

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 \sim 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

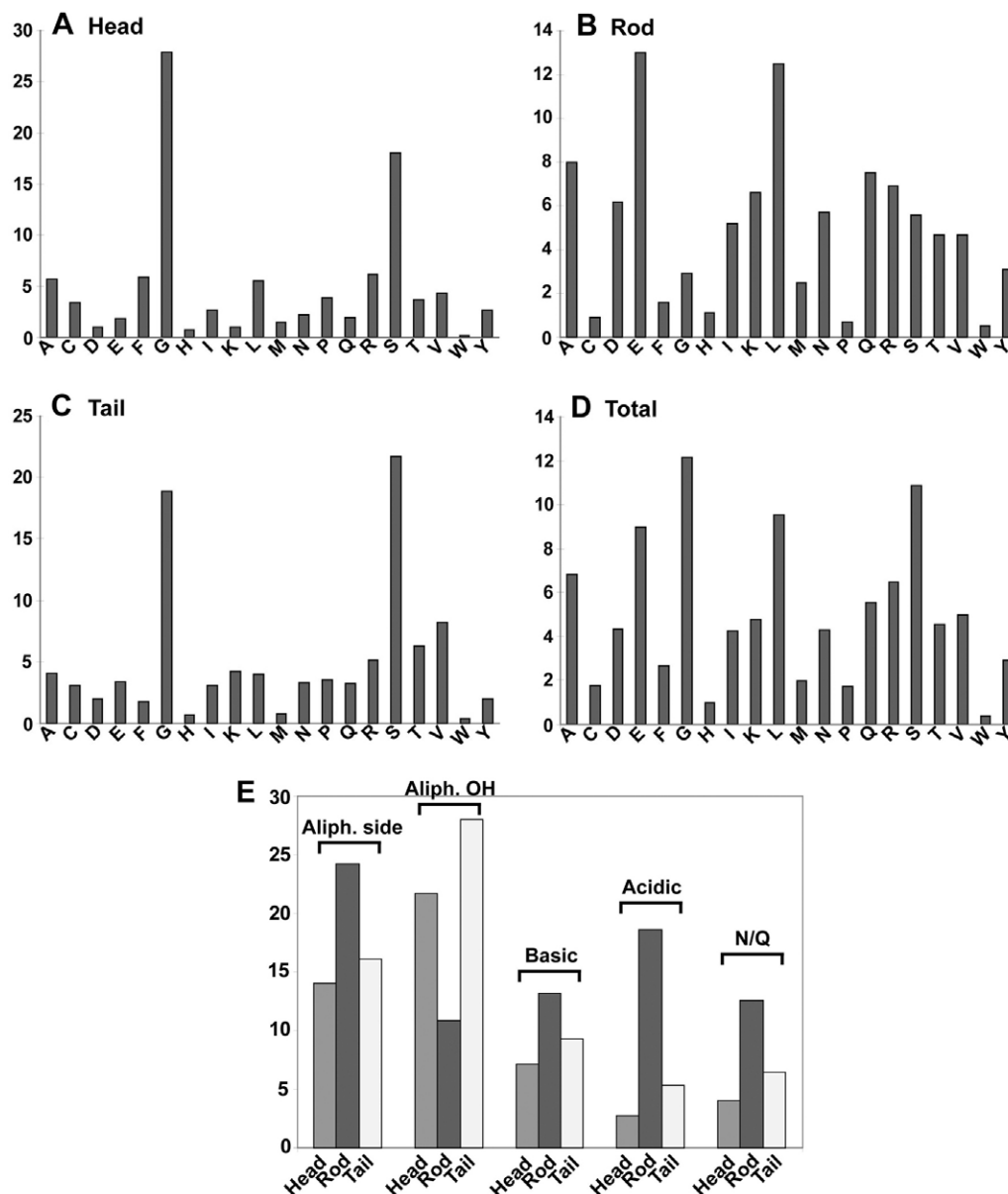


Fig. 1. Amino acid composition differs significantly between the keratin head, rod and tail domains. To study the amino acid composition of keratin domains, the human keratin sequences K1–K10, K12–K14, K16–K20, K23, K24, K31, K35, K81, K85 and K86 were used. For each amino acid (denoted by the single 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

