

Unique amino acid signatures that are evolutionarily conserved distinguish simple-type, epidermal and hair keratins

Pavel Strnad^{1,*}, Valentyn Usachov¹, Cedric Debes², Frauke Gräter², David A. D. Parry³ and M. Bishr Omary^{4,*}

¹Department of Internal Medicine I, Center for Internal Medicine, University Medical Center Ulm, Albert-Einstein-Allee 23, D-89081 Ulm, Germany

²Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

³Institute of Fundamental Sciences and the Riddet Institute, Massey University, Palmerston North 4442, New Zealand

⁴Department of Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, MI 48109 USA

*Authors for correspondence (pavel.strnad@uniklinik-ulm.de; mbishr@umich.edu)

Accepted 22 August 2011

Journal of Cell Science 124, 4221–4232

© 2011. Published by The Company of Biologists Ltd

doi: 10.1242/jcs.089516

Summary

Keratins (Ks) consist of central α -helical rod domains that are flanked by non- α -helical head and tail domains. The cellular abundance of keratins, coupled with their selective cell expression patterns, suggests that they diversified to fulfill tissue-specific functions although the primary structure differences between them have not been comprehensively compared. We analyzed keratin sequences from many species: K1, K2, K5, K9, K10, K14 were studied as representatives of epidermal keratins, and compared with K7, K8, K18, K19, K20 and K31, K35, K81, K85, K86, which represent simple-type (single-layered or glandular) epithelial and hair keratins, respectively. We show that keratin domains have striking differences in their amino acids. There are many cysteines in hair keratins but only a small number in epidermal keratins and rare or none in simple-type keratins. The heads and/or tails of epidermal keratins are glycine and phenylalanine rich but alanine poor, whereas parallel domains of hair keratins are abundant in prolines, and those of simple-type epithelial keratins are enriched in acidic and/or basic residues. The observed differences between simple-type, epidermal and hair keratins are highly conserved throughout evolution. Cysteines and histidines, which are infrequent keratin amino acids, are involved in de novo mutations that are markedly overrepresented in keratins. Hence, keratins have evolutionarily conserved and domain-selectively enriched amino acids including glycine and phenylalanine (epidermal), cysteine and proline (hair), and basic and acidic (simple-type epithelial), which reflect unique functions related to structural flexibility, rigidity and solubility, respectively. Our findings also support the importance of human keratin ‘mutation hotspot’ residues and their wild-type counterparts.

Key words: Keratins, Amino acids, Skin, Glandular epithelium

Introduction

The cytoskeleton of eukaryotic cells consists of three filamentous arrays – the microtubules, the actin-containing microfilaments and the intermediate filaments (IFs) (Ku et al., 1999; Alberts et al., 2008). IFs are encoded by more than 70 unique genes, which are typically specifically expressed in unique cell types, whereas microfilaments and microtubules are far more widespread in their expression profiles (Moll et al., 1982; Sun et al., 1984; Ku et al., 1999; Eriksson et al., 2009). For example, desmin, vimentin and neurofilament chains are the major IFs in muscle, mesenchymal and neuronal cells, respectively, whereas keratins serve as markers of epithelial lineages (Coulombe and Omary, 2002; Eriksson et al., 2009). IF subunits have a tripartite structure made up of a central α -helical rod domain that is flanked by non-helical head and tail segments (Geisler et al., 1982; Crewther et al., 1983; Parry et al., 2007; Herrmann et al., 2009). The rod domains are largely but not totally responsible for the polymerization of the IF subunits into filamentous structures through coiled-coil formation and, as such, represent the most highly conserved IF region (length ~310 amino acids) (Parry et al., 1985; Parry and Steinert, 1995; Herrmann et al., 2009). By contrast, the head and tail domains exhibit large variations in

sequence and size and are thought to modify the interaction of IF subunits through post-translational modifications. The flanking head and tail domains are the preferential regions that undergo these modifications (Omary et al., 1998; Omary et al., 2006; Hyder et al., 2008), albeit rod domain IF post-translational modifications (e.g. sumoylation) are emerging (Zhang and Sarge, 2008; Snider et al., 2011). Accordingly, the rod domain has probably evolved from a common ancestor through conservative substitutions and a deletion of a six heptad subdomain, whereas the terminal domains have probably arisen through tandem duplications. Consequently, when compared on an inter-species basis, the chains tend to exhibit conservation in length rather than in primary structure (Klinge et al., 1987).

The largest family of filament-forming IF chains are the keratins (Ks). There are 54 unique keratin genes, which encode polypeptides that can be subdivided into type I (K9–K20, K23–K28, K31–K40) and type II keratins (K1–K8, K71–K86). Both chain types are expressed in a cell-specific manner (Coulombe and Omary, 2002; Gu and Coulombe, 2007; Omary et al., 2009). During filament formation, the type I and II keratins assemble in an equimolar ratio to build obligate heteropolymers (Quinlan et al., 1984; Parry et al., 1985; Steinert, 1990; Hatzfeld and

Weber, 1990; Herrling and Sparrow, 1991; Coulombe and Omary, 2002) and, therefore, each cell type has a characteristic type I-type II keratin expression pattern (Moll et al., 2008). For example, simple (single-layered and glandular) epithelia typically contain K8 and K18 chains with variable amounts of K7, K19 and K20, whereas the major keratin pairs (heterodimers) in the epidermis, K5/K14 and K1/K10, are in the basal and suprabasal layers, respectively (Moll et al., 2008). Twenty-six keratins including K31, K35, K81, K85 and K86 are largely restricted to the hair follicle (Schweizer et al., 2007). In addition, minor keratin amounts are found in non-epithelial cells (Knapp and Franke, 1989). Furthermore, there is increasing evidence of homodimer formation (possibly transient) as well as promiscuity in the pairing of type I and II chains (Hatzfeld and Franke, 1985; Quinlan et al., 1986; Smith and Parry, 2007; Langbein et al., 2010). However, the functional significance of keratin homodimers is unclear.

The abundance of keratin polypeptides, coupled with their cell-specific expression suggests that keratins have evolved to fulfill specific requirements for specialized epithelial tissues. To support this hypothesis, several studies have demonstrated that the loss of a particular keratin polypeptide is only partially compensated for by the expression of an alternative keratin. For example, ectopic expression of K16 or K18 only partially rescues the epidermal phenotype observed in K10 knockout mice (Hutton et al., 1998; Paladini and Coulombe, 1999). However, K19 can rescue K18 loss (Hesse et al., 2000; Hesse et al., 2005), and replacement of the K14 with a chimeric keratin that consists of the K14 rod domain and the K10 head and tail domains predisposed mice to an accelerated development of skin cancer (Chen et al., 2006). Ectopic keratin expression can even lead to a gain-of-function phenotype as shown in mice overexpressing K16 in the skin, epidermal K14 in the liver or K1 in the pancreatic β -cell (Blessing et al., 1993; Takahashi et al., 1994; Albers et al., 1995). In that context, ectopic expression of a keratin pair, which to our knowledge has not been reported to date, is likely to be more physiologically informative.

Despite this and other lines of evidence, the structural differences between the various keratins have not been clearly defined, although the sequence characteristics of some individual chains have been reported previously [for summaries see the following references (Steinert et al., 1985; Parry and North, 1998; Parry and Steinert, 1995; Parry, 2005)]. In order to understand more thoroughly the amino acid and structural differences between keratins, we analyzed the sequences of the major keratins that are found in simple-type epithelia, epidermis and hair from a range of species. We demonstrate that keratins from these three tissue types have striking differences in their primary amino acid structure and that these differences are likely to contribute to functional associations in cells and tissues.

Results

In order to determine how the amino acid composition differs in keratin head, rod and tail domains, we first analyzed 25 unique human keratin sequences (Fig. 1; supplementary material Table S1). Our results indicate that glutamate, leucine, alanine and glutamine are much more abundant in the rod domain, and serine and glycine are most frequently found in both the head and tail domains (Fig. 1; $P<0.001$ for all comparisons). Lysine and arginine are also common, particularly in both rod and tail (Fig. 1). There were also significant differences in amino acid

composition when comparing type I and type II keratins, but these were mostly restricted to specific domains rather than the overall keratin sequence (supplementary material Table S3). For example, type II keratins were found to have a higher content of basic amino acids in the head and rod domains, but a much lower content in the tail. Similar findings were noted for specific amino acids. For example, type II keratins have a higher isoleucine content in the head domain compared with the type I keratins, whereas both keratin subtypes have comparable isoleucine levels in their rods and tails (supplementary material Table S3). Tryptophan is particularly uncommon in keratins (Fig. 1; supplementary material Table S3), even when considering the fact that it is a rare amino acid that makes up $\sim 1.2\%$ of total amino acids (http://www.ncbi.nlm.nih.gov/Class/Structure/aa_aa_explorer.cgi). Of note, tryptophan is completely absent from heads and tails of type II keratins.

For further analysis, keratins were divided into subgroups to reflect their tissue-specific expression. To that end, K1, K2, K5, K9, K10 and K14 were analyzed as prototype epidermal keratins, whereas K7, K8, K18, K19, K20 and K31, K35, K81, K85, K86 represented the simple-type (i.e. single layer) epithelial and hair keratins, respectively. We did not separately analyze epidermal pairs (e.g. K1/K10 versus K5/14) in order to allow us to work with larger categories of keratins. When the sizes of keratin polypeptides were compared, the epidermal keratins proved to be significantly ($P<0.01$) longer than both the simple-type epithelial and the hair keratins – a direct consequence of their much longer head and tail domains (supplementary material Fig. S1).

We also analyzed whether amino acid composition differed between human keratin subtypes. Although only minor differences were noted in the overall percentage of amino acid subgroups such as acidic, basic or aromatic amino acids (Fig. 2; supplementary material Table S4), some amino acids did differ substantially between the three major keratin subtypes we examined (Fig. 3). For example, hair keratins have many cysteines, but there are significantly fewer in epidermal keratins and they are almost never found in simple-type epithelial keratins ($P<0.0001$ for both comparisons). In particular, the prototypic simple-type epithelial human keratins K8/K18 together with K19 contain no cysteine residues, whereas K7 and K20 have only one cysteine each (supplementary material Table S1). By contrast, glycines and glutamates were, respectively, the most and least abundant amino acids in epidermal keratins ($P<0.005$ for all comparisons; Fig. 3; supplementary material Table S8). K1/K10 (the keratin pair expressed in suprabasal epidermal layers) were particularly glycine rich and alanine poor, whereas K5 and K14 (the keratins of basal epidermal layers) were more similar to simple-type epithelial keratins with respect to their alanine and glycine contents (supplementary material Table S1).

Given that the amino acid composition differs in keratin heads, rods and tails (Fig. 1), we analyzed which keratin domains were particularly responsible for the observed differences in the overall amino acid content. Notably, distinct variations were observed within both heads and tails, but no striking differences were seen in the rod domains (Fig. 2; supplementary material Tables S5–S8). The only exception to this ‘rule’ was cysteine, which was fairly scarce in all simple-type epithelial and/or epidermal keratin domains when compared with hair keratins (Fig. 4B). However, the lower

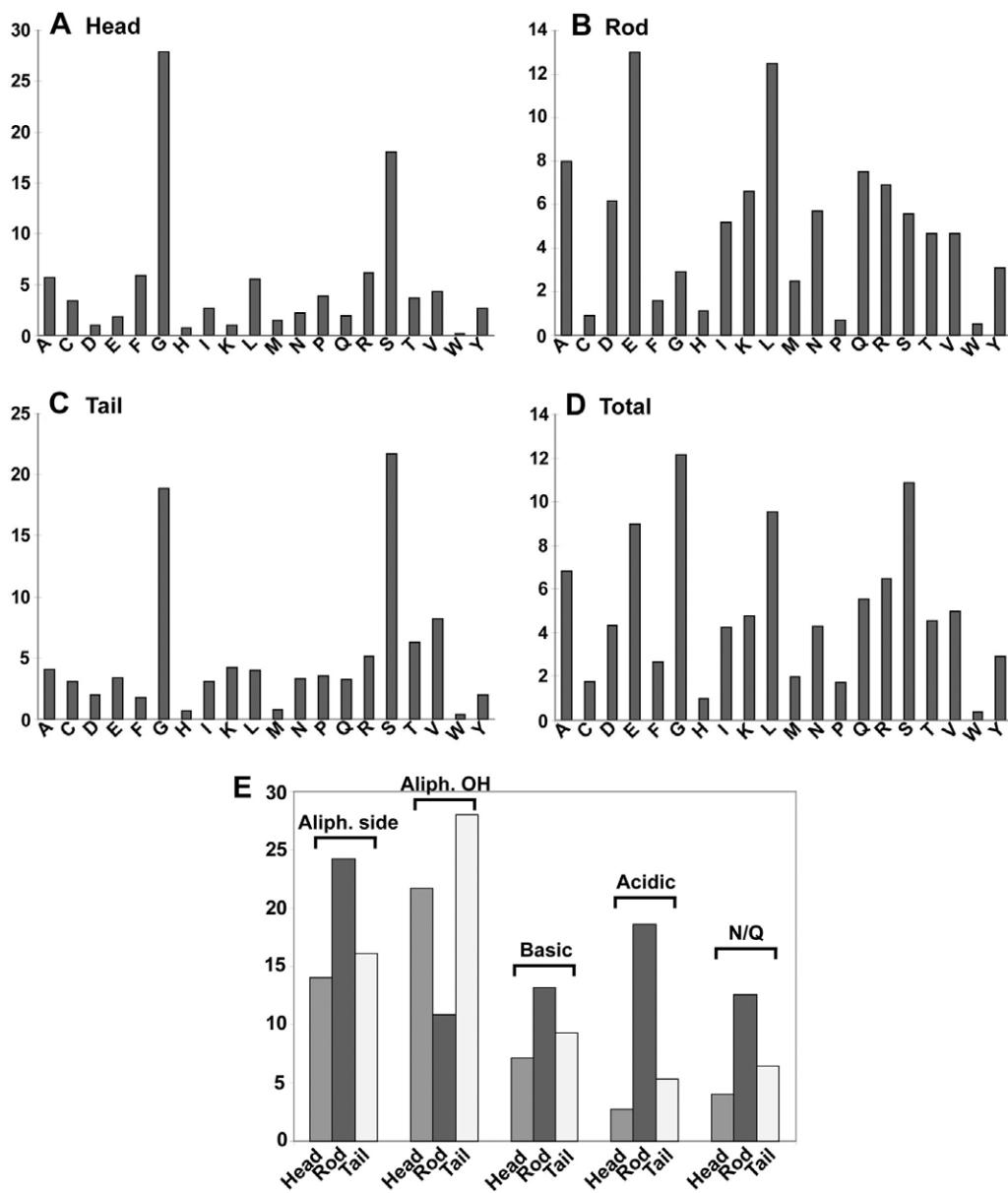


Fig. 1. Amino acid composition differs significantly between the keratin head, rod and tail domains. To study the amino acid composition of keratin domains, the human keratin sequences K1–K10, K12–K14, K16–K20, K23, K24, K31, K35, K81, K85 and K86 were used. For each amino acid (denoted by the single letter abbreviation), the percentage found in each domain (**A**, head; **B**, rod and **C**, tail) and in the complete keratin sequence (**D**) was calculated. The various groupings of amino acids used here are based on standard nomenclature (Berg et al., 2007). Note that D, E, I, L, Q are more abundant in the rod domain than in the head or the tail ($P<0.001$ for all comparisons), whereas F and G are more frequent in the head than in rod and tail domains ($P<0.05$ for all comparisons). In the head domain L is rare but is more common in both the rod and tail, whereas S is infrequent in the rod domain but relatively abundant in the head and tail. (**E**) Distribution of the amino acids groups. Amino acids with aliphatic side chains: isoleucine (I), leucine (L), methionine (M), valine (V); amino acids with hydroxyl groups: serine (S), threonine (T); aromatic amino acids: phenylalanine (F), tryptophan (W), tyrosine (Y); basic amino acids: arginine (R), lysine (K); acidic amino acids: aspartate (D), glutamate (E); other amino acids: alanine (A), asparagine (N), cysteine (C), glutamine (Q), glycine (G), histidine (H), proline (P).

alanine and the higher glycine content in epidermal keratins as well as the relative abundance of prolines in the hair keratins was restricted to heads and tails (Fig. 4; supplementary material Tables S5–S8). The differences in glutamate and tyrosine content were noted in the tail domain only (Fig. 4; supplementary material Table S7).

