounding sequence of a peptide should be considered when y or b ions of corresponding methyl-Lys are not observed (*i.e.* GK<sup>m</sup> has the same residual mass as AK). Possible changes of 14 or 28 Da when one amino acid is replaced by another include: Gly to Ala, Ser to Thr, Val to Leu or Ile, Thr to Asp, Asn to Gln or Lys, and Asp to Glu (+14 Da) and Ala to Val, Ser to Asp, and Thr to Glu (+28 Da). Replacement by two amino acids should also be considered: Lys<sup>m</sup> by VV (142 Da), Arg<sup>m</sup> by GL/LG or AV/VA (170 Da), Lys<sup>dm</sup> by GV/VG (156 Da), Arg<sup>dm</sup> by AL/LA or SP/PS (184 Da), and Lys<sup>tm</sup> (where superscript tm indicates trimethylation/acetylation (+42 Da); this assignment is tentative because it cannot be distinguished from acetylation (+42 Da), and also

the possibility of carbamylation cannot be totally eliminated) by GL/LG or AV/VA (170 Da). To eliminate the above possibilities, extensive BLAST searches were carried out with all peptides in Table III as the query sequences with plausible replacements for the residues in bold. Blast results did not list any sequences of such single amino acid replacement. In addition, as can be seen in the MS/MS spectra of posttranslationally modified peptides (Supplemental Fig. 1), the observed y and b series ions countered such possibilities. The novel observation of Arg double methylation (LASYLTR<sup>dm</sup>VR) could also be interpreted as a single methylation of the homologous peptide (LASYLDR<sup>m</sup>VR) from cytokeratin 18. However, clear observation of y<sub>3</sub> and b<sub>6</sub> ions corresponding to R<sup>dm</sup>VR and LASYLT fragments in its MS/MS spectrum, whereas those for LASYLDR<sup>m</sup>VR were absent (Supplemental Fig. 1), gives confidence in the assignment of double methylation of Arg.

A striking finding is that the proteins comprising the hair shafts are derived from nearly the entire cell. This finding can be rationalized by ultrastructural examination of hair after extensive detergent extraction, permitting visualization of the intricate remaining features of each cell type (6). Most germane, little loss of protein content is evident by extraction of cuticle cells. The marginal band (A layer) at the outer border, the finely textured material in the outer half (exocuticle), and the amorphous large grained material in the inner half (endocuticle) appear largely unaffected. Large granules in the endocuticle, some of which resemble mitochondria in appearance, are reminiscent of the "dust bin hypothesis" for the origin and variability of keratinocyte-cross-linked envelopes (34) and could account for the variety of proteins observed. By contrast, extracted cortical cells are largely devoid of internal content. This observation indicates that the keratins are largely solubilized, but the cross-linking of a substantial fraction into insoluble material is evident biochemically and can be important functionally in epidermal cells (35). This phenomenon likely reflects the close association of the keratins with a variety of linker proteins (including desmoplakin, plakophilin, plakoglobin, and desmoglein 4) that facilitate connecting the cytoskeleton with desmosomes and whose absence has pathological consequences (36). In the extracted hair, cortical cell boundaries are clearly visible, indicative of at most only low level cross-linking of the cytoskeleton with junctional proteins that are prominent in the tabulation. Consistent with characterization of the role of trichohyalin as a transglutaminase substrate (37), the medulla cells where it is largely found exhibit large deposits of insoluble material, including nuclear remnants, after extraction. This material could be a source of proteins associated with chromatin, such as histones, although the contribution of the medulla to the total material presumably is much lower for human hair than for mouse hair in view of the prominence of the medulla in the latter (38).

Among the common desmosomal junctional proteins, only

desmocollins seem to be lacking in the hair shaft. Whether this has functional significance or simply reflects detection difficulties is uncertain. In view of the variable protein composition of desmosomes in the hair follicle (39) and the possibility that remodeling may occur, as reported for inner root sheath desmocollin (40), the paucity of desmocollins could influence adhesiveness among cells of the hair shaft. Because expression of members of this family are important for regulating intercellular adhesion (41, 42), the possibility that hair cell adhesive properties are modulated by substitution or augmentation with another protein may merit consideration. This speculation is prompted by the observation that the hypothetical protein A030011M19 appears highly abundant (764 MS/MS) in this regard next only to desmoplakin (1,114 MS/MS) outside the keratins. Although of unknown function and with little sequence homology to other proteins in GenBank™, it is highly conserved among mammals and appears to be an integral membrane protein according to Kyte-Doolittle hydrophathy plotting (not shown).