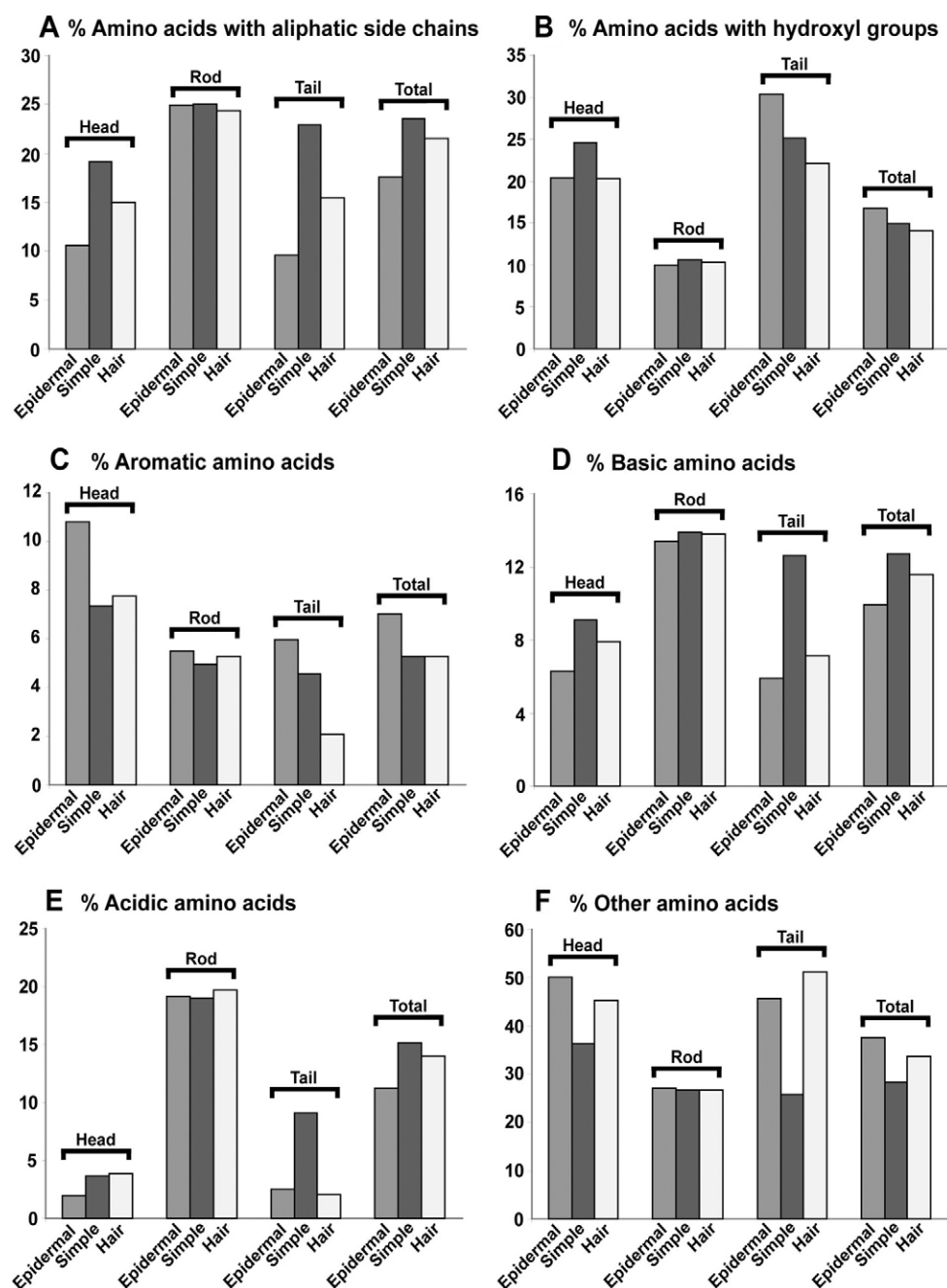


Fig. 2. Amino acid composition of human epidermal, simple-type epithelial and hair keratins. The percentage share of selected amino acid subgroups (A–F) within keratin domains and the whole keratin sequence was calculated. Grouping of amino acids as shown in Fig. 1. K1, K2, K5, K9, K10 and K14 sequences 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