To test whether the differences observed in human keratins were conserved evolutionarily, we analyzed the composition of

selected epidermal, simple-type epithelial and hair keratins in human, cow, mouse, opossum, chicken, frog and zebrafish (supplementary material Table S9). This analysis confirmed that the rareness of cysteine residues in simple-type epithelial keratins was conserved throughout evolution (Fig. 5; supplementary material Table S10). For example, human, mouse and cow K8, K18 and K19 contain no cysteine, whereas only a few cysteines were found in simple-type epithelial sequences from lower

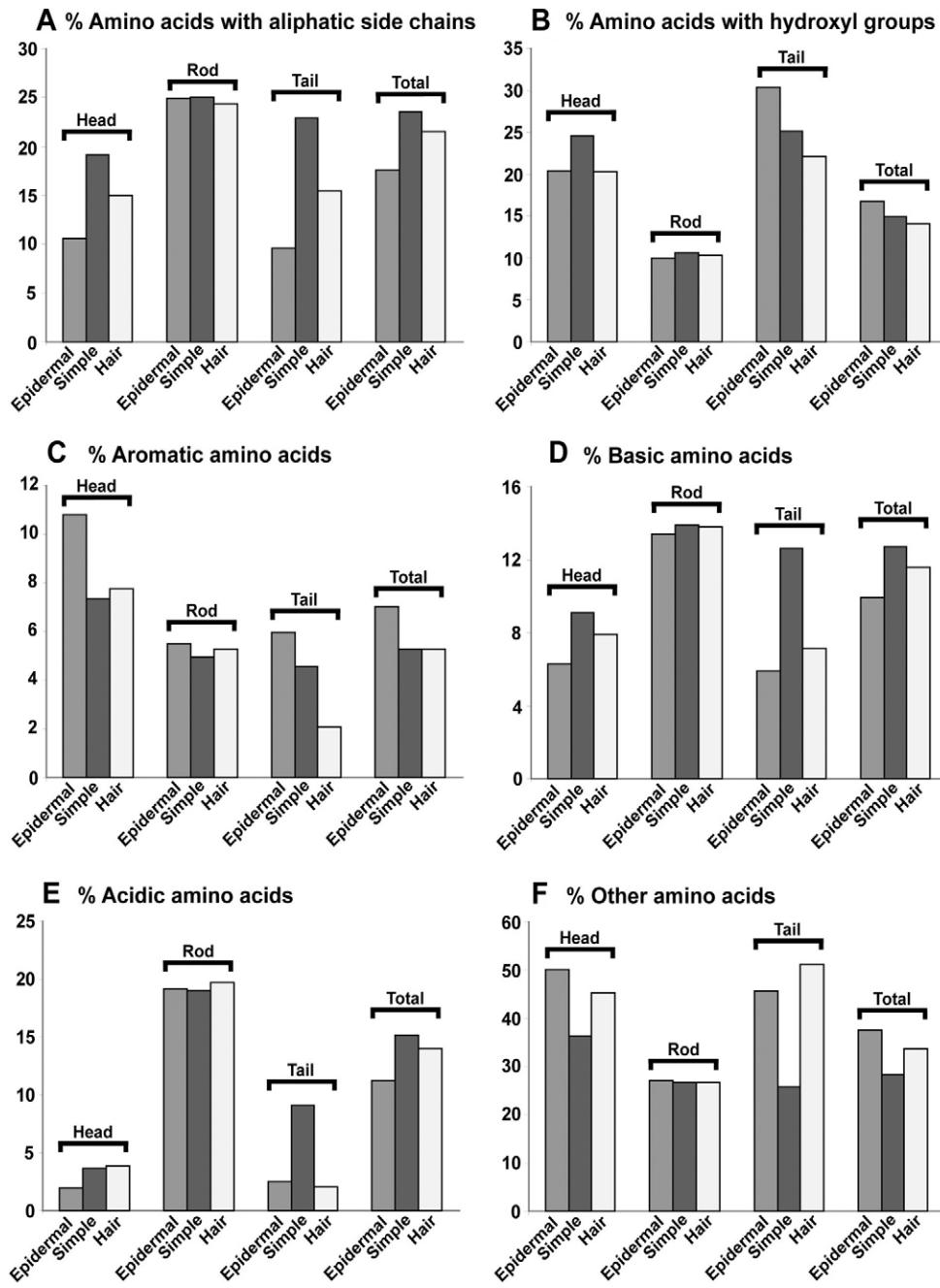


Fig. 2. Amino acid composition of human epidermal, simple-type epithelial and hair keratins. The percentage share of selected amino acid subgroups (A–F) within keratin domains and the whole keratin sequence was calculated. Grouping of amino acids as shown in Fig. 1. K1, K2, K5, K9, K10 and K14 sequences were analyzed for epidermal keratins, and K31, K35, K81, K85, K86 and K8, K9, K18, K19, K20 were used for hair and simple-type epithelial keratins, respectively. There are large differences between the amino acid composition of the head and tail domains, whereas the rods have largely identical compositions. Of note, amino acids with aliphatic side chains and those that are basic are the most and least abundant in heads and tails of simple-type epithelial and epidermal keratins, respectively. The difference in acidic amino acids and amino acids with aliphatic hydroxyl groups is most prominent in the tail domain (aliphatic: epidermal>hair; acidic: simple-type>hair/epidermal), whereas aromatic amino acids are most abundant in the heads and tails of epidermal keratins.

species. Within the analyzed simple-type epithelial keratins, only frog K19 and cow K20 contained two cysteines, and all other sequences had, at most, one cysteine residue (supplementary material Table S10). Similarly, the abundance of glycine within the heads and tails of epidermal keratins was also conserved across the species studied (Fig. 6A; supplementary material Table S11), whereas the higher glycine content of suprabasal K1/K10 compared with the basal K5/K14 was not. For example, in opossum and chicken, K5 contained 105–112 glycines, in contrast to the 44–62 glycines in K1 (supplementary material Table S11). The heads and tails of hair keratins, across all analyzed species, contained more proline than their epidermal and simple-type epithelial counterparts (Fig. 6B; supplementary

material Table S12). Of note, K2, K14 and K20 tails contain no proline in any of the analyzed species (supplementary material Table S12).

The lower abundance of alanine in the heads of epidermal keratins (Fig. 4) was conserved across species, but no obvious differences in alanine content were noted in the tails of non-mammalian species (supplementary material Fig. S2, Table S13). When compared throughout the species, no striking differences were observed between the simple-type epithelial, epidermal and hair keratins with regards to glutamate, phenylalanine and tyrosine contents (supplementary material Tables S14–S16).

The high proportion of acidic amino acids in the tails of simple-type epithelial keratins (Fig. 2) was conserved across

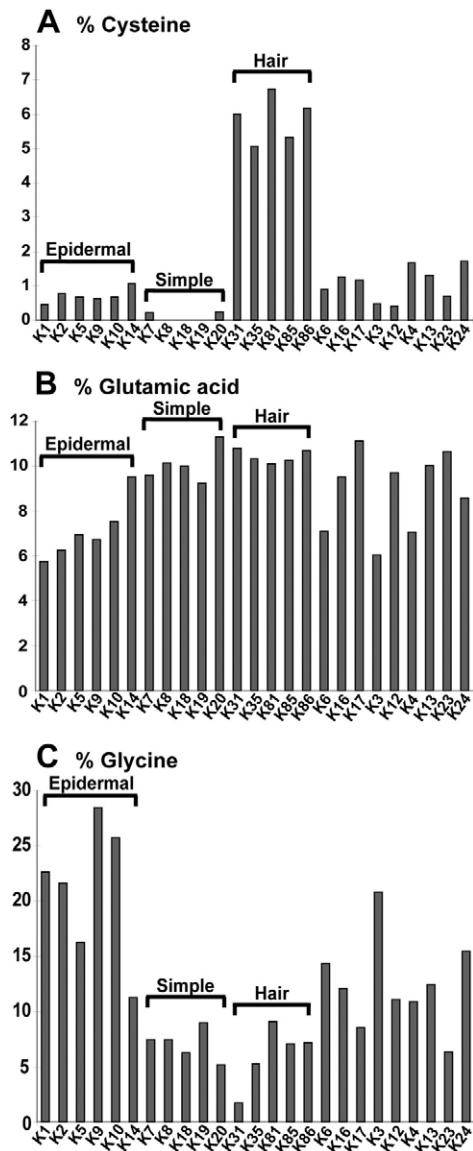


Fig. 3. Percentage of selected amino acids in human keratins. The percentage share of (A) cysteine, (B) glutamic acid and (C) glycine was calculated for the whole keratin sequence. Note that epidermal, simple-type epithelial and hair keratins have different amino acid compositions. For example, cysteine residues are frequent in the hair keratins (i.e. K31, K35, K81, K85 and K86) but rare and almost never present in the epidermal and simple-type epithelial keratins, respectively ($P < 0.0001$ for comparison between hair and simple-type epithelial or epidermal keratins). Compared with other sequence subgroups, glutamic acid is the least abundant, and glycine is the most abundant residue in epidermal keratins. K, keratin.

species (supplementary material Fig. S3, Table S17), with K20 having the most (supplementary material Table S17). Basic amino acids were also more abundant in the tails of simple-type epithelial keratins, and this was a consequence mainly of the basic-amino-acid-rich K18 and K20 chains (supplementary material Fig. S4, Table S18). Aromatic amino acids were more abundant in the heads of epidermal keratins (particularly K9/K10), whereas they were rarely seen in tails of hair keratins (supplementary material Fig. S5, Table S19).

Discussion

We have systematically analyzed selected keratin amino acid sequences in a range of species, and demonstrated that keratins in hair, epidermis and simple-type epithelia have significant variations in their amino acid distributions (Fig. 7). As expected, the differences were primarily in the more variable head and tail domains, whereas the amino acid composition of the rod domain remains relatively uniform and typical of a two-stranded coiled-coil rod (Conway and Parry, 1990; Parry et al., 2008). Despite the largely conserved rod domain structure, previous data indicated that rod domains are not completely interchangeable. For example, the rod domain of K16 contains a proline residue that interferes with tetramer formation (Wawersik et al., 1997). In addition, the tail of K14 was shown to interact with the K5 rod domain, but not with the rod domain of K8 (Lee and Coulombe, 2009), which also suggests that keratin domains are associated rather than autonomous, though this will require further experimental validation.

The keratin rod domains contain many charged residues, but relatively few apolar ones. Indeed, this is expected of any rod-like structure that has a large surface area to internal volume ratio. Such a shape allows the (numerous) charged residues to be located on the surface of the structure and in a position where they can be hydrated while simultaneously shielding the (fewer) apolar residues internally from the aqueous environment. Of note is the relatively high content of charged residues in the short tails of simple-type epithelial keratins, which predestines them to a more cylindrical or extended shape.

Among the unique amino acid differences across the three major keratin categories we analyzed, the most striking difference was seen in the cysteine content. Hair keratins are particularly cysteine rich, which is in good agreement with the limited previously published data on sheep and human keratins (Fraser and Parry, 2007). There is indirect evidence for cysteine crosslinking of hair keratins (Fraser and Parry, 2007) and there is also evidence that esophageal K4 and K13 become disulfide crosslinked during terminal differentiation (Pang et al., 1993). In addition, there is increasing evidence to show that sheep wool acts mechanically as a lightly crosslinked gel, thereby requiring the preferential formation of intramolecular rather than intermolecular disulphide bonds (Parry and Fraser, 1985; Parry and Steinert, 1995). Indeed, it has been shown previously that 97–98% of cysteines in wool form disulphide bridges, thereby executing an important stabilizing function (Fraser et al., 1988). Disulphide bond formation might be further facilitated by the high proline content observed in hair keratins (Fig. 7). Prolines are known to cause the chain to fold back on itself by inducing kinks. They also, as opposed to glycines, provide mechanical rigidity because of the lack of one dihedral degree of freedom in the protein backbone. Of the various keratin subtypes, the hair keratins are the mechanically most challenged. Indeed, one of the primary roles of hair is to protect an animal from mechanical abrasion as it moves through its normal habitat as well as providing a mechanism that allows its overall temperature to be controlled. Therefore, the presence of numerous disulphide bond covalent links gives strength to the hair.

At the other end of the scale, the simple epithelia are the least mechanically challenged keratins. It is tempting to speculate that forced or natural mutant cysteine residues within simple-type epithelial keratins limit their pliability and contribute to the gain-of-function phenotype seen in transgenic mice ectopically

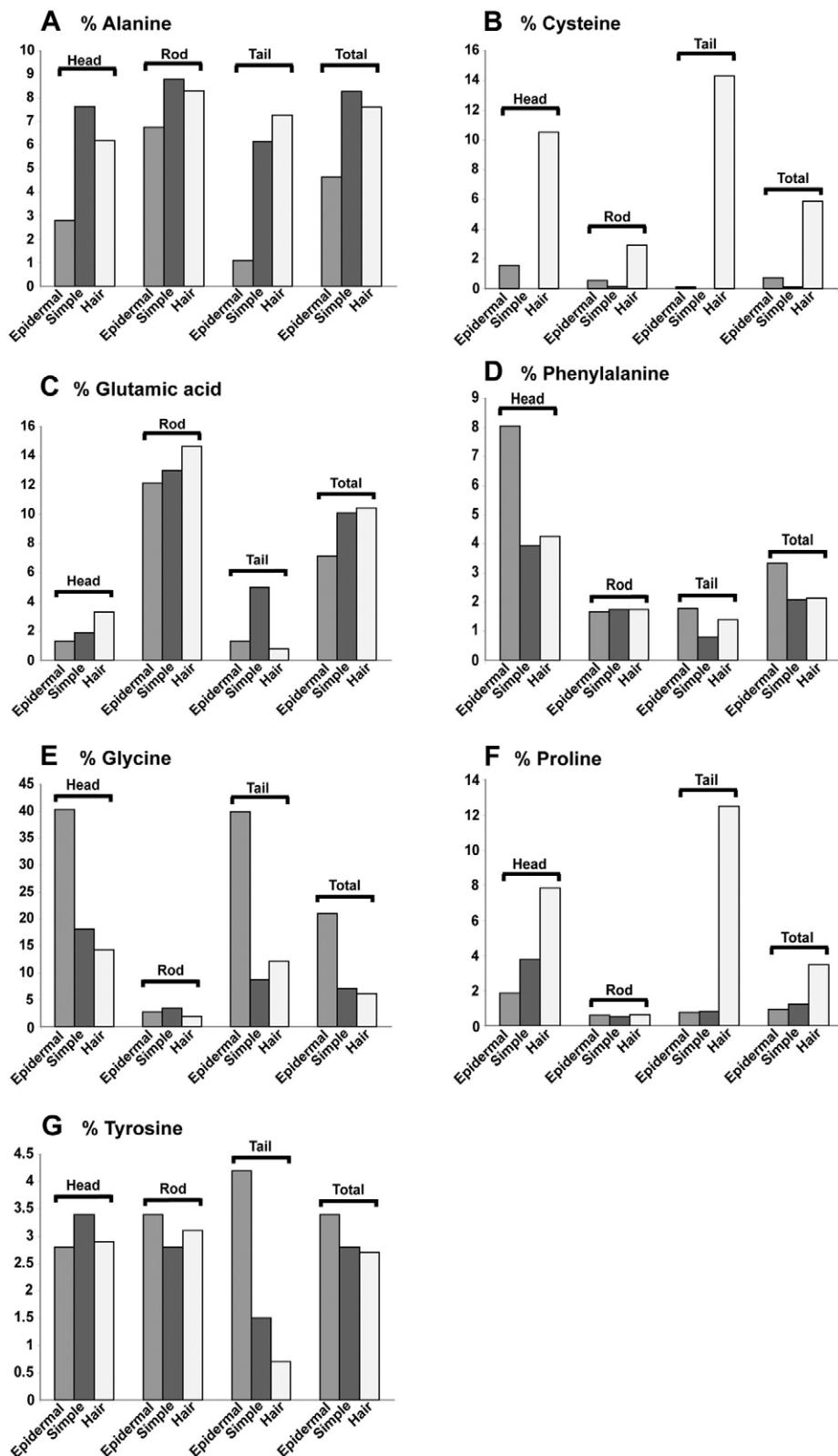


Fig. 4. Distribution of selected amino acids within epidermal, simple-type epithelial and hair keratins. The percentage share of (A) alanine (B) cysteine, (C) glutamic acid, (D) phenylalanine, (E) glycine (F) proline and (G) tyrosine within keratin domains and the whole keratin sequence is shown. Chain designation is as given in Fig. 3. Of note, cysteines are abundant within hair keratins, but rare in epidermal keratins and almost never present in simple-type epithelial keratins. Compared with other keratin subgroups, the heads and tails of epidermal keratins are relatively alanine poor and glycine rich, whereas prolines are most abundant in the heads and tails of hair keratins. Also, phenylalanines are most abundant in the heads of epidermal keratins, whereas no striking differences are seen in other subdomains. Finally, tyrosines are particularly common in tails of epidermal keratins, but less frequent in tails of simple-type epithelial and hair keratins.

expressing epidermal keratins in simple-type epithelial tissues (Albers et al., 1995). This is further supported by the human variants in K8/K18 leading to introduction of cysteines into an otherwise cysteine-free K8. To date, three of these have been

described (Fig. 8A) (Ku et al., 2001; Ku et al., 2005; Tao et al., 2007) and were shown biochemically to lead to disulphide bridge formation (Ku et al., 2001; Ku et al., 2005; Tao et al., 2007), thereby limiting the ability of K8/K18 filaments to reorganize in

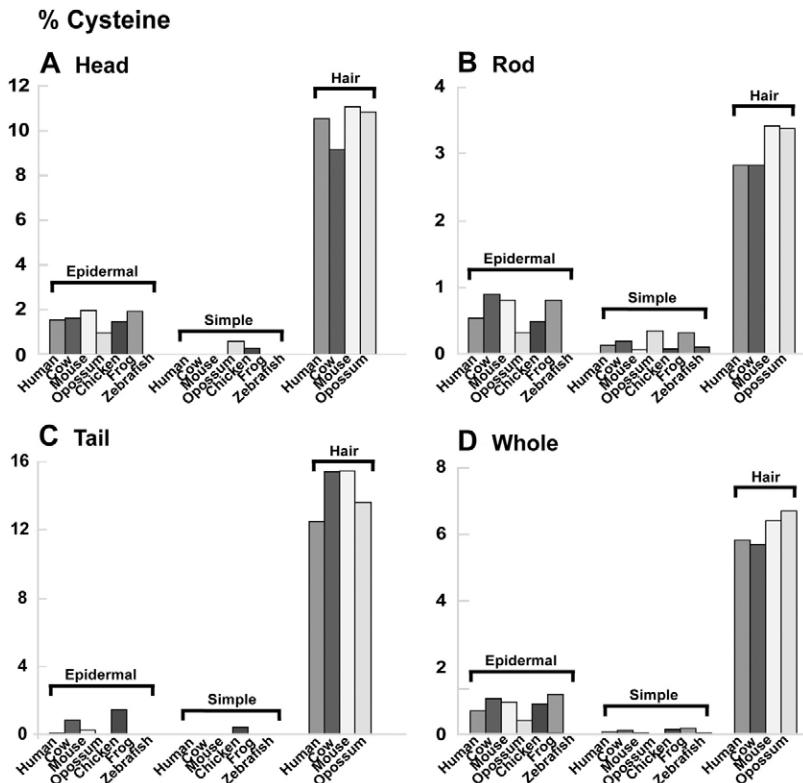


Fig. 5. Cysteine content of epidermal, simple-type epithelial and hair keratins among multiple species. The average cysteine content within keratin domains (A–C) and the whole keratin sequence (D) in selected species. Chain designation is as shown in Fig. 3. Note that irrespective of the analyzed species, hair keratins have a high cysteine content, whereas cysteines are rarely seen in epidermal and almost never in simple-type epithelial keratins.

transfected cells under oxidative stress conditions (Ku et al., 2001). K8 G62C is the most frequently observed cysteine mutation and is associated with the development and progression of acute (Strnad et al., 2010) and chronic (Omáry et al., 2009) human liver disease. In addition, animals overexpressing K8 G62C are susceptible to Fas-induced liver apoptosis, probably because of a conformational change (Tao et al., 2006) that interferes with K8 phosphorylation at Ser74 and thereby shunts phosphorylation to other pro-apoptotic proteins, thus promoting apoptosis (Ku and Omáry, 2006).

Of note, the relative scarcity of cysteine residues is not restricted to simple-type epithelial keratins but is observed in all human epithelial keratins (relative frequency <2% in all analyzed polypeptides; supplementary material Table S1B), suggesting that increased levels of cysteine might be ‘toxic’ to epithelial keratins. This is supported by the observation that de novo cysteine introduction is seen in ~5% of disease-associated keratin missense mutations, whereas loss of a cysteine residue is comparatively rare (Fig. 8A). When the number of reported individuals is considered, the pathogenic role of cysteine becomes even more obvious, as it represents the second most commonly de novo introduced amino acid (www.interfil.org). This is mainly due to the K14 R125C variant, which is responsible for ~40% of epidermolysis bullosa simplex cases (Owens and Lane, 2004; Coulombe et al., 2009). The abundance of this K14 R125C variant, which causes a very severe disease phenotype (Owens and Lane, 2004; Coulombe et al., 2009), is at least in part due to its location in a methylated CpG (C-phosphate–G) DNA sequence (Sommer, 1992; Mohrenweiser, 1994; Pfeifer, 2006). The molecular consequences of K14 R125C/H variants still remain to be elucidated. Although these

variants do not interfere with early assembly stages or obstruct the formation of IFs, they affect higher order structure formation (Ma et al., 2001; Herrmann et al., 2002).