The high abundance of keratins in the detergent-insoluble material in the hair shaft probably reflects their well known and extensive interactions with other cellular components, including those in intercellular junctions and even the nuclear periphery (33, 43). Because the keratins are largely extractable from cells of the cortex without solubilizing the cell borders (6), keratins in the insoluble cross-linked material must be derived largely from cells of the cuticle and medulla. Which proteins are actually connected directly is not known, but keratins are likely connected to a large number of their neighbors and may even require other proteins, particularly Gln donors, to participate extensively in the cross-linking through Lys residues. Isolating cross-linked peptides would permit assignment of partners and their sites of interaction. This information would also be of interest in view of the abundance of chromatin proteins such as histones H2A and H2B in the insoluble material. The finding that keratins *in situ* can be cross-linked chemically to DNA (44) raises the question whether they could be cross-linked enzymatically to nuclear matrix or histones. Although a responsible isozyme has not yet been identified, transglutaminase cross-linking in the nucleus likely accounts for the nuclear deposits in cells of the medulla, a phenomenon also evident in corneocytes of the nail plate (45).

Although the extraction of cuticle protein is extensive in hair from individuals afflicted with TGM1-negative familial recessive lamellar ichthyosis, considerable material resists extraction (46). This finding suggests that other transglutaminase activity participates in the stabilization process. Observation of TGM3 protein among those identified in all three hair samples implicates this isozyme as an important contributor to the cross-linking. Intuitively because of its largely membrane-bound localization, TGM1 would be expected to cross-link proteins at the cell periphery, whereas the soluble TGM3 would attach cytoplasmic proteins to each other. The com-

plementary action of two (or more) such enzymes with differing substrate specificities would maximize the resulting stabilization. From this perspective, deficiency in TGM1 activity could easily account for the lack of a marginal band at the outer edge of cuticle cells in hair from a TGM1-deficient patient with congenital ichthyosiform erythroderma (47). However, the envelope-like margins of cortical cells observed even in cases of severe lamellar ichthyosis (TGM1-negative) indicate these are not TGM1-dependent. Moreover the activity leading to stabilization of nuclear material remains uncertain because the appearance of medulla cells from such patients has not been reported.

Present findings offer the prospect of several valuable directions for future investigation. Refinements in the present results include improving the yield of peptides from technically difficult proteins such as the KAPs. Although the location in the mature hair of many of the proteins identified can be inferred from information in the living cells, these could be confirmed by immunohistochemistry possibly assisted by preparations of individual cell types (especially cuticle). Because the cross-linking specificity of transglutaminases relies primarily on available Gln residues, these could be subject to mapping to understand whether other proteins in hair serve the function of Gln-rich trichohyalin in the medulla (and involucrin in epidermis) to connect numerous proteins by virtue of their Lys residues or possibly with polyamines (48). Moreover noninvasive diagnostic applications may become evident from examination of hair samples from individuals with certain genetic hair or skin diseases such as the ichthyoses. Specific perturbations in hair structure visible after detergent extraction, as in the case of trichothiodystrophy (46) and congenital ichthyosiform erythroderma (47), suggest in addition that further understanding of hair structure will result.

\* This work was supported in part by United States Public Health Service Grants R01 AR27130, P42 ES04699, and P30 ES05707. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

¶ Both authors contributed equally to this work.

|| To whom correspondence should be addressed: Dept. of Environmental Toxicology, University of California, One Shields Ave., Davis, CA 95616-8588. Tel.: 530-752-5176; Fax: 530-752-3394; E-mail: rhrice@ucdavis.edu.

## REFERENCES

1. Powell, B. C., and Rogers, G. E. (1997) The role of keratin proteins and their genes in the growth, structure and properties of hair, in *Formation and Structure of Human Hair* (Jolles, P., Zahn, H., and Hocker, H., eds) pp. 59-148, BurkhauserVerlag, Basel
2. Bertolino, A. P., and O'Guin, W. M. (2003) Differentiation of the hair shaft, in *Disorders of Hair Growth: Diagnosis and Treatment* (Olsen, E. A., ed), 2nd Ed., pp. 23-48, McGraw-Hill, Medical Publishing Division, New York
3. Chung, S. I., and Folk, J. E. (1972) Transglutaminase from hair follicle of