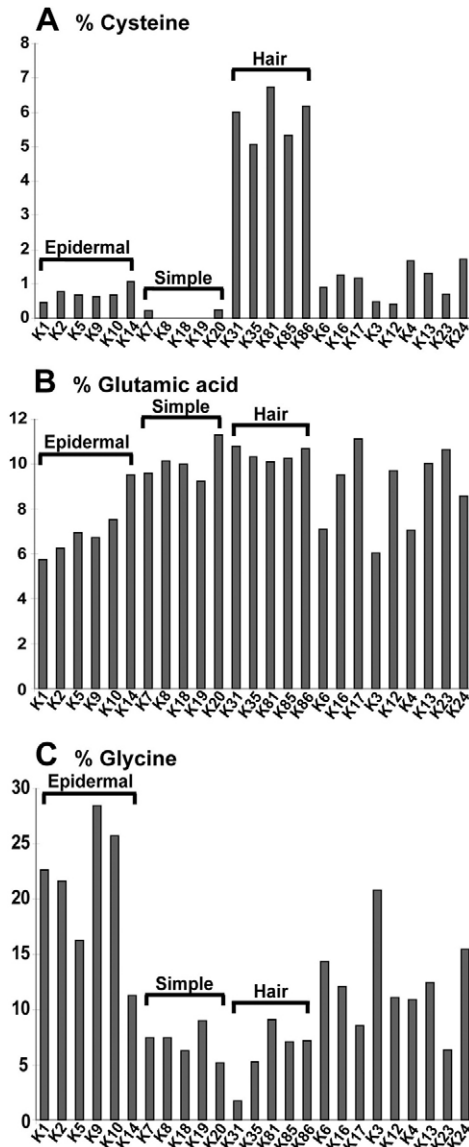


Fig. 3. Percentage of selected amino acids in human keratins. The percentage share of (A) cysteine, (B) glutamic acid and (C) glycine was calculated for the whole keratin sequence. Note that epidermal, simple-type epithelial and hair keratins have different amino acid compositions. For example, cysteine 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