In the case of K18, the introduction of a mutation (K19 R90C) that is homologous to the K14 R125C is similarly detrimental because it leads to keratin network disruption, development of mild chronic hepatitis and to susceptibility to toxin-induced liver injury (Ku et al., 2007), but part of the phenotype can be rescued by supplementation with a wild-type K18 allele (Hesse et al., 2007). Therefore, better understanding of the keratin amino acid composition is likely to help predict the significance of newly described genetic variants. For example, human keratins are generally histidine poor (supplementary material Table S1), whereas a de novo histidine introduction is the most commonly noted disease-associated substitution (www.interfil.org), suggesting that histidine might constitute another disrupting amino acid when introduced into keratins. In support of this, the mutations that lead to a new histidine markedly outnumber the mutations that result in a loss of histidine (Fig. 8B). Specific examples include K8 Y54H, which is found in patients with liver disease, and when transfected into cells it destabilizes keratin filaments after exposure to heat or okadaic acid stress (Ku et al., 2001). Similarly, K10 N154H causes epidermolytic hyperkeratosis and destabilizes filament formation in vitro (Chipev et al., 1994). Also, K14 R125H causes severe epidermolysis bullosa simplex and interferes with the formation of higher order structures (Ma et al., 2001; Herrmann et al., 2002).

Tryptophan is another amino acid that is rarely found in keratins. However, unlike the situation of cysteine and histidine, amino acid substitutions leading to de novo tryptophan are uncommon. The only described new tryptophan mutation to date,

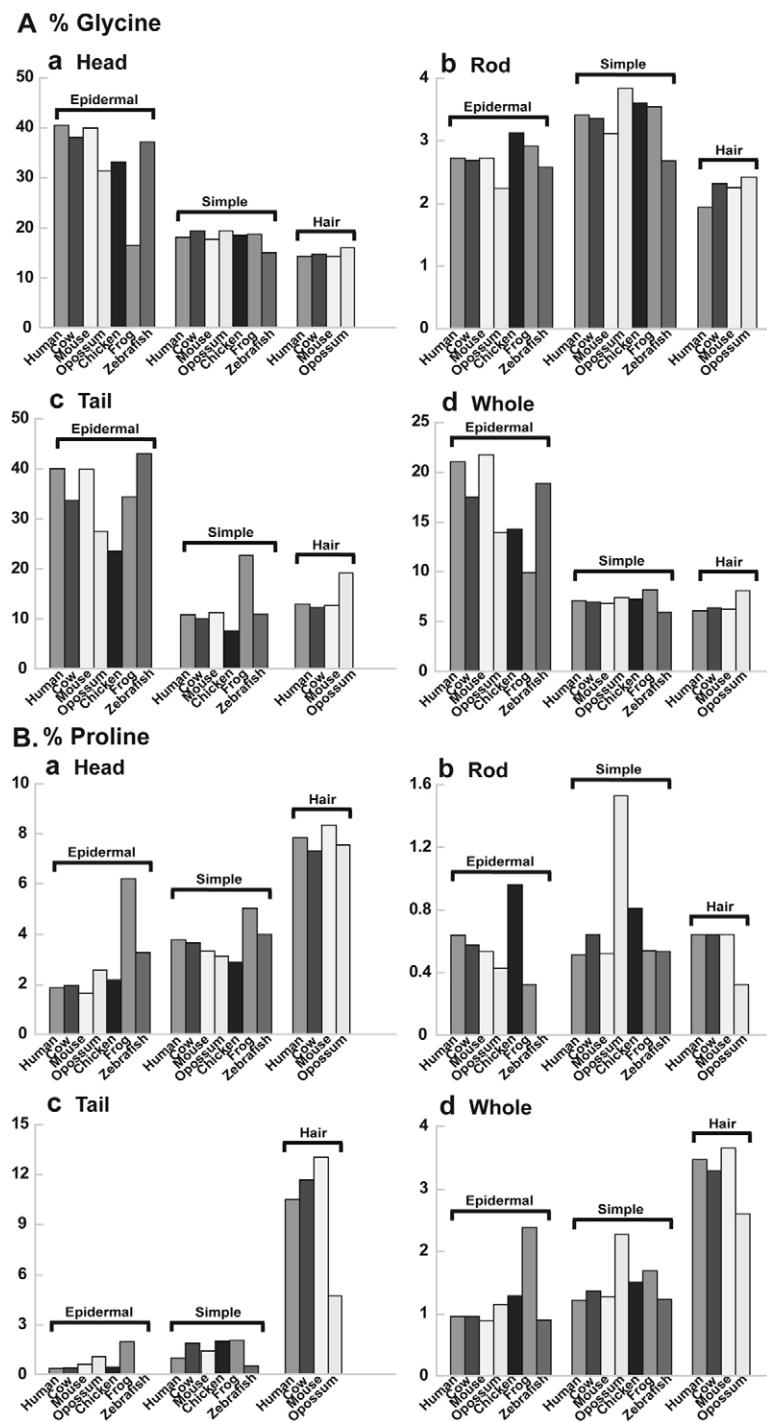


Fig. 6. Glycine and proline content of epidermal, simple-type epithelial and hair keratins in a range of species. The average glycine (A) and proline (B) content within keratin domains and the whole keratin sequence in selected species. Chain designation is as shown in Fig. 3. Note that, independent of the analyzed species, glycines are most abundant in the heads and tails of epidermal keratins, and prolines are found most frequently in the heads and tails of hair keratins.

K9 R162W, is a prevalent genetic defect reported in humans with epidermolytic palmoplantar keratoderma (Navsaria et al., 1995). Further data are needed to evaluate potential gain- or loss-of-function effects of such de novo tryptophan residues.

In addition to primary amino acid structure, unique post-translational modifications could contribute to specific functions of keratin subtypes. We analyzed selected simple epithelial keratin phosphorylation sites (K8 S24, K8 S74, K18 S34 and K18 S53), given that they represent the major and best characterized

post-translational modifications of keratins (Omari et al., 2006). Some of these sites are relatively conserved (as serine/threonine) across some epidermal and hair keratins (K8 S24 and S74) but others (e.g. K18 S34 and S53) are not (Omari et al., 2006). We also examined whether the higher frequency of basic amino acids in simple epithelial keratins (Fig. 2D) might represent higher [R/K]x(x)S/T phosphorylated motifs, but this is not the case (not shown). Therefore, we hypothesize that the high frequency of charged amino acids in the head and tail domains plays a role in

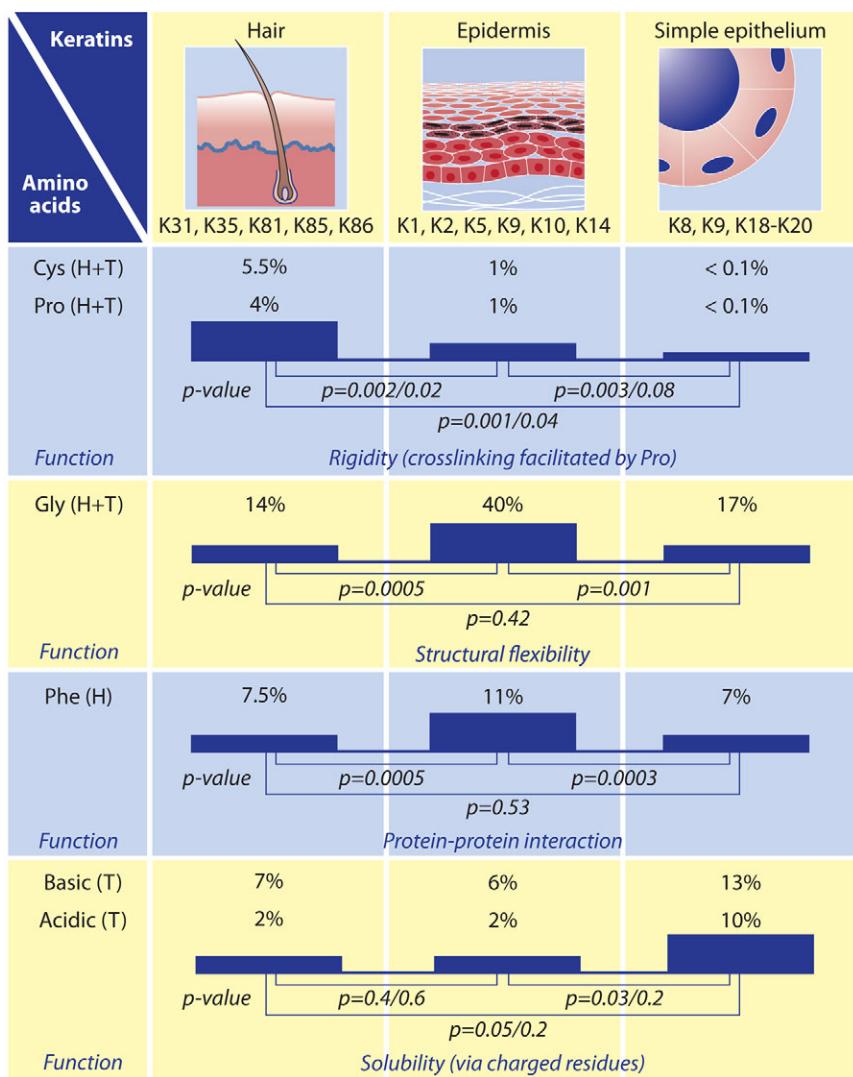


Fig. 7. Summary of the major differences in amino acid composition of hair, epidermal and simple-type epithelial keratin structural domains. The schematic summarizes the major differences in amino acid content of hair, epidermal and simple-type epithelial keratins together with their respective *P*-values. Note that different amino acid content is seen predominantly in the keratin head (H) and tail (T) domains, whereas rods have a largely conserved amino acid composition. The different amino acid content is likely to affect keratin rigidity [through cysteine (Cys) crosslinking, which is facilitated by proline (Pro) residues], structural flexibility [through increased glycine (Gly) content], protein–protein interactions (through phenylalanine (Phe) content) as well as keratin solubility (through the presence of charged residues). The content of charged residues was determined as a sum of acidic (aspartic acid, glutamic acid) and basic (arginine, lysine) residues. In the first row, the first and second *P*-value refers to cysteine and proline residues, respectively, and in the last row, *P*-values refer to the differences in content of basic and acidic amino acids, respectively.

the increased solubility of simple epithelial keratins (Fig. 7) compared with epidermal and hair keratins (Omary et al., 1998), but this needs to be tested experimentally. Additional layers of complexity not addressed by our analysis include the possibility that specific keratin residues can provide unique functional platforms that regulate post-translational modifications and interaction with associated proteins.

Epidermal keratin heads and tails were shown to contain high amounts of glycine and aromatic residues. Such a composition is consistent with the glycine loop hypothesis proposed by Steinert and colleagues for K1/K10 and would be expected to lead to formation of a highly flexible, partly disordered, apolar structure (Fig. 7) (Steinert et al., 1991). In support of this, a high degree of conformational flexibility of the epidermal heads and tails was previously demonstrated using NMR (Mack et al., 1988). The epidermis must be a flexible structure while acting as a key part of the barrier between internal organs of the animal and its environment. The partially unstructured glycine-rich heads and tails of the epidermal keratins would be able to readily interact between themselves in many ways but always through the formation

of strong non-specific apolar interactions facilitated by the ability of these domains to adopt a variety of conformations. By contrast, simple-type epithelial keratins have a high content of charged and aliphatic residues and very low cysteine content. This composition is consistent with a globular structure that is not as excessively flexible as epidermal keratins or as rigid as those in hair. In conclusion, the comparison of amino acid sequences of keratins revealed striking differences between epidermal, simple-type epithelial and hair keratins (Fig. 7) and provided insights into the adaptive evolution of keratins that took place to fulfill the diverse functional requirements of various tissues. Further studies are needed to analyze whether similar principles apply to the evolution of non-keratin IFs, which also show a marked tissue-specific distribution.

Materials and Methods

The amino acid composition of human keratins was determined using the current NCBI reference sequences (for accession numbers, see supplementary material Table S1A). The segmentation of the sequences into head, rod and tail domains was based on data available at www.interfil.org. For each domain, as well as the whole keratin sequence, the absolute and relative amino acid compositions were determined. K1, K2, K5, K9, K10, K14 were analyzed as the major representatives

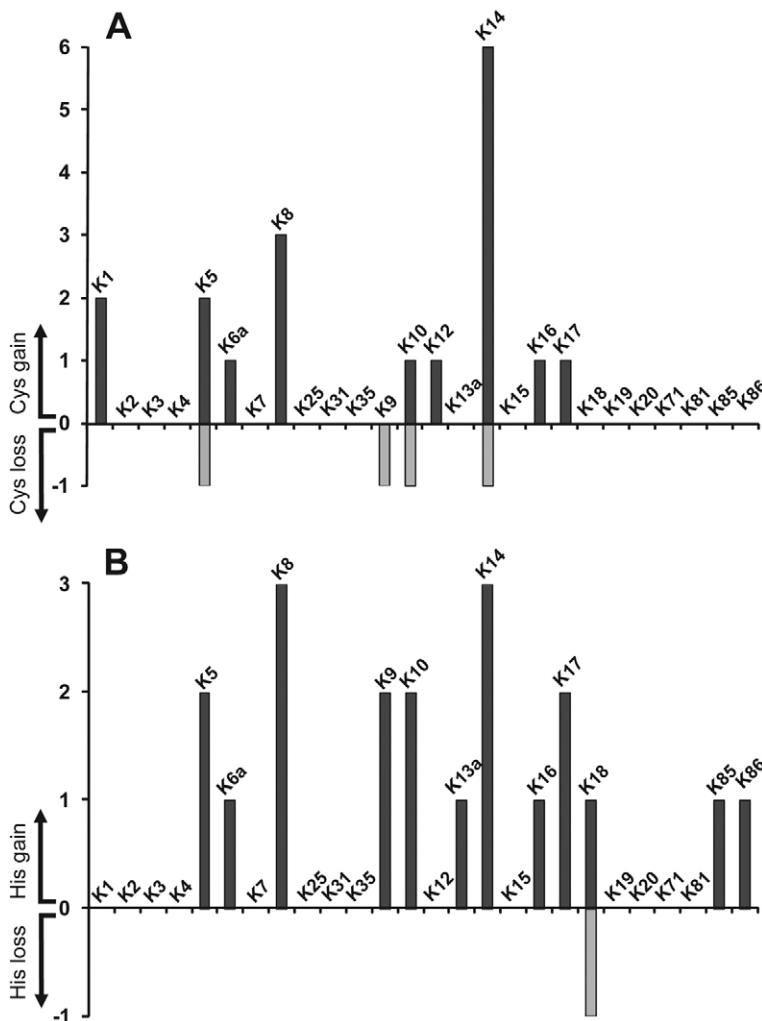


Fig. 8. Overview of missense substitutions within selected human keratins and the resulting loss or gain of cysteine (A) and histidine (B) residues. Data are based on the Human Intermediate Filament database (www.interfil.org), which lists 184 and 169 unique missense substitutions for the analyzed type I and type II keratins, respectively.

of epidermal keratins, and K7, K8, K18, K19, K20 and K31, K35, K81, K85, K86 were studied as representatives of simple-type epithelial and hair keratins, respectively. Relative amino acid compositions were used to compute the average amino acid composition for each keratin subgroup. Amino acids were further divided into acidic, basic and other groupings on the basis of standard nomenclature (Berg et al., 2007).

To compare the amino acid composition in a range of species, sequences from human, cow, mouse, opossum, chicken, frog and zebrafish were chosen because they represent divergent species for which a sufficient quantity of amino acid sequence data has been deposited in the NCBI database. Reference sequences were used wherever available (for accession numbers, see supplementary material Table S9). For every species and every keratin, an alignment with its human counterpart was performed and the sequence was used for further analysis only when the similarity was at least 60% over at least half of the corresponding human sequence. Sequences lacking either head or tail domains were excluded from further analysis (supplementary material Table S2). The amino acid composition for a given keratin subgroup was then computed as an average of the relative amino acid composition of all the keratin sequences remaining.

To determine the level of statistical significance, a two-tailed Student's *t*-test was used and $P < 0.05$ were considered statistically significant. The graphs were plotted with Microsoft Office Excel software (version 2003; Microsoft Corporation) and were further modified and assembled into figures using Adobe Illustrator CS3 software (version 13.0.0; Adobe Systems Incorporated).

Funding

This work was supported by the National Institutes of Health [grant numbers DK52951, DK47918 to M.B.O.]; the Department of Veterans Affairs [Affairs Merit Award to M.B.O.]; the Emmy Noether Program of the German Research Foundation [grant number STR 1095/2-1 to

P.S.]; and the Klaus Tschira Foundation [institutional support to F.G.]. Deposited in PMC for release after 12 months.

Supplementary material available online at
<http://jcs.biologists.org/lookup/suppl/doi:10.1242/jcs.089516/-DC1>

References

- Albers, K. M., Davis, F. E., Perrone, T. N., Lee, E. Y., Liu, Y. and Vore, M. (1995). Expression of an epidermal keratin protein in liver of transgenic mice causes structural and functional abnormalities. *J. Cell Biol.* **128**, 157–169.
- Alberts, A. J., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2008). The cytoskeleton. In *Molecular Biology of the Cell* (ed. M. Anderson and S. Granum), pp. 965–1052. New York: Garland Science.
- Berg, J. M., Tymoczko, J. L. and Stryer, L. (2007). Protein composition and function. In *Biochemistry* (ed. P. Zimmerman), pp. 25–64. New York: W. H. Freeman.
- Blessing, M., Nanney, L. B., King, L. E., Jones, C. M. and Hogan, B. L. (1993). Transgenic mice as a model to study the role of TGF-beta-related molecules in hair follicles. *Genes Dev.* **7**, 204–215.
- Chen, J., Cheng, X., Merched-Sauvage, M., Caulin, C., Roop, D. R. and Koch, P. J. (2006). An unexpected role for keratin 10 end domains in susceptibility to skin cancer. *J. Cell Sci.* **119**, 5067–5076.
- Chiappe, C. C., Yang, J. M., DiGiovanna, J. J., Steinert, P. M., Marekov, L., Compton, J. G. and Bale, S. J. (1994). Preferential sites in keratin 10 that are mutated in epidermolytic hyperkeratosis. *Am. J. Hum. Genet.* **54**, 179–190.
- Conway, J. F. and Parry, D. A. D. (1990). Structural features in the heptad substructures and longer range repeats of two-stranded alpha-fibrous proteins. *Int. J. Biol. Macromol.* **12**, 328–334.
- 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.