- guinea pig. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 303–307
4. Harding, H. J., and Rogers, G. E. (1972) Formation of the  $\epsilon$ -( $\gamma$ -glutamyl) lysine cross-link in hair proteins. Investigation of transamidases in hair follicles. *Biochemistry* **11**, 2858–2863
  5. Zahn, H., Fohles, J., Nienhaus, M., Schwan, A., and Spel, M. (1980) Wool as a biological composite structure. *Ind. Eng. Chem. Prod. Res. Dev.* **19**, 496–501
  6. Rice, R. H., Wong, V. J., and Pinkerton, K. E. (1994) Ultrastructural visualization of cross-linked protein features in epidermal appendages. *J. Cell Sci.* **107**, 1985–1992
  7. Steinert, P. M., and Marekov, L. N. (1997) Direct evidence that involucrin is a major early isopeptide cross-linked component of the keratinocyte cornified cell envelope. *J. Biol. Chem.* **272**, 2021–2030
  8. Rice, R. H., Wong, V. J., Williams, M. L., Price, V. H., Hohl, D., Sundberg, J. P., and Pinkerton, K. E. (1999) Hair shaft defects visualized after detergent extraction. *Exp. Dermatol.* **8**, 308–310
  9. Wolters, D. A., Washburn, M. P., and Yates, J. R., III (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**, 5683–5690
  10. Langbein, L., Rogers, M. A., Winter, H., Praetzel, S., Beckhaus, U., Rackwitz, H.-R., and Schweizer, J. (1999) The catalog of human hair keratins. I. Expression of the nine type I members in the hair follicle. *J. Biol. Chem.* **274**, 19874–19884
  11. Whittock, N. V., Wan, H., Morley, S. M., Garzon, M. C., Kristal, L., Hyde, P., McLean, W. H. I., Pulkkinen, L., Uitto, J., Christiano, A. M., Eady, R. A. J., and McGrath, J. A. (2002) Compound heterozygosity for non-sense and mis-sense mutations in desmoplakin underlies skin fragility/woolly hair syndrome. *J. Investig. Dermatol.* **118**, 232–238
  12. Chavanas, S., Bodemer, C., Rochat, A., Hamel-Teillac, D., Ali, M., Irvine, A. D., Bonafé, J.-L., Wilkinson, J., Taieb, A., Barrandon, Y., Harper, J. I., de Prost, Y., and Hovnanian, A. (2000) Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome. *Nat. Genet.* **25**, 141–142
  13. Schiffman, G. (1966) Immunological methods for characterizing polysaccharides. *Methods Enzymol.* **8**, 79–85
  14. Rice, R. H., Means, G. E., and Brown, W. D. (1977) Stabilization of bovine trypsin by reductive methylation. *Biochim. Biophys. Acta* **492**, 316–321
  15. Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J., and Yates, J. R., III (1998) Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal. Biochem.* **263**, 93–101
  16. Eng, J. K., McCormack, A. L., and Yates, J. R., III (1994) 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
  17. Tabb, D. L., McDonald, W. H., and Yates, J. R., III (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21–26
  18. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., III (1999) Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682
  19. Chen, Y., Kwon, S. W., Kim, S. C., and Zhao, Y. (2005) Integrated approach for manual evaluation of peptides identified by searching protein sequence databases with tandem mass spectra. *J. Proteome Res.* **4**, 998–1005
  20. Paizs, B., and Suhai, S. (2005) Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **24**, 508–548
  21. Liu, H., Sadygov, R. G., and Yates, J. R., III (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201
  22. Yin, T., and Green, K. J. (2004) Regulation of desmosome assembly and adhesion. *Semin. Cell Dev. Biol.* **15**, 665–677
  23. Morrison, D. G., Bansal, M. P., Kirtirel, F., and Medina, D. (1989) Distribution of selenium-binding proteins in different tissues of the mouse and rat. *In Vivo* **3**, 167–172
  24. Kizawa, K., Toyoda, M., Ito, M., and Morohashi, M. (2005) Aberrantly differentiated cells in benign pilomatixoma reflect the normal hair follicle: immunohistochemical analysis of Ca<sup>2+</sup>-binding S100A2, S100A3 and S100A6 proteins. *Br. J. Dermatol.* **152**, 314–320
  25. Pietas, A., Schluns, K., Marenholz, I., Schafer, B. W., Heizmann, C. W., and Petersen, I. (2002) Molecular cloning and characterization of the human S100A14 gene encoding a novel member of the S100 family. *Genomics* **79**, 513–522
  26. Ku, N.-O., Zhou, X., Toivola, D. M., and Omary, M. B. (1999) The cytoskeleton of digestive epithelia in health and disease. *Am. J. Physiol.* **277**, G1108–G1137
  27. Wilker, E., and Yaffe, M. B. (2004) 14-3-3 proteins—a focus on cancer and human disease. *J. Mol. Cell. Cardiol.* **37**, 633–642
  28. Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S. P. (2003) A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* **21**, 921–926