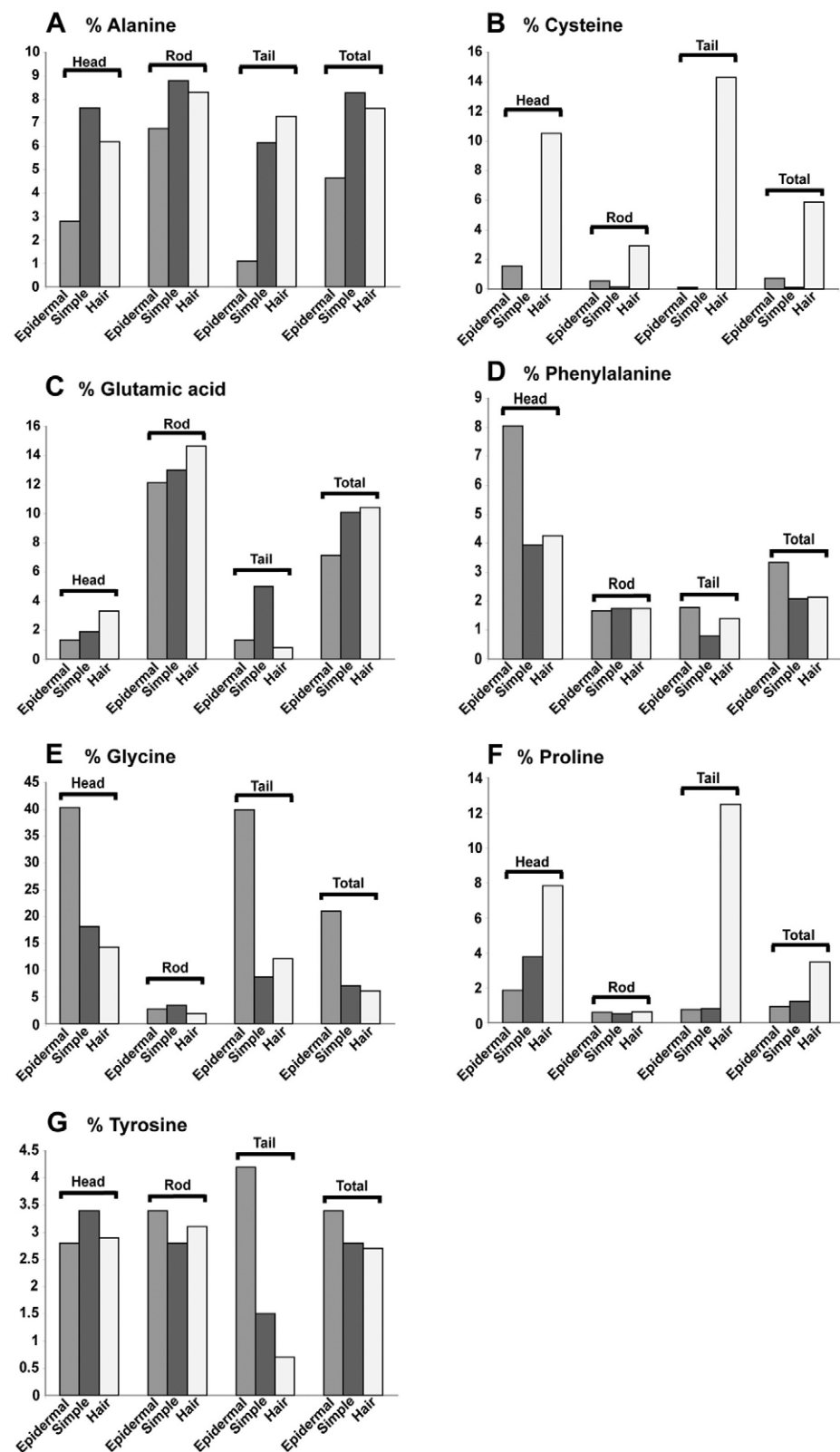


Fig. 4. Distribution of selected amino acids within epidermal, simple-type epithelial and hair keratins. The percentage share of (A) alanine (B) cysteine, (C) glutamic acid, (D) phenylalanine, (E) glycine (F) proline and (G) tyrosine within keratin domains and the whole keratin sequence is shown. Chain 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