- Coulombe, P. A., Kerns, M. L. and Fuchs, E.** (2009). Epidermolysis bullosa simplex: a paradigm for disorders of tissue fragility. *J. Clin. Invest.* **119**, 1784-1793.
- Crewhouse, W. G., Dowling, L. M., Parry, D. A. D. and Steinert, P. M.** (1983). The structure of Intermediate Filaments. *Int. J. Biol. Macromol.* **5**, 267-274.
- Eriksson, J. E., Thomas, T., Grin, B., Helfand, B., Mendez, M., Pallari, H.-M. and Goldman, R. D.** (2009). Introducing intermediate filaments: from discovery to disease. *J. Clin. Invest.* **119**, 1763-1771.
- Fraser, R. D. B. and Parry, D. A. D.** (2007). Structural changes in the trichocyte intermediate filaments accompanying the transition from the reduced to the oxidized form. *J. Struct. Biol.* **159**, 36-45.
- Fraser, R. D. B., MacRae, T. P., Sparrow, L. G. and Parry, D. A. D.** (1988). Disulphide bonding in alpha keratin. *Int. J. Biol. Macromol.* **10**, 106-112.
- Geisler, N., Kaufmann, E. and Weber, K.** (1982). Proteinchemical characterization of three structurally distinct domains along the protofilament unit of desmin 10 nm filaments. *Cell* **30**, 277-286.
- Gu, L. H. and Coulombe, P. A.** (2007). Keratin function in skin epithelia: a broadening palette with surprising shades. *Curr. Opin. Cell. Biol.* **19**, 13-23.
- Hatzfeld, M. and Franke, W. W.** (1985). Pair formation and promiscuity of cytokeratins: formation in vitro of heterotypic complexes and intermediate-sized filaments by homologous and heterologous recombinations of purified polypeptides. *J. Cell Biol.* **101**, 1826-1841.
- Hatzfeld, M. and Weber, K.** (1990). The coiled-coil of *in vitro* assembled keratin filaments is a heterodimer of type I and II keratins: Use of site-specific mutagenesis and recombinant protein expression. *J. Cell Biol.* **110**, 1199-1210.
- Herrling, J. and Sparrow, L. G.** (1991). Interaction of intermediate filament proteins from wool. *Int. J. Biol. Macromol.* **13**, 115-119.
- Herrmann, H., Wedig, T., Porter, R. M., Lane, E. B. and Aeby, U.** (2002). Characterization of early assembly intermediates of recombinant human keratins. *J. Struct. Biol.* **137**, 82-96.
- Herrmann, H., Strelkov, S. V., Burkhardt, P. and Aeby, U.** (2009). Intermediate filaments: primary determinants of cell architecture and plasticity. *J. Clin. Invest.* **119**, 1772-1783.
- Hesse, M., Franz, T., Tamai, Y., Taketo, M. M., Magin, T. M.** (2000). Targeted deletion of keratins 18 and 19 leads to trophoblast fragility and early embryonic lethality. *EMBO J.* **19**, 5060-5070.
- Hesse, M., Watson, E. D., Schwaluk, T. and Magin, T. M.** (2005). Rescue of keratin 18/19 doubly deficient mice using aggregation with tetraploid embryos. *Eur. J. Cell Biol.* **84**, 355-361.
- Hesse, M., Grund, C., Herrmann, H., Bröhl, D., Franz, T., Omary, M. B. and Magin, T. M.** (2007). A mutation of keratin 18 within the coil 1A consensus motif causes widespread keratin aggregation but cell type-restricted lethality in mice. *Exp. Cell Res.* **313**, 3127-3140.
- Hutton, E., Paladini, R. D., Yu, Q. C., Yen, M.-Y., Coulombe, P. A. and Fuchs, E.** (1998). Functional differences between keratins of stratified and simple epithelia. *J. Cell Biol.* **143**, 1-13.
- Hyder, C. L., Pallari, H. M., Kochin, V. and Eriksson, J. E.** (2008). Providing cellular signposts-post-translational modifications of intermediate filaments. *FEBS Lett.* **582**, 2140-2148.
- Klinge, E. M., Sylvestre, Y. R., Freedberg, I. M. and Blumenberg, M.** (1987). Evolution of keratin genes: different protein domains evolve by different pathways. *J. Mol. Evol.* **24**, 319-329.
- Knapp, A. C. and Franke, W. W.** (1989). Spontaneous losses of control of cytokeratin gene expression in transformed, non-epithelial human cells occurring at different levels of regulation. *Cell* **59**, 67-79.
- Ku, N. O. and Omary, M. B.** (2006). A disease- and phosphorylation-related nonmechanical function for keratin 8. *J. Cell Biol.* **174**, 115-125.
- 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. Gastrointest. Liver Physiol.* **277**, G1108-G1137.
- Ku, N. O., Gish, R., Wright, T. L. and Omary, M. B.** (2001). Keratin 8 mutations in patients with cryptogenic liver disease. *N. Engl. J. Med.* **344**, 1580-1587.
- Ku, N. O., Lim, J. K., Krams, S. M., Esquivel, C. O., Keeffe, E. B., Wright, T. L., Parry, D. A. D. and Omary, M. B.** (2005). Keratins as susceptibility genes for end-stage liver disease. *Gastroenterology* **129**, 885-893.
- Ku, N. O., Strnad, P., Zhong, B. H., Tao, G. Z. and Omary, M. B.** (2007). Keratins let liver live: mutations predispose to liver disease and crosslinking generates Mallory-Denk bodies. *Hepatology* **46**, 1639-1649.
- Langbein, L., Yoshida, H., Praetzel-Wunder, S., Parry, D. A. and Schweizer, J.** (2010). The keratins of the human beard hair medulla: the riddle in the middle. *J. Invest. Dermatol.* **130**, 55-73.
- Lee, C. H. and Coulombe, P. A.** (2009). Self-organization of keratin intermediate filaments into cross-linked networks. *J. Cell Biol.* **186**, 409-421.
- Ma, L., Yamada, S., Wirtz, D. and Coulombe, P. A.** (2001). A 'hot-spot' mutation alters the mechanical properties of keratin filament networks. *Nat. Cell Biol.* **3**, 503-506.
- Mack, J. W., Torchia, D. A. and Steinert, P. M.** (1988). Solid-state NMR studies of the dynamics and structure of mouse keratin intermediate filaments. *Biochemistry* **27**, 5418-5426.
- Mohrenweiser, H.** (1994). International Commission for Protection Against Environmental Mutagens and Carcinogens. Working paper no. 5. Impact of the molecular spectrum of mutational lesions on estimates of germinal gene-mutation rates. *Mutat. Res.* **304**, 119-137.
- Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. and Krepler, R.** (1982). The catalog of human cytokeratins – patterns of expression in normal epithelia, tumors and cultured cells. *Cell* **31**, 11-24.
- Moll, R., Divo, M. and Langbein, L.** (2008). The human keratins: biology and pathology. *Histochem. Cell Biol.* **129**, 705-733.
- Navasaria, H. A., Swensson, O., Ratnavel, R. C., Shamsher, M., McLean, W. H., Lane, E. B., Griffiths, D., Eady, R. A. and Leigh, I. M.** (1995). Ultrastructural changes resulting from keratin-9 gene mutations in two families with epidermolytic palmoplantar keratoderma. *J. Invest. Dermatol.* **104**, 425-429.
- Omary, M. B., Ku, N. O., Liao, J. and Price, D.** (1998). Keratin modifications and solubility properties in epithelial cells and in vitro. *Subcell Biochem.* **31**, 105-140.
- Omary, M. B., Ku, N. O., Tao, G. Z., Toivola, D. M. and Liao, J.** (2006). Heads-and-tails of intermediate filament phosphorylation: multiple sites and functional insights. *Trends Biochem. Sci.* **31**, 383-394.
- Omary, M. B., Ku, N. O., Strnad, P. and Hanada, S.** (2009). Toward unraveling the complexity of simple epithelial keratins in health and disease. *J. Clin. Invest.* **119**, 1794-1805.
- Owens, D. W. and Lane, E. B.** (2004). Keratin mutations and intestinal pathology. *J. Pathol.* **204**, 377-385.
- Paladini, R. D. and Coulombe, P. A.** (1999). The functional diversity of epidermal keratins revealed by the partial rescue of the keratin 14 null phenotype by keratin 16. *J. Cell Biol.* **146**, 1185-1201.
- Pang, Y. Y., Schermer, A., Yu, J. and Sun, T. T.** (1993). Suprabasal change and subsequent formation of disulfide-stabilized homo- and hetero-dimers of keratins during esophageal epithelial differentiation. *J. Cell Sci.* **104**, 727-740.
- Parry, D. A. D.** (2005). Microdissection of the sequence and structure of intermediate filament chains. *Adv. Protein Chem.* **70**, 113-142.
- Parry, D. A. D. and Fraser, R. D. B.** (1985). Intermediate filament structure. I. Analysis of IF protein sequence data. *Int. J. Biol. Macromol.* **7**, 203-213.
- Parry, D. A. D. and Steinert, P. M.** (1995). *Intermediate Filament Structure*, pp. 1-183. Heidelberg: Springer-Verlag.
- Parry, D. A. D. and North, A. C. T.** (1998). Hard α -keratin intermediate filament chains: substructure of the N- and C-terminal domains and the predicted structure and function of the C-Terminal domains of Type I and Type II chains. *J. Struct. Biol.* **122**, 67-75.
- Parry, D. A. D., Steven, A. C. and Steinert, P. M.** (1985). The coiled-coil molecules of intermediate filaments consist of two parallel chains in exact axial register. *Biochem. Biophys. Res. Commun.* **127**, 1012-1018.
- Parry, D. A. D., Strelkov, S. V., Burkhardt, P., Aeby, U. and Herrmann, H.** (2007). Towards a molecular description of intermediate filament structure and assembly. *Exp. Cell Res.* **313**, 2204-2216.
- Parry, D. A. D., Fraser, R. D. B. and Squire, J. M.** (2008). Fifty years of coiled-coils and α -helical bundles: a close relationship between sequence and structure. *J. Struct. Biol.* **163**, 258-269.
- Pfeifer, G. P.** (2006). Mutagenesis at methylated CpG sequences. *Curr. Top. Microbiol. Immunol.* **301**, 259-281.
- Quinlan, R. A., Cohlberg, J. A., Schiller, D. L., Hatzfeld, M. and Franke, W. W.** (1984). Heterotypic tetramer (A2D2) complexes of non-epidermal keratins isolated from cytoskeletons of rat hepatocytes and hepatoma cells. *J. Mol. Biol.* **178**, 365-388.
- Quinlan, R. A., Hatzfeld, M., Franke, W. W., Lustig, A., Schultheiss, T. and Engel, J.** (1986). Characterization of dimer subunits of intermediate filament proteins. *J. Mol. Biol.* **192**, 337-349.
- Schweizer, J., Langbein, L., Rogers, M. A. and Winter, H.** (2007). Hair follicle-specific keratins and their diseases. *Exp. Cell Res.* **313**, 2010-2020.
- Smith, T. A. and Parry, D. A. D.** (2007). Sequence analyses of Type I and Type II chains in human hair and epithelial keratin intermediate filaments: promiscuous obligate heterodimers. Type II template for molecule formation and a rationale for heterodimer formation. *J. Struct. Biol.* **158**, 344-357.
- Snider, N. T., Weerasinghe, S. V. W., Iniguez-Lluhi, J. A., Herrmann, H. and Omary, M. B.** (2011). Keratin hypersumoylation alters filament dynamics and is a marker for human liver disease and keratin mutation. *J. Biol. Chem.* **286**, 2273-2284.
- Sommer, S. S.** (1992). Assessing the underlying pattern of human germline mutations: lessons from the factor IX gene. *FASEB J.* **6**, 2767-2774.
- Steinert, P. M.** (1990). The two-chain coiled-coil molecule of native epidermal keratin intermediate filaments is a Type I-Type II heterodimer. *J. Biol. Chem.* **265**, 8766-8774.
- Steinert, P. M., Parry, D. A. D., Idler, W. W., Johnson, L. D., Steven, A. C. and Roop, D. R.** (1985). Amino acid sequences of mouse and human epidermal type II keratins of Mr 67,000 provide a systematic basis for the structural and functional diversity of the end domains of keratin intermediate filament subunits. *J. Biol. Chem.* **260**, 7142-7149.
- Steinert, P. M., Mack, J. W., Korge, B. P., Gan, S.-Q., Haynes, S. R. and Steven, A. C.** (1991). Glycine loops in proteins: Their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. *Int. J. Biol. Macromol.* **13**, 130-139.
- Strnad, P., Zhou, Q., Hanada, S., Lazzeroni, L. C., Zhong, B. H., So, P., Davern, T. J., Lee, W. M., Acute Liver Failure Study Group and Omary, M. B.** (2010). Keratin variants predispose to acute liver failure and adverse outcome: race and ethnic associations. *Gastroenterology* **139**, 828-835.
- Sun, T.-T., Eichner, R., Cooper, D., Schermer, A., Nelson, W. G. and Weiss, R. A.** (1984). In *The Cancer Cell*, Vol. 1 (ed. A. Levine, W. Topp, G. Van de Woude and J. Watson), pp. 169-176. New York: Cold Spring Harbor Laboratory Press.

- Takahashi, K., Folmer, J. and Coulombe, P. A. (1994). Increased expression of keratin 16 causes anomalies in cytoarchitecture and keratinization in transgenic mouse skin. *J. Cell Biol.* **127**, 505-520.
- Tao, G. Z., Nakamichi, I., Ku, N.-O., Wang, J., Frolkis, M., Gong, X., Zhu, W., Pytela, R. and Omary, M. B. (2006). Bispecific and human disease-related anti-keratin rabbit monoclonal antibodies. *Exp. Cell Res.* **312**, 411-422.
- Tao, G. Z., Strnad, P., Zhou, Q., Kamal, A., Zhang, L., Madani, N. D., Kugathasan, S., Brant, S. R., Cho, J. H., Omary, M. B. and Duerr, R. H. (2007). Analysis of keratin polypeptides 8 and 19 variants in inflammatory bowel disease. *Clin. Gastroenterol. Hepatol.* **5**, 857-864.
- Wawersik, M., Paladini, R. D., Noensie, E. and Coulombe, P. A. (1997). A proline residue in the alpha-helical rod domain of type I keratin 16 destabilizes keratin heterotetramers. *J. Biol. Chem.* **272**, 32557-32565.
- Zhang, Y.-Q. and Sarge, K. D. (2008). Sumoylation regulates lamin A function and is lost in lamin A mutants associated with familial cardiomyopathies. *J. Cell Biol.* **182**, 35-39.

Table S1a. Absolute amino acid composition of selected human keratins.

AA, Amino acid; IF, Intermediate filament; K, Keratin; Amino acids are denoted by their single-letter-abbreviations.

IF Subtype	Protein name	Accession number	AA Length	Amino acid description (single letter code)																			
				A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Type II	K1	NP_006112.3	644	24	3	27	37	21	146	4	30	28	42	8	28	7	36	38	90	24	26	1	24
	K2	NP_000414.2	639	31	5	25	40	25	138	8	29	35	45	7	25	8	31	32	84	17	35	1	18
	K3	NP_476429.2	628	49	3	25	38	30	131	5	28	28	48	12	26	9	33	32	70	18	24	1	18
	K4	NP_002263.2	594	42	10	25	42	22	65	4	26	38	56	12	24	14	34	34	66	30	37	1	12
	K5	NP_000415.2	590	36	4	24	41	26	96	3	21	29	51	13	28	9	27	37	68	26	34	2	15
	K6A	NP_005545.1	564	42	5	24	40	17	81	5	27	30	52	8	24	8	30	36	60	26	30	2	17
	K7	NP_005547.3	469	53	1	24	45	12	35	3	22	25	51	10	19	8	27	36	46	18	23	2	9
	K8	NP_002264.1	483	33	0	21	49	12	36	2	25	32	49	16	23	6	27	32	60	22	22	1	15
	K81	NP_002272.2	505	39	34	17	51	12	46	4	25	24	45	6	21	12	22	35	45	20	33	1	13
	K85	NP_002274.1	507	44	27	16	52	13	36	4	23	25	42	7	20	14	20	41	54	19	33	2	15
	K86	NP_002275.1	486	40	29	17	51	11	35	4	22	24	41	6	21	12	21	37	42	21	37	1	14
Type I	K9	NP_000217.2	623	19	4	24	42	17	177	8	22	26	41	11	25	3	33	24	87	18	14	1	27
	K10	NP_000412.3	584	22	4	19	44	18	150	8	19	23	47	6	25	3	26	24	89	16	14	2	25
	K12	NP_000214.1	494	38	2	27	48	12	55	2	18	20	57	15	22	7	29	31	54	21	22	2	12
	K13a	NP_705694.2	458	34	6	19	46	12	57	3	19	18	46	13	19	7	22	30	42	29	20	2	14
	K14	NP_000517.2	472	30	5	24	45	12	53	4	18	23	48	15	20	4	25	32	52	22	24	2	14
	K16	NP_005548.2	473	31	6	20	45	13	57	3	19	17	50	11	14	6	32	33	61	20	19	2	14
	K17	NP_000413.1	432	31	5	20	48	7	37	3	19	20	49	10	19	4	29	32	38	25	20	2	14
	K18	NP_000215.1	430	38	0	24	43	8	27	6	23	21	45	10	18	4	24	35	37	30	24	2	11
	K19	NP_002267.2	400	36	0	23	37	8	36	6	17	14	44	9	13	5	25	31	39	21	20	2	14
	K20	NP_061883.1	424	23	1	19	48	6	22	9	18	27	52	10	22	4	33	29	30	27	30	1	13
	K23	NP_056330.3	422	22	3	19	45	8	27	16	23	30	40	13	18	13	27	26	34	26	17	4	11
	K24	NP_061889.2	525	39	9	25	45	18	81	2	21	23	41	11	20	2	23	30	73	24	22	2	14
	K31	NP_002268.2	416	22	25	15	45	8	7	3	14	12	52	2	30	22	30	30	40	24	23	2	10
	K35	NP_002271.3	455	37	23	20	47	7	24	2	13	17	54	8	19	20	26	32	49	18	23	3	13

Table S1b. Relative amino acid composition of selected human keratins.

AA, Amino acid; IF, Intermediate filament; K, Keratin. Amino acids are denoted by their single-letter-abbreviations.

Numbers are shown as percent composition for the listed amino acids.