  29. Freitas, M. A., Sklenar, A. R., and Parthun, M. R. (2004) Application of mass spectrometry to the identification and quantification of histone post-translational modifications. *J. Cell. Biochem.* **92**, 691–700
  30. Pesavento, J. J., Kim, Y. B., Taylor, G. K., and Kelleher, N. L. (2004) Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry. *J. Am. Chem. Soc.* **126**, 3386–3387
  31. Cobb, J. A., and Roberts, D. M. (2000) Structural requirements for N-trimethylation of lysine 115 of calmodulin. *J. Biol. Chem.* **275**, 18969–18975
  32. Hake, S. B., Garcia, B. A., Duncan, E. M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D. P., Allis, C. D., and Hunt, D. F. (2006) Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J. Biol. Chem.* **281**, 559–568
  33. Coulombe, P. A., and Omary, M. B. (2002) ‘Hard’ and ‘soft’ principles defining the structure, function and regulation of keratin intermediate filaments. *Curr. Opin. Cell Biol.* **14**, 110–122
  34. Reichert, U., Michel, S., and Schmidt, R. (1993) The cornified envelope: a key structure of terminally differentiating keratinocytes, in *Molecular Biology of the Skin: The Keratinocyte* (Darmon, M., and Blumenberg, M., eds) pp. 107–150, Academic Press, San Diego, CA
  35. Candi, E., Tarcsa, E., DiGiovanna, J. J., Compton, J. G., Marekov, L. N., and Steinert, P. M. (1998) A highly conserved lysine residue on the head domain of type II keratins is essential for the attachment of keratin intermediate filaments to the cornified cell envelope through isopeptide crosslinking by transglutaminases. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2067–2072
  36. Cheong, J. E. L., Wessagowit, V., and McGrath, J. A. (2005) Molecular abnormalities of the desmosomal protein desmoplakin in human disease. *Exp. Dermatol.* **30**, 261–266
  37. Rothnagel, J. A., and Rogers, G. E. (1986) Trichohyalin, an intermediate filament-associated protein of the hair follicle. *J. Cell Biol.* **102**, 1419–1429
  38. Rice, R. H., Wong, V. J., Pinkerton, K. E., and Sundberg, J. P. (1999) Cross-linked features of mouse pelage hair resistant to detergent extraction. *Anat. Rec.* **254**, 231–237
  39. Kurzen, H., Moll, I., Moll, R., Schafer, S., Simics, E., Amagai, M., Wheelock, M. J., and Franke, W. W. (1998) Compositionally different desmosomes in the various compartments of the human hair follicle. *Differentiation* **63**, 295–304
  40. Donetti, E., Boschini, E., Cerini, A., Selleri, S., Rumio, C., and Barajon, I. (2004) Desmocollin 1 expression and desmosomal remodeling during terminal differentiation of human anagen hair follicle: an electron microscopic study. *Exp. Dermatol.* **13**, 289–297
  41. Getsios, S., Amargo, E. V., Dusek, R. L., Ishii, K., Sheu, L., Gotsel, L. M., and Green, K. J. (2004) Coordinated expression of desmoglein 1 and desmocollin 1 regulates intercellular adhesion. *Differentiation* **72**, 419–433
  42. Hardman, M. J., Liu, K. J., Avilion, A. A., Merritt, A., Brennan, K., Garrod, D. R., and Byrne, C. (2005) Desmosomal cadherin misexpression alters  $\beta$ -catenin stability and epidermal differentiation. *Mol. Cell. Biol.* **25**, 969–978
  43. Djabali, K. (1999) Cytoskeletal proteins connecting intermediate filaments to cytoplasmic and nuclear periphery. *Histol. Histopathol.* **14**, 501–509
  44. Holth, L. T., Chadee, D. N., Spencer, V. A., Samuel, S. K., Safneck, J. R., and Davie, J. R. (1998) Chromatin, nuclear matrix and the cytoskeleton: role of cell structure in neoplastic transformation (review). *Int. J. Oncol.* **13**, 827–837
  45. Rice, R. H., Crumrine, D., Hohl, D., Munro, C. S., and Elias, P. M. (2003) Cross-linked envelopes in nail plate in lamellar ichthyosis. *Br. J. Dermatol.* **149**, 1050–1054

46. Rice, R. H., Wong, V. J., Price, V. H., Hohl, D., and Pinkerton, K. E. (1996) Cuticle cell defects in lamellar ichthyosis hair and anomalous hair shaft syndromes visualized after detergent extraction. *Anat. Rec.* **246**, 433–440
47. Rice, R. H., Crumrine, D., Uchida, Y., Gruber, R., and Elias, P. M. (2005) Structural changes in epidermal scale and appendages as indicators of defective TGM1 activity. *Arch. Dermatol. Res.* **297**, 127–133
48. Piacentini, M., Martinet, N., Beninati, S., and Folk, J. E. (1988) Free and protein-conjugated polyamines in mouse epidermal cells. *J. Biol. Chem.* **263**, 3790–3794
49. Taylor, S. W., Fahy, S., Zhang, B., Glenn, G. M., Warnock, D. E., Wiley, S., Murphy, A. N., Gaucher, S. P., Capaldi, R. A., Gibson, B. W., and Ghosh, S. S. (2003) Characterization of the human heart mitochondrial proteome. *Nat. Biotechnol.* **21**, 281–286