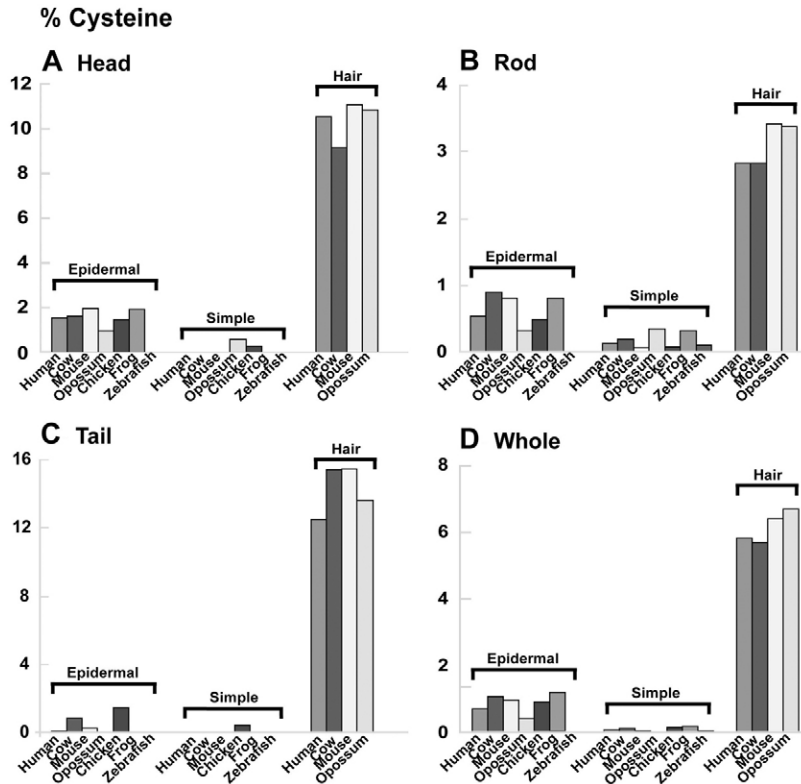


Fig. 5. Cysteine content of epidermal, simple-type epithelial and hair keratins among multiple species. The average cysteine content within keratin domains (A–C) and the whole keratin sequence (D) in selected species. Chain designation is as shown in Fig. 3. Note that irrespective of the analyzed species, 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 (Omary 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 Omary, 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,