IF Subtype	Protein name	Relative amino acid composition (%)																			
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Type II	K1	3.7	0.5	4.2	5.7	3.3	22.7	0.6	4.7	4.3	6.5	1.2	4.3	1.1	5.6	5.9	14.0	3.7	4.0	0.2	3.7
	K2	4.9	0.8	3.9	6.3	3.9	21.6	1.3	4.5	5.5	7.0	1.1	3.9	1.3	4.9	5.0	13.1	2.7	5.5	0.2	2.8
	K3	7.8	0.5	4.0	6.1	4.8	20.9	0.8	4.5	4.5	7.6	1.9	4.1	1.4	5.3	5.1	11.1	2.9	3.8	0.2	2.9
	K4	7.1	1.7	4.2	7.1	3.7	10.9	0.7	4.4	6.4	9.4	2.0	4.0	2.4	5.7	5.7	11.1	5.1	6.2	0.2	2.0
	K5	6.1	0.7	4.1	6.9	4.4	16.3	0.5	3.6	4.9	8.6	2.2	4.7	1.5	4.6	6.3	11.5	4.4	5.8	0.3	2.5
	K6A	7.4	0.9	4.3	7.1	3.0	14.4	0.9	4.8	5.3	9.2	1.4	4.3	1.4	5.3	6.4	10.6	4.6	5.3	0.4	3.0
	K7	11.3	0.2	5.1	9.6	2.6	7.5	0.6	4.7	5.3	10.9	2.1	4.1	1.7	5.8	7.7	9.8	3.8	4.9	0.4	1.9
	K8	6.8	0.0	4.3	10.1	2.5	7.5	0.4	5.2	6.6	10.1	3.3	4.8	1.2	5.6	6.6	12.4	4.6	4.6	0.2	3.1
	K81	7.7	6.7	3.4	10.1	2.4	9.1	0.8	5.0	4.8	8.9	1.2	4.2	2.4	4.4	6.9	8.9	4.0	6.5	0.2	2.6
	K85	8.7	5.3	3.2	10.3	2.6	7.1	0.8	4.5	4.9	8.3	1.4	3.9	2.8	3.9	8.1	10.7	3.7	6.5	0.4	3.0
	K86	8.2	6.2	3.5	10.5	2.3	7.2	0.8	4.5	4.9	8.4	1.2	4.3	2.5	4.3	7.6	8.6	4.3	7.6	0.2	2.9
Type I	K9	3.0	0.6	3.9	6.7	2.7	28.4	1.3	3.5	4.2	6.6	1.8	4.0	0.5	5.3	3.9	14.0	2.9	2.2	0.2	4.3
	K10	3.8	0.7	3.3	7.5	3.1	25.7	1.4	3.3	3.9	8.0	1.0	4.3	0.5	4.5	4.1	15.2	2.7	2.4	0.3	4.3
	K12	7.7	0.4	5.5	9.7	2.4	11.1	0.4	3.6	4.0	11.5	3.0	4.5	1.4	5.9	6.3	10.9	4.3	4.5	0.4	2.4
	K13a	7.4	1.3	4.1	10.0	2.6	12.4	0.7	4.1	3.9	10.0	2.8	4.1	1.5	4.8	6.6	9.2	6.3	4.4	0.4	3.1
	K14	6.4	1.1	5.1	9.5	2.5	11.2	0.8	3.8	4.9	10.2	3.2	4.2	0.8	5.3	6.8	11.0	4.7	5.1	0.4	3.0
	K16	6.6	1.3	4.2	9.5	2.7	12.1	0.6	4.0	3.6	10.6	2.3	3.0	1.3	6.8	7.0	12.9	4.2	4.0	0.4	3.0
	K17	7.2	1.2	4.6	11.1	1.6	8.6	0.7	4.4	4.6	11.3	2.3	4.4	0.9	6.7	7.4	8.8	5.8	4.6	0.5	3.2
	K18	8.8	0.0	5.6	10.0	1.9	6.3	1.4	5.3	4.9	10.5	2.3	4.2	0.9	5.6	8.1	8.6	7.0	5.6	0.5	2.6
	K19	9.0	0.0	5.8	9.3	2.0	9.0	1.5	4.3	3.5	11.0	2.3	3.3	1.3	6.3	7.8	9.8	5.3	5.0	0.5	3.5
	K20	5.4	0.2	4.5	11.3	1.4	5.2	2.1	4.2	6.4	12.3	2.4	5.2	0.9	7.8	6.8	7.1	6.4	7.1	0.2	3.1
	K23	5.2	0.7	4.5	10.7	1.9	6.4	3.8	5.5	7.1	9.5	3.1	4.3	3.1	6.4	6.2	8.1	6.2	4.0	0.9	2.6
	K24	7.4	1.7	4.8	8.6	3.4	15.4	0.4	4	4.4	7.8	2.1	3.8	0.4	4.4	5.7	13.9	4.6	4.2	0.4	2.7
	K31	5.3	6.0	3.6	10.8	1.9	1.7	0.7	3.4	2.9	12.5	0.5	7.2	5.3	7.2	7.2	9.6	5.8	5.5	0.5	2.4
	K35	8.1	5.1	4.4	10.3	1.5	5.3	0.4	2.9	3.7	11.9	1.8	4.2	4.4	5.7	7.0	10.8	4.0	5.1	0.7	2.9

Table S2. Keratin subdomain information for sequences used for multiple species comparison. Amino acid length of head, rod and tail domains are indicated as first, second and third number, respectively.

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	180-313-151		188-313-136	146-313-98	143-313-0		
K2	178-313-148	176-313-128	199-313-195				
K5	168-313-109	169-313-119	162-313-105	168-313-107	174-313-112	150-313-102	154-311-93
K9	153-312-158	131-312-85	131-312-300				
K10	146-314-124	127-314-85	135-314-112	103-314-84	143-311-139		
K14	115-311-46	121-311-45	121-311-52		109-311-47	187-311-0	
K7	91-312-66	92-312-62	85-312-60	102-287-0			
K8	91-311-81	98-311-69	97-311-82		93-295-72	99-311-97	115-311-94
K18	80-312-38	79-312-38	72-313-38		77-312-105	55-312-11	84-311-36
K19	80-311-9	79-311-9	83-311-9	86-311-10	94-312-17	104-308-0	49-311-1
K20	70-311-43	48-310-42	77-291-63		81-311-44		
K31	56-311-49	56-311-49	56-311-49				
K35	97-311-47	97-311-47	97-311-47				
K81	106-311-88	106-311-83	106-311-64	105-311-91			
K85	123-311-73	123-311-73	123-311-73				
K86	106-311-69	106-311-69	106-311-69	107-311-76			

Table S3a. Relative amino acid composition of human type I and type II keratins. The percentage share of the respective amino acid subgroups was calculated for the keratin subdomains of human type I and type II keratins. Grouping of amino acids is based on Stryer's biochemistry, 5th edition. AS, Amino acids with aliphatic side chains (isoleucine, leucine, methionine, valine); AO, Amino acids with hydroxyl groups (serine, threonine); AR, Aromatic amino acids (phenylalanine, tryptophan; tyrosine); BA, Basic amino acids (arginine, lysine); AC, Amino acids with acidic side chains (aspartate, glutamate); OT, Amino acids not belonging to any of the predefined subgroups (alanine, asparagine, cysteine, glutamine, glycine, histidine, proline). Human keratin sequences K1-K8/ K81/K85/K86 and K9-K10/K12-K14/ K16-K20/K23/K24/ K31/K35 were used for type II and type I keratins, respectively. Note that the amino acid distribution substantially differs between keratin subdomains. For example, type II keratins display higher content of basic amino acids in the head and rod domains, but less acidic amino acids in the tail.

	Sub-type	Relative amino acid composition (%)					
		AS	AO	AR	BA	AC	OT
Head	I	12.6	24.5	9.5	5.8	2.2	45.6
	II	15.9	18.2	7.8	8.9	3.5	45.6
Rod	I	25.0	11.1	4.9	12.7	19.3	27.1
	II	24.6	9.3	5.7	14.6	19.1	26.7
Tail	I	16.7	26.5	3.7	11.7	7.9	33.5
	II	15.4	30.2	5.1	6.3	2.0	41.0

Table S3b. Relative amino acid composition of keratin head, rod and tail domains differs between human type I and type II keratins. The percentage share of the respective amino acid subgroups was calculated for the keratin subdomains of human type I and type II keratins and amino acids are displayed in single-letter-code. Human keratin sequences K1-K8/ K81/K85/K86 and K9-K10/K12-K14/ K16-K20/K23/K24/ K31/K35 were used for type II and type I keratins, respectively. Note that the amino acid distribution substantially differs between keratin subdomains. For example, type II keratins display higher isoleucine content in the head than type I keratins, while both keratin subtypes display comparable isoleucine percentage in the rod and tail.

	Sub-type	Relative amino acid composition (%)																			
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Head	I	5.5	3.1	1.0	1.2	6.2	30.3	0.8	1.7	0.6	5.7	1.9	2.0	2.9	1.1	5.2	20.8	3.7	3.3	0.4	2.9
	II	5.9	3.8	1.0	2.5	5.5	24.8	0.5	4.0	1.5	5.2	1.1	2.4	5.2	3.0	7.4	14.4	3.8	5.5	0.0	2.3
Rod	I	7.4	0.8	6.1	13.2	1.1	3.2	1.3	5.0	5.6	13.0	2.6	5.6	0.9	7.8	7.0	5.8	5.3	4.5	0.5	3.3
	II	8.7	0.9	6.4	12.7	2.3	2.6	1.0	5.4	7.8	11.9	2.4	5.8	0.3	7.2	6.8	5.4	4.0	4.9	0.4	3.0
Tail	I	3.9	2.8	2.9	5.1	1.4	12.2	1.0	3.4	5.8	4.4	0.9	4.1	4.8	4.8	5.9	19.0	7.5	8.0	0.8	1.5
	II	4.4	3.4	0.9	1.1	2.4	27.4	0.4	2.7	2.2	3.6	0.6	2.3	1.9	1.1	4.1	25.2	5.0	8.5	0.0	2.7

Table S4a. Relative amino acid composition of selected human keratins. The percentage share of the respective amino acid subgroups was calculated for the indicated human type I and type II keratins.

Grouping of amino acids is based on Stryer's biochemistry, 5th edition. AS, Amino acids with aliphatic side chains (isoleucine, leucine, methionine, valine); AO, Amino acids with hydroxyl groups (serine, threonine); AR, Aromatic amino acids (phenylalanine, tryptophan; tyrosine); BA, Basic amino acids (arginine, lysine); AC, Amino acids with acidic side chains (aspartate, glutamate); OT, Amino acids not belonging to any of the predefined subgroups (alanine, asparagine, cysteine, glutamine, glycine, histidine, proline). IF, Intermediate filament; K, Keratin.

IF Subtype	Protein name	Relative amino acid composition (%)					
		AS	AO	AR	BA	AC	OT
Type II	K1	16.5	17.7	7.1	10.2	9.9	38.5
	K2	18.2	15.8	6.9	10.5	10.2	38.5
	K3	17.8	14.0	7.8	9.6	10.0	40.8
	K4	22.1	16.2	5.9	12.1	11.3	32.5
	K5	20.2	15.9	7.3	11.2	11.0	34.4
	K6A	20.7	15.2	6.4	11.7	11.3	34.6
	K7	22.6	13.6	4.9	13.0	14.7	31.1
	K8	23.2	17.0	5.8	13.3	14.5	26.3
	K81	21.6	12.9	5.1	11.7	13.5	35.2
	K85	20.7	14.4	5.9	13.0	13.4	32.5
	K86	21.8	13.0	5.3	12.6	14.0	33.6
Type I	K9	14.1	16.9	7.2	8.0	10.6	43.2
	K10	14.7	18.0	7.7	8.0	10.8	40.8
	K12	22.7	15.2	5.3	10.3	15.2	31.4
	K13a	21.4	15.5	6.1	10.5	14.2	32.3
	K14	22.2	15.7	5.9	11.7	14.6	29.9
	K16	20.9	17.1	6.1	10.6	13.7	31.5
	K17	22.7	14.6	5.3	12.0	15.7	29.6
	K18	23.7	15.6	4.9	13.0	15.6	27.2
	K19	22.5	15.0	6.0	11.3	15.0	30.3
	K20	25.9	13.4	4.7	13.2	15.8	26.9
	K23	22.0	14.2	5.5	13.3	15.2	29.9
	K24	18.1	18.5	6.5	10.1	13.3	33.5
	K31	21.9	15.4	4.8	10.1	14.4	33.4
	K35	21.5	14.7	5.1	10.8	14.7	33.2

Table S4b. Relative amino acid composition of human type I and type II keratins domains.

The percentage share of the respective amino acid subgroups was calculated for the indicated type I and type II keratin domains. Grouping of amino acids is based on Stryer's biochemistry, 5th edition. AS, Amino acids with aliphatic side chains (isoleucine, leucine, methionine, valine); AO, Amino acids with hydroxyl groups (serine, threonine); AR, Aromatic amino acids (phenylalanine, tryptophan; tyrosine); BA, Basic amino acids (arginine, lysine); AC, Amino acids with acidic side chains (aspartate, glutamate); OT, Amino acids not belonging to any of the predefined subgroups (alanine, asparagine, cysteine, glutamine, glycine, histidine, proline). IF, Intermediate filament; K, Keratin.

IF Subtype	Protein name	Relative amino acid composition (%)																	
		Head						Rod						Tail					
		AS	AO	AR	BA	AC	OT	AS	AO	AR	BA	AC	OT	AS	AO	AR	BA	AC	OT
Type II	K1	11.7	16.1	10.0	8.3	2.8	51.1	24.3	11.8	5.8	13.7	18.5	25.9	6.0	31.8	6.6	5.3	0.7	49.7
	K2	12.9	14.6	9.0	7.3	2.2	53.9	26.5	8.0	5.8	14.1	18.8	26.8	6.8	33.8	6.8	6.8	1.4	44.6
	K3	11.1	13.6	10.1	5.6	2.0	57.6	25.4	9.5	5.7	13.7	18.4	27.3	8.7	27.0	9.6	5.2	0.9	48.7
	K4	18.5	16.6	6.6	11.4	4.7	42.2	25.9	10.9	5.4	14.1	17.9	25.9	15.7	38.6	5.7	5.7	1.4	32.9
	K5	14.3	20.2	9.5	7.7	2.4	45.8	24.6	8.3	6.1	15.3	19.2	26.5	16.5	31.2	7.3	4.6	0.9	39.4
	K6A	16.0	19.6	7.4	8.0	2.5	46.6	24.3	8.3	5.8	15.3	18.8	27.5	17.0	31.8	6.8	5.7	1.1	37.5
	K7	22.0	18.7	5.5	12.1	5.5	36.3	23.4	9.3	5.1	14.4	19.9	27.9	19.7	27.3	3.0	7.6	3.0	39.4
	K8	20.9	27.5	6.6	9.9	3.3	31.9	25.1	9.6	5.5	15.1	19.9	24.8	18.5	33.3	6.2	9.9	6.2	25.9
	K81	15.1	17.0	7.5	8.5	4.7	47.2	24.1	9.0	5.8	14.8	19.6	26.7	20.5	21.6	0.0	4.5	2.3	51.1
	K85	17.1	19.5	6.5	10.6	4.1	42.3	23.2	8.7	6.1	15.4	19.6	27.0	16.4	30.1	4.1	6.8	2.7	39.7
	K86	15.1	17.0	7.5	8.5	4.7	47.2	23.8	9.0	5.8	15.1	19.9	27.0	23.9	25.4	0.0	7.5	1.5	41.7
Type I	K9	5.2	22.2	13.1	5.2	1.3	52.9	25.0	11.2	4.8	11.2	19.2	28.5	1.3	22.8	6.3	4.4	2.5	62.7
	K10	6.2	23.3	13.7	4.8	1.4	50.7	24.2	11.5	5.4	11.5	19.1	28.3	0.8	28.2	6.5	3.2	0.8	60.5
	K12	17.6	23.2	7.2	4.8	2.4	44.8	24.1	10.5	5.1	12.4	19.7	28.3	25.9	24.1	1.9	11.1	18.5	18.5
	K13a	11.5	14.4	11.5	2.9	2.9	56.7	26.0	11.5	4.8	12.8	19.6	25.3	11.9	47.6	2.4	11.9	2.4	23.8
	K14	13.0	26.1	9.6	4.3	1.7	45.2	25.1	9.0	5.1	14.5	20.3	26.0	26.1	34.8	2.2	10.9	8.7	17.4
	K16	12.8	23.9	9.4	4.3	1.7	47.9	26.0	10.0	4.8	13.2	19.6	26.4	6.7	48.9	6.7	8.9	4.4	24.4
	K17	14.3	32.1	7.1	6.0	2.4	38.1	25.1	9.0	5.1	12.9	19.6	28.3	21.6	21.6	2.7	18.9	13.5	21.6
	K18	16.3	27.5	7.5	7.5	1.3	40.0	25.6	11.2	4.8	14.1	19.6	24.7	23.7	26.3	0.0	15.8	13.2	21.1
	K19	15.0	26.3	10.0	6.3	2.5	40.0	24.4	11.9	4.8	12.5	18.6	27.7	22.2	22.2	11.1	11.1	0.0	33.3
	K20	21.4	22.9	7.1	10.0	5.7	32.9	26.4	10.9	4.5	13.2	17.0	28.0	30.2	16.3	2.3	18.6	23.3	9.3
	K23	8.3	23.6	8.3	8.3	0.0	51.4	24.8	11.6	5.5	13.9	19.4	24.8	25.0	17.5	0.0	17.5	10.0	30.0
	K24	6.4	28.6	10.7	5.0	1.4	47.9	22.5	12.0	5.4	11.4	19.0	29.7	21.7	27.5	2.9	14.5	11.6	21.7
	K31	12.5	21.4	8.9	3.6	3.6	50.0	25.7	13.8	4.5	11.6	18.6	25.7	8.2	18.4	2.0	8.2	0.0	63.3
	K35	15.5	26.8	8.2	8.2	2.1	39.2	25.4	10.9	4.2	11.9	20.6	27.0	8.5	14.9	4.3	8.5	2.1	61.7

Table S5a. Absolute amino acid composition of selected human keratin head domains. AA, Amino acid; IF, Intermediate filament; K, Keratins. Amino acids are denoted by their single-letter-abbreviations.