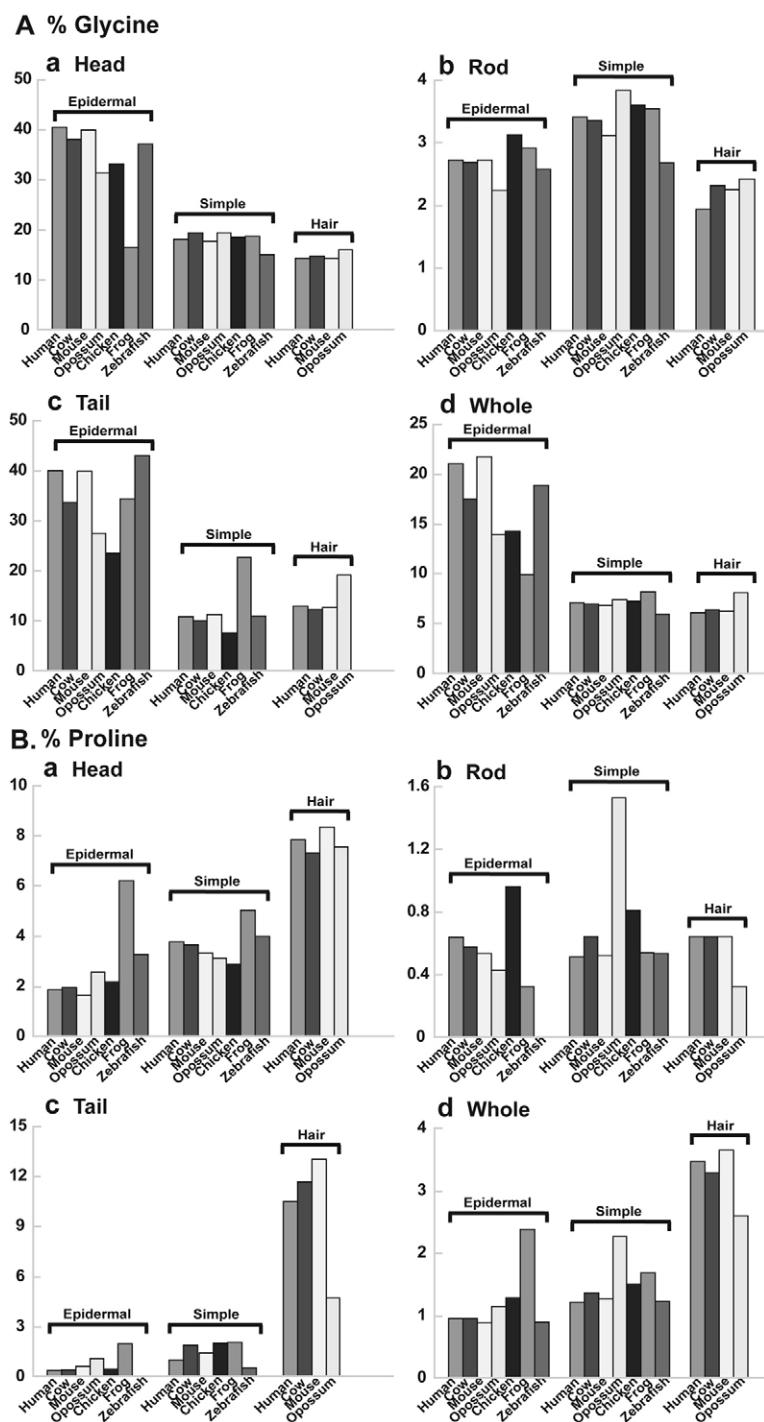


Fig. 6. Glycine and proline content of epidermal, simple-type epithelial and hair keratins in a range of species. The average glycine (A) and proline (B) content within keratin domains and the whole keratin sequence in selected species. Chain designation is as shown in Fig. 3. Note that, independent 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 (Omary 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 (Omary 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] \times (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

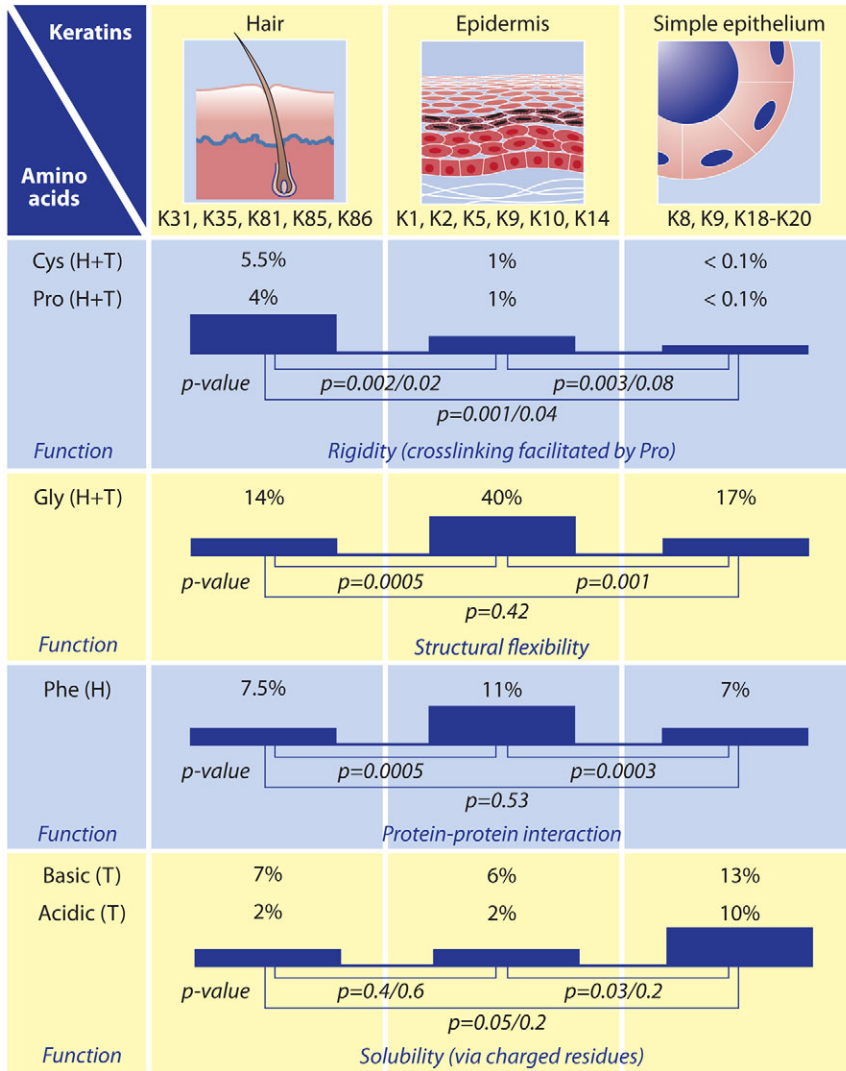


Fig. 7. Summary of the major differences in amino acid composition of hair, epidermal and simple-type epithelial keratin structural domains. The schematic summarizes the major differences in amino acid content of hair, epidermal and simple-type epithelial keratins together with their respective *P*-values. 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 rods 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