IF Subtype	Protein name	AA Length	Amino acid description (single letter code)																			
			A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Type II	K1	180	3	2	1	4	13	72	0	9	3	5	1	4	5	6	12	25	4	6	0	5
	K2	178	5	3	1	3	15	72	2	5	4	7	1	3	6	5	9	24	2	10	0	1
	K3	198	15	2	1	3	16	77	1	6	3	6	3	4	7	8	8	25	2	7	0	4
	K4	211	16	8	5	5	11	41	1	8	8	13	5	4	13	6	16	26	9	13	0	3
	K5	168	9	2	1	3	13	50	0	6	1	8	1	4	7	5	12	26	8	9	0	3
	K6A	163	12	3	1	3	9	46	1	8	1	10	1	3	7	4	12	26	6	7	0	3
	K7	91	10	0	2	3	3	11	1	3	0	9	1	1	6	4	11	15	2	7	0	2
	K8	91	4	0	1	2	3	14	0	4	2	5	2	3	4	4	7	19	6	8	0	3
	K81	106	6	11	1	4	5	19	1	5	1	5	1	3	8	2	8	12	6	5	0	3
	K85	123	7	12	1	4	4	17	0	5	2	7	1	6	8	2	11	18	6	8	0	4
	K86	106	6	12	1	4	4	18	1	5	1	5	1	3	8	2	8	12	6	5	0	4
Type I	K9	153	4	2	1	1	13	73	0	2	0	4	1	1	0	1	8	33	1	1	0	7
	K10	146	0	2	1	1	15	70	1	1	2	5	1	1	0	0	5	34	0	2	0	5
	K12	125	8	1	2	1	7	40	0	2	0	11	6	3	3	1	6	27	2	3	0	2
	K13a	104	5	4	2	1	8	47	0	0	0	7	1	1	0	2	3	12	3	4	0	4
	K14	115	5	3	1	1	7	42	0	3	1	7	2	0	1	1	4	26	4	3	0	4
	K16	117	6	4	1	1	8	44	0	3	1	7	2	0	1	1	4	24	4	3	0	3
	K17	84	3	3	1	1	3	25	0	2	1	8	1	0	0	1	4	22	5	1	0	3
	K18	80	9	0	0	1	3	17	0	3	0	3	3	2	2	2	6	16	6	4	0	3
	K19	80	7	0	1	1	4	20	1	1	0	4	1	1	2	1	5	18	3	6	0	4
	K20	70	2	0	3	1	3	12	2	2	0	5	2	3	2	2	7	11	5	6	0	2
	K23	72	5	1	0	0	4	17	3	1	0	3	1	2	8	1	6	13	4	1	2	0
	K24	140	9	5	1	1	13	52	0	0	0	3	2	0	0	1	7	38	2	4	0	2
	K31	56	2	9	0	2	3	3	1	1	0	3	1	6	7	0	2	10	2	2	1	1
	K35	97	10	5	0	2	4	17	0	1	4	9	2	1	5	0	4	22	4	3	1	3

Table S5b. Relative amino acid composition of selected human keratin head domains

IF Subtype	Protein name	Relative amino acid composition (%)																			
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Type II	K1	1.7	1.1	0.6	2.2	7.2	40.0	0.0	5.0	1.7	2.8	0.6	2.2	2.8	3.3	6.7	13.9	2.2	3.3	0.0	2.8
	K2	2.8	1.7	0.6	1.7	8.4	40.4	1.1	2.8	2.2	3.9	0.6	1.7	3.4	2.8	5.1	13.5	1.1	5.6	0.0	0.6
	K3	7.6	1.0	0.5	1.5	8.1	38.9	0.5	3.0	1.5	3.0	1.5	2.0	3.5	4.0	4.0	12.6	1.0	3.5	0.0	2.0
	K4	7.6	3.8	2.4	2.4	5.2	19.4	0.5	3.8	3.8	6.2	2.4	1.9	6.2	2.8	7.6	12.3	4.3	6.2	0.0	1.4
	K5	5.4	1.2	0.6	1.8	7.7	29.8	0.0	3.6	0.6	4.8	0.6	2.4	4.2	3.0	7.1	15.5	4.8	5.4	0.0	1.8
	K6A	7.4	1.8	0.6	1.8	5.5	28.2	0.6	4.9	0.6	6.1	0.6	1.8	4.3	2.5	7.4	16.0	3.7	4.3	0.0	1.8
	K7	11.0	0.0	2.2	3.3	3.3	12.1	1.1	3.3	0.0	9.9	1.1	1.1	6.6	4.4	12.1	16.5	2.2	7.7	0.0	2.2
	K8	4.4	0.0	1.1	2.2	3.3	15.4	0.0	4.4	2.2	5.5	2.2	3.3	4.4	4.4	7.7	20.9	6.6	8.8	0.0	3.3
	K81	5.7	10.4	0.9	3.8	4.7	17.9	0.9	4.7	0.9	4.7	0.9	2.8	7.5	1.9	7.5	11.3	5.7	4.7	0.0	2.8
	K85	5.7	9.8	0.8	3.3	3.3	13.8	0.0	4.1	1.6	5.7	0.8	4.9	6.5	1.6	8.9	14.6	4.9	6.5	0.0	3.3
	K86	5.7	11.3	0.9	3.8	3.8	17.0	0.9	4.7	0.9	4.7	0.9	2.8	7.5	1.9	7.5	11.3	5.7	4.7	0.0	3.8
Type I	K9	2.6	1.3	0.7	0.7	8.5	47.7	0.0	1.3	0.0	2.6	0.7	0.7	0.0	0.7	5.2	21.6	0.7	0.7	0.0	4.6
	K10	0.0	1.4	0.7	0.7	10.3	47.9	0.7	0.7	1.4	3.4	0.7	0.7	0.0	0.0	3.4	23.3	0.0	1.4	0.0	3.4
	K12	6.4	0.8	1.6	0.8	5.6	32.0	0.0	1.6	0.0	8.8	4.8	2.4	2.4	0.8	4.8	21.6	1.6	2.4	0.0	1.6
	K13a	4.8	3.8	1.9	1.0	7.7	45.2	0.0	0.0	0.0	6.7	1.0	1.0	0.0	1.9	2.9	11.5	2.9	3.8	0.0	3.8
	K14	4.3	2.6	0.9	0.9	6.1	36.5	0.0	2.6	0.9	6.1	1.7	0.0	0.9	0.9	3.5	22.6	3.5	2.6	0.0	3.5
	K16	5.1	3.4	0.9	0.9	6.8	37.6	0.0	2.6	0.9	6.0	1.7	0.0	0.9	0.9	3.4	20.5	3.4	2.6	0.0	2.6
	K17	3.6	3.6	1.2	1.2	3.6	29.8	0.0	2.4	1.2	9.5	1.2	0.0	0.0	1.2	4.8	26.2	6.0	1.2	0.0	3.6
	K18	11.3	0.0	0.0	1.3	3.8	21.3	0.0	3.8	0.0	3.8	3.8	2.5	2.5	2.5	7.5	20.0	7.5	5.0	0.0	3.8
	K19	8.8	0.0	1.3	1.3	5.0	25.0	1.3	1.3	0.0	5.0	1.3	1.3	2.5	1.3	6.3	22.5	3.8	7.5	0.0	5.0
	K20	2.9	0.0	4.3	1.4	4.3	17.1	2.9	2.9	0.0	7.1	2.9	4.3	2.9	2.9	10.0	15.7	7.1	8.6	0.0	2.9
	K23	6.9	1.4	0	0	5.6	23.6	4.2	1.4	0	4.2	1.4	2.8	11.1	1.4	8.3	18.1	5.6	1.4	2.8	0
	K24	6.4	3.6	0.7	0.7	9.3	37.1	0	0	0	2.2	1.4	0	0	0.7	5	27.1	1.4	2.9	0	1.4
	K31	3.6	16.1	0.0	3.6	5.4	5.4	1.8	1.8	0.0	5.4	1.8	10.7	12.5	0.0	3.6	17.9	3.6	3.6	1.8	1.8

	K35	10.3	5.2	0.0	2.1	4.1	17.5	0.0	1.0	4.1	9.3	2.1	1.0	5.2	0.0	4.1	22.7	4.1	3.1	1.0	3.1
--	-----	------	-----	-----	-----	-----	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	-----	-----	-----	-----

AA, Amino acid; IF, Intermediate filament; K, Keratins. Amino acids are denoted by their single-letter-abbreviations.

Table S6. Absolute amino acid composition of selected human keratin rod domains.

IF Subtype	Protein name	AA Length	Amino acid description (single letter code)																			
			A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Type II	K1	313	20	0	26	32	7	5	2	19	23	37	7	23	1	30	20	23	14	13	1	10
	K2	313	22	2	23	36	6	8	5	21	27	37	6	20	2	25	17	14	11	19	1	11
	K3	315	28	0	24	34	7	10	4	18	24	41	8	21	1	22	19	16	14	13	1	10
	K4	313	25	1	20	36	7	5	3	14	28	40	7	19	1	27	16	18	16	20	1	9
	K5	313	25	2	23	37	9	7	3	14	25	36	12	23	1	22	23	11	15	15	2	8
	K6A	313	28	2	23	36	7	9	3	16	26	38	7	19	1	24	22	11	15	15	2	9
	K7	312	35	1	20	42	8	10	2	16	24	37	7	15	1	23	21	17	12	13	2	6
	K8	311	25	0	18	44	7	10	1	19	25	38	12	19	0	22	22	20	10	9	1	9
	K81	311	29	3	15	46	7	8	3	18	22	36	5	14	1	19	24	18	10	16	1	10
	K85	311	33	9	14	47	7	8	4	15	23	34	6	13	1	16	25	18	9	17	2	10
	K86	313	29	10	15	47	7	8	3	17	22	35	5	14	1	19	25	18	10	17	1	10
Type I	K9	312	13	2	23	37	3	12	4	19	24	36	10	23	3	32	11	18	17	13	1	11
	K10	314	22	2	18	42	2	10	3	18	19	42	5	24	2	26	17	20	16	11	2	13
	K12	315	29	1	22	40	4	13	2	14	15	43	7	18	3	23	24	19	14	12	2	10
	K13a	312	26	2	16	45	3	7	3	18	18	39	12	17	4	20	22	18	18	12	2	10
	K14	311	25	2	20	43	4	9	4	13	21	38	12	19	2	20	24	14	14	15	2	10
	K16	311	24	2	19	42	3	11	3	15	15	42	9	14	4	24	26	17	14	15	2	10
	K17	311	28	2	18	43	4	11	2	15	16	41	9	19	3	23	24	14	14	13	2	10
	K18	312	28	0	20	41	5	8	5	18	18	40	6	14	2	20	26	17	18	16	2	8
	K19	311	28	0	22	36	4	16	5	16	13	38	8	10	3	24	26	19	18	14	1	10
	K20	311	21	1	14	39	3	9	7	13	21	45	8	18	2	29	20	16	18	16	1	10
	K23	310	14	1	19	41	4	9	12	18	25	35	11	14	4	23	18	17	19	13	2	11
	K24	316	29	3	20	40	3	20	2	19	18	36	8	18	2	20	18	19	19	8	2	12
	K31	311	16	7	15	43	4	2	2	12	11	48	1	20	3	30	25	25	18	19	1	9
	K35	311	22	10	19	45	2	4	2	11	12	43	6	16	4	26	25	21	13	19	2	9

AA, Amino acid; K, Keratins. Amino acids are denoted by their single-letter-abbreviations.

Table S7a. Absolute amino acid composition of selected human keratin tail domains.

AA, Amino acid; IF, Intermediate filament; K, Keratins. Amino acids are denoted by their single-letter-abbreviations.

IF Subtype	Protein name	AA Length	Amino acid description (single letter code)																			
			A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Type II	K1	151	1	1	0	1	1	69	2	2	2	0	0	1	1	0	6	42	6	7	0	9
	K2	148	4	0	1	1	4	58	1	3	4	1	0	2	0	1	6	46	4	6	0	6
	K3	115	6	1	0	1	7	44	0	4	1	1	1	1	1	3	5	29	2	4	0	4
	K4	70	1	1	0	1	4	19	0	4	2	3	0	1	0	1	2	22	5	4	0	0
	K5	109	2	0	0	1	4	39	0	1	3	7	0	1	1	0	2	31	3	10	0	4
	K6A	88	2	0	0	1	1	26	1	3	3	4	0	2	0	2	2	23	5	8	0	5
	K7	66	8	0	2	0	1	14	0	3	1	5	2	3	1	0	4	14	4	3	0	1
	K8	81	4	0	2	3	2	12	1	2	5	6	2	1	2	1	3	21	6	5	0	3
	K81	88	4	14	1	1	0	19	0	2	1	4	0	4	3	1	3	15	4	12	0	0
	K85	73	4	6	1	1	2	11	0	3	0	1	0	1	5	2	5	18	4	8	0	1
	K86	67	5	7	1	0	0	9	0	0	1	1	0	4	3	0	4	12	5	15	0	0
Type I	K9	158	2	0	0	4	1	92	4	1	2	1	0	1	0	0	5	36	0	0	0	9
	K10	124	0	0	0	1	1	70	4	0	2	0	0	0	1	0	2	35	0	1	0	7
	K12	54	1	0	3	7	1	2	0	2	5	3	2	1	1	5	1	8	5	7	0	0
	K13a	42	3	0	1	0	1	3	0	1	0	0	0	1	3	0	5	12	8	4	0	0
	K14	46	0	0	3	1	1	2	0	2	1	3	1	1	1	4	4	12	4	6	0	0
	K16	45	1	0	0	2	2	2	0	1	1	1	0	0	1	7	3	20	2	1	0	1
	K17	37	0	0	1	4	0	1	1	2	3	0	0	0	1	5	4	2	6	6	0	1
	K18	38	1	0	4	1	0	2	1	2	3	2	1	2	0	2	3	4	6	4	0	0
	K19	9	1	0	0	0	0	0	0	0	1	2	0	2	0	0	0	2	0	0	1	0
	K20	43	0	0	2	8	0	1	0	3	6	2	0	1	0	2	2	3	4	8	0	1
	K23	40	3	1	0	4	0	1	1	4	5	2	1	2	1	3	2	4	3	3	0	0
	K24	69	1	1	4	4	2	9	0	2	5	2	1	2	0	2	5	16	3	10	0	0
	K31	49	4	9	0	0	1	2	0	1	1	1	0	4	12	0	3	5	4	2	0	0
	K35	47	5	8	1	0	1	3	0	1	1	2	0	2	11	0	3	6	1	1	0	1

Table S7b. Relative amino acid composition of selected keratin tail domains.

AA, Amino acid; IF, Intermediate filament; K, Keratins. Amino acids are denoted by their single-letter-abbreviations.

IF Subtype	Protein name	Relative amino acid composition (%)																			
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Type II	K1	0.7	0.7	0.0	0.7	0.7	45.7	1.3	1.3	1.3	0.0	0.0	0.7	0.7	0.0	4.0	27.8	4.0	4.6	0.0	6.0
	K2	2.7	0.0	0.7	0.7	2.7	39.2	0.7	2.0	2.7	0.7	0.0	1.4	0.0	0.7	4.1	31.1	2.7	4.1	0.0	4.1
	K3	5.2	0.9	0.0	0.9	6.1	38.3	0.0	3.5	0.9	0.9	0.9	0.9	2.6	4.3	25.2	1.7	3.5	0.0	3.5	
	K4	1.4	1.4	0.0	1.4	5.7	27.1	0.0	5.7	2.9	4.3	0.0	1.4	0.0	1.4	2.9	31.4	7.1	5.7	0.0	0.0
	K5	1.8	0.0	0.0	0.9	3.7	35.8	0.0	0.9	2.8	6.4	0.0	0.9	0.9	0.0	1.8	28.4	2.8	9.2	0.0	3.7
	K6A	2.3	0.0	0.0	1.1	1.1	29.5	1.1	3.4	3.4	4.5	0.0	2.3	0.0	2.3	2.3	26.1	5.7	9.1	0.0	5.7
	K7	12.1	0.0	3.0	0.0	1.5	21.2	0.0	4.5	1.5	7.6	3.0	4.5	1.5	0.0	6.1	21.2	6.1	4.5	0.0	1.5
	K8	4.9	0.0	2.5	3.7	2.5	14.8	1.2	2.5	6.2	7.4	2.5	1.2	2.5	1.2	3.7	25.9	7.4	6.2	0.0	3.7
	K81	4.5	15.9	1.1	1.1	0.0	21.6	0.0	2.3	1.1	4.5	0.0	4.5	3.4	1.1	3.4	17.0	4.5	13.6	0.0	0.0
	K85	5.5	8.2	1.4	1.4	2.7	15.1	0.0	4.1	0.0	1.4	0.0	1.4	6.8	2.7	6.8	24.7	5.5	11.0	0.0	1.4
	K86	7.5	10.4	1.5	0.0	0.0	13.4	0.0	0.0	1.5	1.5	0.0	6.0	4.5	0.0	6.0	17.9	7.5	22.4	0.0	0.0
	K9	1.3	0.0	0.0	2.5	0.6	58.2	2.5	0.6	1.3	0.6	0.0	0.6	0.0	0.0	3.2	22.8	0.0	0.0	0.0	5.7
	K10	0.0	0.0	0.0	0.8	0.8	56.5	3.2	0.0	1.6	0.0	0.0	0.0	0.8	0.0	1.6	28.2	0.0	0.8	0.0	5.6
	K12	1.9	0.0	5.6	13.0	1.9	3.7	0.0	3.7	9.3	5.6	3.7	1.9	1.9	9.3	1.9	14.8	9.3	13.0	0.0	0.0
Type I	K13a	7.1	0.0	2.4	0.0	2.4	7.1	0.0	2.4	0.0	0.0	0.0	2.4	7.1	0.0	11.9	28.6	19.0	9.5	0.0	0.0
	K14	0.0	0.0	6.5	2.2	2.2	4.3	0.0	4.3	2.2	6.5	2.2	2.2	2.2	8.7	8.7	26.1	8.7	13.0	0.0	0.0
	K16	2.2	0.0	0.0	4.4	4.4	4.4	0.0	2.2	2.2	2.2	0.0	0.0	2.2	15.6	6.7	44.4	4.4	2.2	0.0	2.2
	K17	0.0	0.0	2.7	10.8	0.0	2.7	2.7	5.4	8.1	0.0	0.0	0.0	2.7	13.5	10.8	5.4	16.2	16.2	0.0	2.7
	K18	2.6	0.0	10.5	2.6	0.0	5.3	2.6	5.3	7.9	5.3	2.6	5.3	0.0	5.3	7.9	10.5	15.8	10.5	0.0	0.0
	K19	11.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.1	22.2	0.0	22.2	0.0	0.0	0.0	22.2	0.0	0.0	11.1	0.0
	K20	0.0	0.0	4.7	18.6	0.0	2.3	0.0	7.0	14.0	4.7	0.0	2.3	0.0	4.7	4.7	7.0	9.3	18.6	0.0	2.3
	K23	7.5	2.5	0	10	0	2.5	2.5	10	12.5	5	2.5	5	2.5	7.5	5	10	7.5	7.5	0	0
	K24	1.4	1.4	5.8	5.8	2.9	13	0	2.9	7.2	2.9	1.4	2.9	0	2.9	7.2	23.2	4.4	14.5	0	0
	K31	8.2	18.4	0.0	0.0	2.0	4.1	0.0	2.0	2.0	2.0	0.0	8.2	24.5	0.0	6.1	10.2	8.2	4.1	0.0	0.0
	K35	10.6	17.0	2.1	0.0	2.1	6.4	0.0	2.1	2.1	4.3	0.0	4.3	23.4	0.0	6.4	12.8	2.1	2.1	0.0	2.1

Table S8. Relative amino acid composition of human epidermal, simple-type epithelial and hair keratins. Epid, epidermal. Amino acids are denoted by their single-letter-abbreviations.

		Relative amino acid composition (%)																			
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Head	Epid.	2.8	1.5	0.7	1.3	8.0	40.4	0.3	2.7	1.1	3.9	0.8	1.3	1.9	1.8	5.2	18.4	2.0	3.2	0.0	2.8
	Simple	7.6	0.0	1.8	1.9	3.9	18.2	1.0	3.1	0.4	6.3	2.2	2.5	3.8	3.1	8.7	19.1	5.4	7.5	0.0	3.4
	Hair	6.2	10.5	0.5	3.3	4.2	14.3	0.7	3.3	1.5	6.0	1.3	4.5	7.9	1.1	6.3	15.6	4.8	4.5	0.6	2.9
Rod	Epid.	6.8	0.5	7.1	12.1	1.7	2.7	1.1	5.5	7.4	12.0	2.8	7.0	0.6	8.3	6.0	5.3	4.6	4.6	0.5	3.4
	Simple	8.8	0.1	6.0	13.0	1.7	3.4	1.3	5.3	6.5	12.7	2.6	4.9	0.5	7.6	7.4	5.7	4.9	4.4	0.4	2.8
	Hair	8.3	2.9	5.0	14.7	1.7	1.9	0.9	4.7	5.8	12.6	1.5	5.0	0.6	7.1	8.0	6.4	3.9	5.7	0.5	3.1
Tail	Epid.	1.1	0.1	1.2	1.3	1.8	39.9	1.3	1.5	2.0	2.4	0.4	1.0	0.8	1.6	3.9	27.4	3.0	5.3	0.0	4.2
	Simple	6.2	0.0	4.1	5.0	0.8	8.7	0.8	3.9	8.1	9.4	1.6	7.1	0.8	2.2	4.5	17.4	7.7	8.0	2.2	1.5
	Hair	7.3	14.0	1.2	0.5	1.4	12.1	0.0	2.1	1.4	2.7	0.0	4.9	12.5	0.8	5.7	16.5	5.6	10.6	0.0	0.7
Total	Epid.	4.6	0.7	4.1	7.1	3.3	21.0	1.0	3.9	4.6	7.8	1.8	4.3	1.0	5.0	5.3	13.1	3.5	4.2	0.3	3.4
	Simple	8.3	0.1	5.1	10.1	2.1	7.1	1.2	4.7	5.3	10.9	2.5	4.3	1.2	6.2	7.4	9.5	5.4	5.4	0.4	2.8
	Hair	7.6	5.8	3.6	10.4	2.1	6.1	0.7	4.0	4.2	10.0	1.2	4.8	3.5	5.1	7.4	9.7	4.3	6.2	0.4	2.7

Table S9. Accession numbers of sequences used for multi species comparison. K, Keratin.

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	NP_006112.3		NP_032499.2	XP_001381422.1	XP_428851.2		
K2	NP_000414.2	XP_001254016.1	NP_034798.2				
K5	NP_000415.2	NP_001008663.1	NP_081287.1	XP_001362910.1	NP_001001195.1	NP_001072377.1	NP_571231.1
K9	NP_000217.2	XP_002696068.1	NP_957707.2				
K10	NP_000412.3	NP_776802.1	NP_034790.2	XP_001369459.1	XP_418163.2		
K14	NP_000517.2	NP_001160047.1	NP_058654.1		NP_001001311.2	XP_002940701.1	
K7	NP_005547.3	NP_001039876.1	NP_149064.1	XP_001377290.1			
K8	NP_002264.1	NP_001028782.1	NP_112447.2		XP_424502.2	NP_001080525.1	NP_956374.1
K18	NP_000215.1	NP_001179024.1	NP_034794.2		XP_001377493.1	NP_001089734.1	NP_848524.1
K19	NP_002267.2	NP_001015600.1	NP_032497.1	XP_001367988.1	NP_990340.1	NP_001084992.1	XP_002663760.1
K20	NP_061883.1	NP_001095321.1	NP_075745.1		XP_418168.1		
K31	NP_002268.2	NP_001069974.1	NP_034789.2				
K35	NP_002271.3	NP_001069541.1	NP_058576.2				
K81	NP_002272.2	NP_001069674.1	NP_001159629.1	XP_001362739.1			
K85	NP_002274.1	NP_001069381.1	NP_058575.2				
K86	NP_002275.1	NP_001179997.1	NP_034797.1	XP_001377307.1			

Table S10. Cysteine content in keratin domains from multiple species. The tables display cysteine count in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Cysteine content in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	2/180		5/188	2/146	0/143		
K2	3/178	2/176	3/199				
K5	2/168	1/169	2/162	1/168	1/174	1/150	0/154
K9	2/153	4/131	3/131				
K10	2/146	1/127	1/135	1/103	5/143		
K14	3/115	3/121	4/121		2/109	6/187	
K7	0/91	0/92	0/85	0/102			
K8	0/91	0/98	0/97		0/93	0/99	0/115
K18	0/80	0/79	0/72		0/77	0/55	0/84
K19	0/80	0/79	0/83	1/86	1/94	0/104	0/49
K20	0/70	0/48	0/77		0/81		
K31	9/56	8/56	10/56				
K35	5/97	6/97	5/97				
K81	11/106	9/106	12/106	11/105			
K85	12/123	9/123	12/123				
K86	12/106	10/106	12/106	12/107			

b. Cysteine content in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	0/313		3/313	0/313	2/313		
K2	2/313	2/313	4/313				
K5	2/313	2/313	2/313	2/313	1/313	4/313	0/311
K9	2/312	6/312	2/312				
K10	2/314	2/314	2/314	1/314	1/311		
K14	2/311	2/311	2/311		2/311	1/311	
K7	1/312	1/312	0/312	2/287			
K8	0/311	0/311	0/311		0/295	0/311	1/311
K18	0/312	0/312	0/313		0/312	1/312	0/311
K19	0/311	0/311	0/311	0/311	0/312	2/308	0/311
K20	1/311	2/310	1/291		1/311		
K31	7/311	8/311	12/311				
K35	10/311	10/311	10/311				
K81	9/311	9/311	10/311	11/311			
K85	9/311	8/311	11/311				
K86	9/311	9/311	10/311	10/311			

c. Cysteine content in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	1/151		1/136	0/98	-		
K2	0/148	1/128	1/195				
K5	0/109	0/119	0/105	0/107	1/112	0/102	0/93
K9	0/158	3/85	1/300				
K10	0/124	0/85	0/112	0/84	5/139		
K14	0/46	0/45	0/52		0/47	-	
K7	0/66	0/62	0/60	-			
K8	0/81	0/69	0/82		1/72	0/97	0/94
K18	0/38	0/38	0/38		0/105	0/11	0/36
K19	0/9	0/9	0/9	0/10	0/17	-	0/1
K20	0/43	0/42	0/63		0/44		
K31	9/49	10/49	10/49				
K35	8/47	8/47	8/47				
K81	14/88	14/83	10/64	14/91			
K85	6/73	6/73	6/73				
K86	8/69	10/69	11/69	9/76			

d. Cysteine content in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	3/644		9/637	2/557	2/456		
K2	5/639	5/617	8/707				
K5	4/590	3/601	4/580	3/588	3/599	5/565	0/558
K9	4/623	13/528	6/743				
K10	4/584	3/526	3/561	2/501	11/593		
K14	5/472	5/477	6/484		4/467	7/495	
K7	1/469	1/466	0/457	2/389			
K8	0/483	0/478	0/490		1/460	0/507	1/520
K18	0/430	0/429	0/423		0/494	1/378	0/431
K19	0/400	0/399	0/403	1/407	1/423	2/412	0/361
K20	1/424	2/400	1/431		1/436		
K31	25/416	26/416	32/416				
K35	23/455	24/455	23/455				
K81	34/505	32/500	32/481	36/507			
K85	27/507	23/507	29/507				
K86	29/486	29/486	33/486	31/494			

Table S11. Glycine content in keratin domains from multiple species. The tables display glycine count in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Glycine content in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	72/180		71/188	37/146	38/143		
K2	72/178	70/176	83/199				
K5	50/168	54/169	47/162	55/168	63/174	31/150	57/154
K9	73/153	48/131	60/131				
K10	70/146	56/127	64/135	37/103	50/143		
K14	42/115	46/121	46/121		38/109	23/187	
K7	11/91	11/92	10/85	17/102			
K8	14/91	13/98	15/97		10/93	14/99	18/115
K18	17/80	18/79	12/72		14/77	5/55	11/84
K19	20/80	22/79	21/83	19/86	30/94	34/104	8/49
K20	12/70	10/48	15/77		11/81		
K31	3/56	5/56	5/56				
K35	17/97	17/97	18/97				
K81	19/106	19/106	17/106	18/105			
K85	17/123	16/123	15/123				
K86	18/106	17/106	17/106	16/107			

b. Glycine content in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	5/313		6/313	6/313	6/313		
K2	8/313	7/313	8/313				
K5	7/313	8/313	8/313	8/313	10/313	6/313	8/311
K9	12/312	10/312	10/312				
K10	10/314	9/314	10/314	7/314	12/311		
K14	9/311	8/311	9/311		11/311	12/311	
K7	10/312	8/312	9/312	10/287			
K8	10/311	10/311	10/311		15/295	10/311	7/311
K18	8/312	8/312	10/313		9/312	12/312	11/311
K19	16/311	16/311	13/311	13/311	12/312	11/308	7/311
K20	9/311	10/310	6/291		8/311		
K31	2/311	6/311	5/311				
K35	4/311	5/311	6/311				
K81	8/311	8/311	8/311	7/311			
K85	8/311	9/311	8/311				
K86	8/311	8/311	8/311	8/311			

c. Glycine content in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	69/151		61/136	19/98	-		
K2	58/148	46/128	91/195				
K5	39/109	47/119	36/105	42/107	39/112	35/102	40/93
K9	92/158	32/85	159/300				
K10	70/124	43/85	63/112	20/84	44/139		
K14	2/46	2/45	2/52		2/47	-	
K7	14/66	13/62	14/60	-			
K8	12/81	8/69	13/82		10/72	22/97	18/94
K18	2/38	2/38	2/38		7/105	0/10	1/36
K19	0/9	0/9	0/9	0/10	3/17	-	0/1
K20	1/43	1/42	2/63		1/44		
K31	2/49	2/49	3/49				
K35	3/47	3/47	3/47				
K81	19/88	15/83	11/64	18/91			
K85	11/73	11/73	12/73				
K86	9/69	12/69	12/69	14/76			

d. Glycine content in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	146/644		138/637	62/557	44/456		
K2	138/639	123/617	182/707				
K5	96/590	109/601	91/580	105/588	112/599	72/565	105/558
K9	177/623	90/528	229/743				
K10	150/584	108/526	137/561	64/501	106/593		
K14	53/472	56/477	57/484		51/467	35/495	
K7	35/469	32/466	33/457	27/389			
K8	36/483	31/478	38/490		35/460	46/507	43/520
K18	27/430	28/429	23/423		30/494	17/378	23/431
K19	36/400	38/399	34/403	32/407	45/423	45/412	15/361
K20	22/424	21/400	23/431		20/436		
K31	7/416	13/416	13/416				
K35	24/455	25/455	27/455				
K81	46/505	42/500	36/481	43/507			
K85	36/507	36/507	35/507				
K86	35/486	37/486	37/486	38/494			

Table S12. Proline content in keratin domains from multiple species. The tables display proline count in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Proline content in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	5/180		5/188	6/146	3/143		
K2	6/178	7/176	5/199				
K5	7/168	7/169	6/162	6/168	7/174	9/150	5/154
K9	0/153	1/131	0/131				
K10	0/146	0/127	0/135	0/103	1/143		
K14	1/115	1/121	1/121		2/109	12/187	
K7	6/91	6/92	5/85	4/102			
K8	4/91	5/98	4/97		4/93	4/99	5/115
K18	2/80	1/79	2/72		2/77	5/55	3/84
K19	2/80	1/79	1/83	2/86	2/94	2/104	2/49
K20	2/70	2/48	2/77		2/81		
K31	7/56	6/56	7/56				
K35	5/97	5/97	5/97				
K81	8/106	7/106	8/106	8/105			
K85	8/123	8/123	11/123				
K86	8/106	8/106	8/106	8/107			

b. Proline content in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	1/313		1/313	1/313	3/313		
K2	2/313	1/313	1/313				
K5	1/313	1/313	1/313	1/313	1/313	1/313	0/311
K9	3/312	2/312	2/312				
K10	2/314	2/314	2/314	2/314	4/311		
K14	3/311	3/311	3/311		4/311	1/311	
K7	1/312	1/312	1/312	6/287			
K8	0/311	0/311	0/311		2/295	0/311	0/311
K18	2/312	3/312	2/313		2/312	2/312	3/311
K19	3/311	3/311	3/311	3/311	4/312	3/308	2/311
K20	2/311	3/310	2/291		2/311		
K31	3/311	3/311	3/311				
K35	4/311	4/311	4/311				
K81	1/311	1/311	1/311	1/311			
K85	1/311	1/311	1/311				
K86	1/311	1/311	1/311	1/311			

c. Proline content in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	1/151		1/136	1/98	-		
K2	0/148	0/128	0/195				
K5	1/109	1/119	2/105	1/107	0/112	2/102	0/93
K9	0/158	0/85	1/300				
K10	1/124	1/85	1/112	2/84	2/139		
K14	0/46	0/45	0/52		0/47	-	
K7	1/66	2/62	2/60	-			
K8	2/81	3/69	2/82		3/72	2/97	1/94
K18	0/38	0/38	0/38		2/105	3/11	0/36
K19	0/9	0/9	2/9	3/10	2/17	-	0/1
K20	0/43	0/42	0/63		0/44		
K31	12/49	10/49	14/49				
K35	11/47	11/47	9/47				
K81	3/88	4/83	3/64	5/91			
K85	5/73	4/73	5/73				
K86	3/69	3/69	4/69	3/76			

d. Proline content in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	7/644		7/637	7/557	6/456		
K2	8/639	8/617	6/707				
K5	9/590	9/601	9/580	8/588	8/599	12/56 5	5/558
K9	3/623	3/528	3/743				
K10	3/584	3/526	3/561	4/501	7/593		
K14	4/472	4/477	4/484		6/467	13/49 5	
K7	8/469	9/466	8/457	10/389			
K8	6/483	8/478	6/490		9/460	6/507	6/520
K18	4/430	4/429	4/423		6/494	10/37 8	6/431
K19	5/400	4/399	6/403	8/407	8/423	5/412	4/361
K20	4/424	5/400	4/431		4/436		
K31	22/416	19/416	24/416				
K35	20/455	20/455	18/455				
K81	12/505	12/500	12/481	14/507			
K85	14/507	13/507	17/507				
K86	12/486	12/486	13/486	12/494			

Table S13. Alanine content in keratin domains from multiple species. The tables display alanine count in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Alanine content in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	3/180		2/188	8/146	4/143		
K2	5/178	2/176	2/199				
K5	9/168	10/169	10/162	8/168	10/174	8/150	7/154
K9	4/153	7/131	2/131				
K10	0/146	1/127	0/135	3/103	12/143		
K14	5/115	6/121	4/121		9/109	6/187	
K7	10/91	7/92	7/85	9/102			
K8	4/91	6/98	5/97		8/93	10/99	7/115
K18	9/80	6/79	7/72		9/77	3/55	6/84
K19	7/80	7/79	3/83	4/86	6/94	3/104	3/49
K20	2/70	5/48	5/77		9/81		
K31	2/56	2/56	3/56				
K35	10/97	10/97	7/97				
K81	6/106	6/106	7/106	5/105			
K85	7/123	8/123	6/123				
K86	6/106	7/106	7/106	8/107			

b. Alanine content in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	20/313		18/313	19/313	15/313		
K2	22/313	22/313	18/313				
K5	25/313	25/313	24/313	21/313	29/313	27/313	31/311
K9	13/312	8/312	10/312				
K10	22/314	18/314	18/314	23/314	12/311		
K14	25/311	24/311	24/311		30/311	21/311	
K7	35/312	33/312	34/312	21/287			
K8	25/311	27/311	25/311		19/295	23/311	27/311
K18	28/312	26/312	26/313		25/312	24/312	21/311
K19	28/311	29/311	26/311	28/311	30/312	21/308	19/311
K20	21/311	20/310	22/291		24/311		
K31	16/311	18/311	18/311				
K35	22/311	20/311	19/311				
K81	29/311	31/311	33/311	28/311			
K85	33/311	32/311	38/311				
K86	29/311	30/311	33/311	27/311			

c. Alanine content in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	1/151		0/136	4/98	-		
K2	4/148	3/128	2/195				
K5	2/109	1/119	2/105	0/107	2/112	6/102	2/93
K9	2/158	0/85	2/300				
K10	0/124	0/85	1/112	6/84	7/139		
K14	0/46	0/45	0/52		3/47	-	
K7	8/66	4/62	3/60	-			
K8	4/81	1/69	2/82		5/72	0/97	3/94
K18	1/38	1/38	1/38		8/105	0/11	1/36
K19	1/9	2/9	1/9	1/10	1/17	-	0/1
K20	0/43	0/42	1/63		1/44		
K31	4/49	5/49	3/49				
K35	5/47	4/47	4/47				
K81	4/88	4/83	6/64	3/91			
K85	4/73	6/73	3/73				
K86	5/69	6/69	8/69	6/76			

d. Alanine content in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	24/644		20/637	31/557	19/456		
K2	31/639	27/617	22/707				
K5	36/590	36/601	36/580	29/588	41/599	41/565	40/558
K9	19/623	15/528	14/743				
K10	22/584	19/526	19/561	32/501	31/593		
K14	30/472	30/477	28/484		42/467	27/495	
K7	53/469	44/466	44/457	30/389			
K8	33/483	34/478	32/490		32/460	33/507	37/520
K18	38/430	33/429	34/423		42/494	27/378	28/431
K19	36/400	38/399	30/403	33/407	37/423	24/412	22/361
K20	23/424	25/400	28/431		34/436		
K31	22/416	25/416	24/416				
K35	37/455	34/455	30/455				
K81	39/505	41/500	46/481	36/507			
K85	44/507	46/507	47/507				
K86	40/486	43/486	48/486	41/494			

Table S14. Glutamate content in keratin domains from multiple species. The tables display glutamate count in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Glutamate content in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	4/180		3/188	4/146	6/143		
K2	3/178	3/176	3/199				
K5	3/168	3/169	3/162	3/168	3/174	5/150	2/154
K9	1/153	1/131	2/131				
K10	1/146	0/127	1/135	1/103	2/143		
K14	1/115	1/121	1/121		1/109	8/187	
K7	3/91	2/92	3/85	4/102			
K8	2/91	2/98	2/97		2/93	3/99	2/115
K18	1/80	1/79	1/72		1/77	0/55	2/84
K19	1/80	1/79	1/83	1/86	1/94	1/104	0/49
K20	1/70	1/48	0/77		1/81		
K31	2/56	2/56	2/56				
K35	2/97	2/97	2/97				
K81	4/106	4/106	4/106	4/105			
K85	4/123	4/123	4/123				
K86	4/106	4/106	4/106	4/107			

b. Glutamate content in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	32/313		32/313	30/313	46/313		
K2	36/313	35/313	34/313				
K5	37/313	35/313	36/313	35/313	39/313	38/313	42/311
K9	37/312	40/312	39/312				
K10	42/314	43/314	43/314	42/314	43/311		
K14	43/311	42/311	44/311		45/311	41/311	
K7	42/312	39/312	41/312	22/287			
K8	44/311	44/311	44/311		41/295	41/311	40/311
K18	41/312	41/312	41/313		42/312	36/312	39/311
K19	36/311	37/311	36/311	35/311	37/312	39/308	39/311
K20	39/311	37/310	36/291		43/311		
K31	43/311	45/311	46/311				
K35	45/311	44/311	41/311				
K81	46/311	45/311	45/311	45/311			
K85	47/311	46/311	47/311				
K86	46/311	45/311	45/311	46/311			

c. Glutamate content in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	1/151		1/136	3/98	-		
K2	1/148	2/128	2/195				
K5	1/109	1/119	1/105	1/107	1/112	1/102	1/93
K9	4/158	3/85	3/300				
K10	1/124	1/85	0/112	2/84	8/139		
K14	1/46	1/45	1/52		3/47	-	
K7	0/66	0/62	0/60	-			
K8	3/81	3/69	3/82		4/72	4/97	4/94
K18	1/38	1/38	1/38		6/105	2/11	4/36
K19	0/9	0/9	0/9	0/10	0/17	-	0/1
K20	8/43	7/42	12/63		9/44		
K31	0/49	0/49	0/49				
K35	0/47	0/47	0/47				
K81	1/88	1/83	1/64	1/91			
K85	1/73	1/73	1/73				
K86	1/69	1/69	1/69	1/76			

d. Glutamate content in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	37/644		36/637	37/557	52/456		
K2	40/639	40/617	39/707				
K5	41/590	39/601	40/580	39/588	43/599	44/565	45/558
K9	42/623	44/528	44/743				
K10	44/584	44/526	44/561	45/501	53/593		
K14	45/472	44/477	46/484		49/467	49/495	
K7	45/469	41/466	44/457	26/389			
K8	49/483	49/478	49/490		47/460	48/507	46/520
K18	43/430	43/429	43/423		49/494	38/378	45/431
K19	37/400	38/399	37/403	36/407	38/423	40/412	39/361
K20	48/424	45/400	48/431		53/436		
K31	45/416	47/416	48/416				
K35	47/455	46/455	43/455				
K81	51/505	50/500	50/481	50/507			
K85	52/507	51/507	52/507				
K86	51/486	50/486	50/486	51/494			