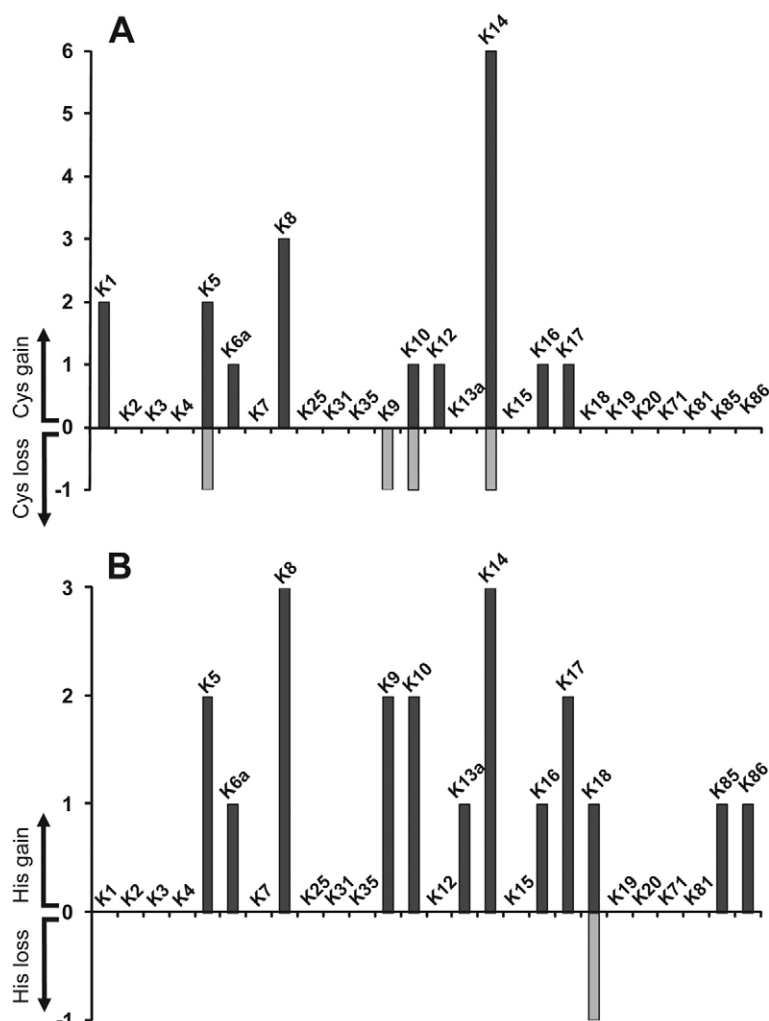


Fig. 8. Overview of missense substitutions within selected human keratins and the resulting loss or gain of cysteine (A) and histidine (B) residues. Data are based on the Human Intermediate Filament database (www.interfil.org), which lists 184 and 169 unique missense substitutions for the analyzed type 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.
- Chipev, 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.
- Crewther, 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 Aebi, U. (2002). Characterization of early assembly intermediates of recombinant human keratins. *J. Struct. Biol.* **137**, 82-96.
- Herrmann, H., Strelkov, S. V., Burkhard, P. and Aebi, 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., Keffe, 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.
- Navsaria, H. A., Swenson, 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. 1. 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., Burkhard, P., Aebi, 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., Schulthess, 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 M_r 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, loricins and single-stranded RNA binding proteins. *Int. J. Biol. Macromol.* **13**, 130-139.
- Strnad, P., Zhou, Q., Hanada, S., Lazzaroni, 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.