Table S15. Phenylalanine content in keratin domains from multiple species. The tables display phenylalanine count in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Phenylalanine content in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	13/180		15/188	6/146	6/143		
K2	15/178	15/176	18/199				
K5	13/168	14/169	13/162	11/168	9/174	7/150	12/154
K9	13/153	8/131	11/131				
K10	15/146	10/127	13/135	7/103	7/143		
K14	7/115	7/121	9/121		6/109	5/187	
K7	3/91	4/92	2/85	6/102			
K8	3/91	3/98	5/97		3/93	5/99	2/115
K18	3/80	2/79	2/72		2/77	3/55	0/84
K19	4/80	4/79	4/83	3/86	6/94	2/104	3/49
K20	3/70	1/48	3/77		2/81		
K31	3/56	4/56	2/56				
K35	4/97	6/97	4/97				
K81	5/106	6/106	5/106	5/105			
K85	4/123	4/123	5/123				
K86	4/106	4/106	4/106	3/107			

b. Phenylalanine content in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	7/313		10/313	7/313	7/313		
K2	6/313	8/313	7/313				
K5	9/313	8/313	8/313	8/313	7/313	6/313	9/311
K9	3/312	4/312	3/312				
K10	2/314	3/314	2/314	3/314	2/311		
K14	4/311	4/311	5/311		4/311	8/311	
K7	8/312	8/312	8/312	12/287			
K8	7/311	6/311	7/311		6/295	7/311	8/311
K18	5/312	4/312	5/313		5/312	3/312	5/311
K19	4/311	6/311	4/311	6/311	4/312	4/308	7/311
K20	3/311	3/310	4/291		5/311		
K31	4/311	3/311	4/311				
K35	2/311	4/311	2/311				
K81	7/311	7/311	7/311	7/311			
K85	7/311	7/311	7/311				
K86	7/311	7/311	7/311	7/311			

c. Phenylalanine content in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	1/151		1/136	2/98	-		
K2	4/148	4/128	4/195				
K5	4/109	6/119	7/105	3/107	2/112	5/102	3/93
K9	1/158	1/85	1/300				
K10	1/124	0/85	1/112	0/84	0/139		
K14	1/46	1/45	2/52		0/47	-	
K7	1/66	2/62	2/60	-			
K8	2/81	2/69	4/82		3/72	5/97	0/94
K18	0/38	0/38	0/38		2/105	0/11	0/36
K19	0/9	0/9	0/9	0/10	0/17	-	0/1
K20	0/43	0/42	0/63		0/44		
K31	1/49	0/49	1/49				
K35	1/47	1/47	1/47				
K81	0/88	0/83	0/64	0/91			
K85	2/73	2/73	2/73				
K86	0/69	0/69	0/69	0/76			

d. Phenylalanine content in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	21/644		26/637	15/557	13/456		
K2	25/639	27/617	29/707				
K5	26/590	28/601	28/580	22/588	18/599	18/565	24/558
K9	17/623	13/528	15/743				
K10	18/584	13/526	16/561	10/501	9/593		
K14	12/472	12/477	16/484		10/467	13/495	
K7	12/469	14/466	12/457	18/389			
K8	12/483	11/478	16/490		12/460	17/507	10/520
K18	8/430	6/429	7/423		9/494	6/378	5/431
K19	8/400	10/399	8/403	9/407	10/423	6/412	10/361
K20	6/424	4/400	7/431		7/436		
K31	8/416	7/416	7/416				
K35	7/455	11/455	7/455				
K81	12/505	13/500	12/481	12/507			
K85	13/507	13/507	14/507				
K86	11/486	11/486	11/486	10/494			

Table S16. Tyrosine content in keratin domains from multiple species. The tables display tyrosine count in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Tyrosine content in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	5/180		2/188	3/146	11/143		
K2	1/178	2/176	2/199				
K5	3/168	3/169	4/162	5/168	5/174	5/150	3/154
K9	7/153	2/131	3/131				
K10	5/146	7/127	6/135	5/103	5/143		
K14	4/115	5/121	2/121		6/109	5/187	
K7	2/91	2/92	3/85	1/102			
K8	3/91	3/98	1/97		1/93	4/99	6/115
K18	3/80	2/79	2/72		3/77	1/55	4/84
K19	4/80	4/79	3/83	4/86	3/94	8/104	1/49
K20	2/70	3/48	3/77		2/81		
K31	1/56	1/56	1/56				
K35	3/97	1/97	3/97				
K81	3/106	3/106	3/106	4/105			
K85	4/123	5/123	4/123				
K86	4/106	5/106	4/106	6/107			

b. Tyrosine content in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	10/313		9/313	9/313	13/313		
K2	11/313	9/313	9/313				
K5	8/313	8/313	9/313	8/313	9/313	10/313	8/311
K9	11/312	10/312	9/312				
K10	13/314	13/314	12/314	12/314	13/311		
K14	10/311	10/311	9/311		10/311	9/311	
K7	6/312	7/312	6/312	8/287			
K8	9/311	10/311	8/311		6/295	11/311	9/311
K18	8/312	8/312	8/313		8/312	7/312	7/311
K19	10/311	9/311	11/311	8/311	10/312	10/308	8/311
K20	10/311	9/310	9/291		8/311		
K31	9/311	9/311	8/311				
K35	9/311	10/311	10/311				
K81	10/311	10/311	10/311	10/311			
K85	10/311	10/311	10/311				
K86	10/311	10/311	10/311	10/311			

c. Tyrosine content in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	9/151		7/136	5/98	-		
K2	6/148	6/128	7/195				
K5	4/109	2/119	1/105	6/107	4/112	3/102	5/93
K9	9/158	3/85	20/300				
K10	7/124	6/85	7/112	3/84	0/139		
K14	0/46	0/45	0/52		1/47	-	
K7	1/66	1/62	1/60	-			
K8	3/81	3/69	3/82		2/72	5/97	10/94
K18	0/38	0/38	0/38		0/105	0/11	0/36
K19	0/9	0/9	0/9	0/10	0/17	-	0/1
K20	1/43	1/42	2/63		0/44		
K31	0/49	1/49	0/49				
K35	1/47	0/47	1/47				
K81	0/88	0/83	0/64	1/91			
K85	1/73	1/73	1/73				
K86	0/69	0/69	0/69	1/76			

d. Tyrosine content in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	24/644		18/637	17/557	24/456		
K2	18/639	17/617	18/707				
K5	15/590	13/601	14/580	19/588	18/599	18/565	16/558
K9	27/623	15/528	32/743				
K10	25/584	26/526	25/561	20/501	18/593		
K14	14/472	15/477	11/484		17/467	14/495	
K7	9/469	10/466	10/457	9/389			
K8	15/483	16/478	12/490		9/460	20/507	25/520
K18	11/430	10/429	10/423		11/494	8/378	11/431
K19	14/400	13/399	14/403	12/407	13/423	18/412	9/361
K20	13/424	13/400	14/431		10/436		
K31	10/416	11/416	9/416				
K35	13/455	11/455	14/455				
K81	13/505	13/500	13/481	15/507			
K85	15/507	16/507	15/507				
K86	14/486	15/486	14/486	17/494			

Table S17. Content of acidic amino acids in keratin domains from multiple species. The tables display count of acidic amino acids (i.e. glutamates and aspartates) in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Content of acidic amino acids in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	5/180		4/188	5/146	15/143		
K2	4/178	4/176	5/199				
K5	4/168	4/169	4/162	4/168	5/174	7/150	3/154
K9	2/153	3/131	2/131				
K10	2/146	1/127	2/135	3/103	2/143		
K14	2/115	2/121	3/121		2/109	11/187	
K7	5/91	4/92	5/85	5/102			
K8	3/91	3/98	3/97		4/93	4/99	3/115
K18	1/80	1/79	1/72		1/77	3/55	2/84
K19	2/80	2/79	2/83	2/86	2/94	2/104	3/49
K20	4/70	1/48	3/77		3/81		
K31	2/56	2/56	2/56				
K35	2/97	2/97	2/97				
K81	5/106	5/106	5/106	5/105			
K85	5/123	5/123	5/123				
K86	5/106	5/106	5/106	5/107			

b. Content of acidic amino acids in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	58/313		58/313	57/313	63/313		
K2	59/313	57/313	60/313				
K5	60/313	60/313	60/313	59/313	59/313	59/313	64/311
K9	60/312	63/312	62/312				
K10	60/314	60/314	60/314	61/314	60/311		
K14	63/311	63/311	63/311		63/311	63/311	
K7	62/312	61/312	60/312	44/287			
K8	62/311	62/311	62/311		53/295	61/311	66/311
K18	61/312	62/312	63/313		63/312	62/312	57/311
K19	58/311	58/311	56/311	57/311	60/312	60/308	63/311
K20	53/311	54/310	49/291		61/311		
K31	58/311	61/311	60/311				
K35	64/311	62/311	63/311				
K81	61/311	60/311	61/311	59/311			
K85	61/311	61/311	61/311				
K86	61/311	60/311	61/311	60/311			

c. Content of acidic amino acids in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	1/151		5/136	1/98	-		
K2	2/148	3/128	4/195				
K5	1/109	1/119	1/105	1/107	1/112	1/102	1/93
K9	4/158	4/85	7/300				
K10	1/124	1/85	1/112	5/84	11/139		
K14	4/46	4/45	4/52		5/47	-	
K7	2/66	1/62	1/60	-			
K8	5/81	5/69	5/82		6/72	6/97	7/94
K18	5/38	5/38	5/38		8/105	3/11	6/36
K19	0/9	1/9	0/9	0/10	3/17	-	0/1
K20	10/43	10/42	15/63		11/44		
K31	0/49	0/49	0/49				
K35	1/47	1/47	1/47				
K81	2/88	2/83	2/64	2/91			
K85	2/73	2/73	2/73				
K86	2/69	3/69	2/69	2/76			

d. Content of acidic amino acids in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	64/644		63/637	67/557	78/456		
K2	65/639	64/617	69/707				
K5	65/590	65/601	65/580	64/588	65/599	67/565	68/558
K9	66/623	70/528	71/743				
K10	63/584	62/526	63/561	69/501	73/593		
K14	69/472	69/477	70/484		70/467	74/495	
K7	69/469	66/466	66/457	49/389			
K8	70/483	70/478	70/490		63/460	71/507	76/520
K18	67/430	68/429	69/423		73/494	68/378	65/431
K19	60/400	61/399	58/403	59/407	65/423	62/412	66/361
K20	67/424	65/400	67/431		75/436		
K31	60/416	63/416	62/416				
K35	67/455	65/455	66/455				
K81	68/505	67/500	68/481	66/507			
K85	68/507	68/507	68/507				
K86	68/486	68/486	68/486	67/494			

Table S18. Content of basic amino acids in keratin domains from multiple species. The tables display count of basic amino acids (i.e. lysines and arginines) in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Content of basic amino acids in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	15/180		14/188	13/146	14/143		
K2	13/178	12/176	14/199				
K5	13/168	13/169	12/162	13/168	14/174	12/150	7/154
K9	8/153	8/131	6/131				
K10	7/146	6/127	5/135	6/103	10/143		
K14	5/115	5/121	6/121		6/109	19/187	
K7	11/91	9/92	10/85	10/102			
K8	9/91	10/98	10/97		10/93	8/99	9/115
K18	6/80	6/79	7/72		6/77	5/55	7/84
K19	5/80	5/79	6/83	5/86	6/94	5/104	5/49
K20	7/70	4/48	7/77		5/81		
K31	2/56	2/56	2/56				
K35	8/97	9/97	10/97				
K81	9/106	11/106	9/106	10/105			
K85	13/123	13/123	13/123				
K86	9/106	10/106	9/106	10/107			

b. Content of basic amino acids in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	43/313		45/313	43/313	46/313		
K2	44/313	45/313	45/313				
K5	48/313	48/313	48/313	48/313	48/313	51/313	50/311
K9	35/312	38/312	37/312				
K10	36/314	37/314	37/314	35/314	37/311		
K14	43/311	43/311	43/311		42/311	47/311	
K7	45/312	47/312	44/312	40/287			
K8	47/311	47/311	47/311		48/295	47/311	49/311
K18	44/312	43/312	43/313		47/312	48/312	44/311
K19	39/311	38/311	40/311	39/311	39/312	41/308	38/311
K20	41/311	40/310	37/291		43/311		
K31	36/311	38/311	38/311				
K35	37/311	36/311	37/311				
K81	46/311	47/311	48/311	47/311			
K85	48/311	47/311	48/311				
K86	47/311	47/311	48/311	47/311			

c. Content of basic amino acids in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	8/151		8/136	8/98	-		
K2	10/148	11/128	13/195				
K5	5/109	5/119	6/105	5/107	6/112	5/102	4/93
K9	7/158	7/85	14/300				
K10	4/124	4/85	4/112	6/84	13/139		
K14	7/46	7/45	7/52		6/47	-	
K7	5/66	6/62	5/60	-			
K8	8/81	8/69	8/82		7/72	8/97	7/94
K18	6/38	6/38	6/38		21/105	1/11	7/36
K19	1/9	1/9	1/9	1/10	4/17	-	0/1
K20	8/43	8/42	12/63		8/44		
K31	4/49	4/49	4/49				
K35	4/47	4/47	4/47				
K81	4/88	4/83	4/64	5/91			
K85	5/73	5/73	5/73				
K86	5/69	4/69	4/69	5/76			

d. Content of basic amino acids in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	66/644		67/637	64/557	60/456		
K2	67/639	68/617	72/707				
K5	66/590	66/601	66/580	66/588	68/599	68/565	61/558
K9	50/623	53/528	57/743				
K10	47/584	47/526	46/561	47/501	60/593		
K14	55/472	55/477	56/484		54/467	66/495	
K7	61/469	62/466	59/457	50/389			
K8	64/483	65/478	65/490		65/460	63/507	65/520
K18	56/430	55/429	56/423		76/494	54/378	58/431
K19	45/400	44/399	47/403	45/407	49/423	46/412	44/361
K20	56/424	52/400	56/431		56/436		
K31	42/416	44/416	44/416				
K35	49/455	49/455	51/455				
K81	59/505	62/500	61/481	62/507			
K85	66/507	65/507	66/507				
K86	61/486	61/486	61/486	62/494			

Table S19. Content of aromatic amino acids in keratin domains from multiple species. The tables display count of aromatic amino acids (i.e. phenylalanines, tyrosines plus tryphophans) in the highlighted keratin domain together with the respective domain amino acid length. K, Keratin.

a. Content of aromatic amino acids in keratin head domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	18/180		17/188	9/146	19/143		
K2	16/178	17/176	20/199				
K5	16/168	17/169	17/162	16/168	14/174	12/150	15/154
K9	20/153	11/131	15/131				
K10	20/146	17/127	19/135	12/103	13/143		
K14	11/115	12/121	11/121		12/109	11/187	
K7	5/91	6/92	5/85	7/102			
K8	6/91	6/98	6/97		4/93	9/99	8/115
K18	6/80	5/79	5/72		6/77	4/55	4/84
K19	8/80	8/79	7/83	7/86	9/94	10/104	4/49
K20	5/70	4/48	6/77		4/81		
K31	5/56	6/56	4/56				
K35	8/97	8/97	8/97				
K81	8/106	9/106	8/106	9/105			
K85	8/123	9/123	9/123				
K86	8/106	9/106	8/106	9/107			

b. Content of aromatic amino acids in keratin rod domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	18/313		20/313	17/313	21/313		
K2	18/313	18/313	17/313				
K5	19/313	18/313	19/313	18/313	18/313	18/313	19/311
K9	15/312	15/312	13/312				
K10	17/314	18/314	16/314	17/314	17/311		
K14	16/311	16/311	16/311		16/311	19/311	
K7	16/312	16/312	16/312	22/287			
K8	17/311	17/311	16/311		15/295	19/311	19/311
K18	15/312	14/312	15/313		15/312	11/312	13/311
K19	16/311	17/311	16/311	16/311	16/312	16/308	17/311
K20	14/311	13/310	14/291		14/311		
K31	14/311	13/311	13/311				
K35	13/311	16/311	14/311				
K81	18/311	18/311	19/311	19/311			
K85	19/311	19/311	19/311				
K86	18/311	18/311	19/311	19/311			

c. Content of aromatic amino acids in keratin tail domain

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	10/151		8/136	7/98	-		
K2	10/148	10/128	11/195				
K5	8/109	8/119	8/105	9/107	6/112	8/102	8/93
K9	10/158	4/85	21/300				
K10	8/124	6/85	8/112	3/84	1/139		
K14	1/46	1/45	2/52		1/47	-	
K7	2/66	3/62	3/60	-			
K8	5/81	5/69	7/82		5/72	10/97	10/94
K18	0/38	0/38	0/38		4/105	0/11	0/36
K19	0/9	0/9	0/9	0/10	0/17	-	0/1
K20	1/43	1/42	2/63		0/44		
K31	1/49	1/49	1/49				
K35	1/47	0/47	1/47				
K81	0/88	0/83	0/64	1/91			
K85	1/73	1/73	1/73				
K86	0/69	0/69	0/69	1/76			

d. Content of aromatic amino acids in the whole keratin sequence

	Human	Cow	Mouse	Opossum	Chicken	Frog	Zebrafish
K1	46/644		45/637	33/557	40/456		
K2	44/639	45/617	48/707				
K5	43/590	43/601	44/580	43/588	38/599	38/565	42/558
K9	45/623	30/528	49/743				
K10	45/584	41/526	43/561	32/501	31/593		
K14	28/472	29/477	29/484		29/467	30/495	
K7	23/469	25/466	24/457	29/389			
K8	28/483	28/478	29/490		24/460	38/507	37/520
K18	21/430	19/429	20/423		25/494	15/378	17/431
K19	24/400	25/399	23/403	23/407	25/423	26/412	21/361
K20	20/424	18/400	22/431		18/436		
K31	20/416	20/416	18/416				
K35	23/455	25/455	24/455				
K81	26/505	27/500	27/481	29/507			
K85	30/507	31/507	31/507				
K86	26/486	27/486	27/486	29/494			