1. ABSTRACT

On human chromosome 1q21, a 2-Mb region called the epidermal differentiation complex comprises many genes encoding structural and regulatory proteins that are of crucial importance for keratinocyte differentiation and stratum corneum properties. Apart from those for involucrin and loricrin, most of the genes are organized in four families: the genes encoding EF-hand calcium-binding proteins of the S100A family, the genes encoding the small proline rich proteins (SPRRs) and the late cornified envelope (LCE) proteins, two families of cornified cell envelope components, and the genes encoding the S100-fused type proteins (SFTPs). This review focuses on the SPRRs, LCE proteins and SFTPs. It describes their structures, their specific functions and, when known, the mechanisms involved in the regulation of their expression. It also highlights their possible involvement in skin diseases.

2. INTRODUCTION

The skin is the largest human organ (~1.8 m² in the adult). At the interface between the body and its outside environment, it provides mechanical protection and defends against the invasion of pathogens, the entry of undesirable chemicals such as allergens or toxins, the disastrous effects of oxidative reagents and ultraviolet (UV) radiation, and uncontrolled loss of water and solutes (Figure 1). These functions, collectively called “barrier functions”, are largely carried out by the epidermis, a stratified squamous epithelium, and particularly by its outermost layer, the stratum corneum (also known as the cornified or horny layer) (1-6). Premature newborns with incomplete acquisition of the epidermal barrier functions (2), common inflammatory skin diseases, such as psoriasis and atopic dermatitis (AD), with decreased barrier functions (6-8), and animal models underscore the essential role of the stratum corneum (9, 10).
Figure 1. Functions of the stratum corneum. The cornified layer is a barrier between the body and the external environment. It prevents the penetration of allergens, pathogens (viruses and bacteria), chemical irritants and UV radiation, and the loss of body fluids. It is also a mechanical protection and an antioxidant barrier.

The stratum corneum is a highly organized and resilient layer approximately 20 µm thick. It is formed by the stacking of the so-called corneocytes, the end-products of keratinocyte terminal differentiation (4). The keratinocyte differentiation is an oriented program of gene expression starting in the basal layer and ending at the surface of the epidermis. During their journey through the spinous and granular layers, the cells turn specific genes on and off, and undergo a series of metabolic and structural changes in sequence (1). For instance, the expression of KRT1 and KRT10, encoding keratin 1 and 10, starts in the spinous layer, whereas filaggrin is detected in the granular layer and lower stratum corneum (11). Finally, the granular keratinocytes undergo a specialized form of programmed cell death called cornification.

Cornification is characterized by the elimination of all organelles and of the nucleus, by the aggregation of intermediate filaments to form an intracellular fibrous matrix, and by the assembly of a resistant protein-shell at the keratinocyte periphery, the cornified cell envelope (CE) (12). At the same time, desmosomes, the intercellular junctional structures, are transformed into corneodesmosomes after the addition of corneodesmosin (13). The CE, together with corneodesmosomes, is critical for the stratum corneum barrier functions since it confers resistance on the layer (12, 13). In addition, it is involved in the structural organization of lipids filling the inter-corneocyte spaces in the form of lamellae after their secretion by lamellar bodies. This extracellular hydrophobic matrix, enriched in cholesterol, ceramides and free fatty acids, plays a major role in the watertightness of the horny layer (14).

A 2-Mb region on the locus 1q21 comprises sixty genes, many of them encoding structural and regulatory proteins that are of crucial importance for keratinocyte differentiation and stratum corneum properties (Figure 2), i.e. intermediate filament-associated proteins (abbreviated IFAP), calcium-binding proteins and proteins involved in the CE formation. The region is known as the “epidermal differentiation complex” (EDC) (15-17).

3. DISCOVERY OF THE EPIDERMAL DIFFERENTIATION COMPLEX

The end of the eighties saw the publication of the first localization in human chromosome 1q21, by fluorescence in-situ hybridization, of genes involved in the very stable ε-(γ-glutamyl)lysine isopeptide bonds between various precursors of a proteinaceous nature, including involucrin and loricrin. This reaction is catalyzed by calcium-dependent enzymes called transglutaminases (EC 2.3.2.13). The CE, together with corneodesmosomes, is critical for the stratum corneum barrier functions since it confers resistance on the layer (12, 13). In addition, it is involved in the structural organization of lipids filling the inter-corneocyte spaces in the form of lamellae after their secretion by lamellar bodies. This extracellular hydrophobic matrix, enriched in cholesterol, ceramides and free fatty acids, plays a major role in the watertightness of the horny layer (14).
late steps of epidermal differentiation. These are the genes encoding loricrin (18), involucrin (19), profilaggrin (20) and trichohyalin (21). From 1992 came the idea that the genes of CE precursors, including loricrin, involucrin and members of the small proline-rich protein (SPRR, formerly SPR) family, may derive from a common gene ancestor (22). The term “functional cluster” of epidermal differentiation genes was then used (23). The name “EDC” was suggested for the first time in 1996 to define this large region of human chromosome 1 and its region of synteny in mouse chromosome 3F1.F2 (15, 24). Since the beginning of the second millennium, the entire sequence of the human genome has been known. This has provided an accurate and definitive view of all the genes clustered in 1q21. Apart from the genes for involucrin and loricrin, most are organized in four families, (a) two families of genes encoding proteins of the CE, namely the late cornified envelope (LCE) proteins and the SPRRs, (b) genes encoding EF-hand calcium-binding proteins of the S100A family, (c) genes encoding the seven known S100-fused type proteins (SFTPs): profilaggrin, hornerin, filaggrin-2 (also known as ifapsoriasin), repetin, cornulin, trichohyalin and trichohyalin-like 1. Analysis of the GenBank and Ensembl databases has revealed the existence of the EDC throughout evolution, in rodents, marsupials and birds but not in fishes (25). The identification of involucrin, cornulin and SPRR genes in Gallus Gallus indicates that they were present in a common ancestor of birds and mammals >300 million years ago (16). Since the function and properties of involucrin (26), loricrin (12) and S100A proteins (27) are well known, this review will focus on the LCEs, SPRRs and SFTPs, the more recently identified and less well characterized protein families.

4. SMALL PROLINE-RICH AND LATE CORNIFIED ENVELOPE PROTEINS

4.1. SPRRs

The SPRRs form a class of ten keratinocyte-specific closely related human proteins, the expression of which is strongly induced during differentiation and after UV or phorbol ester treatments (28-30). Four groups have been defined on the basis of their amino acid sequences and tandem repetitions (Figure 3): SPRR1A and 1B (also known as cornifin A and B, SPRK and 19 kDa pancornulins), SPRR2 (six different proteins, SPRR2A-2G, SPRR2C being a pseudogene), SPRR3 (a single member formerly called esophagin, cornifin-beta or 22 kDa pancornulin) and SPRR4. These are all short polypeptides (72-89 amino acids long, 8-10 kDa predicted molecular mass) except for SPRR3 (169 amino acids, 18.2 kDa). They all consist of a conserved NH2-terminal domain, three to sixteen 9-8 amino acids long tandemly repeated units that are highly enriched in proline residues, and a conserved COOH-terminal domain (28-30). They seem to have little or no secondary structure in solution, the central domain consisting of beta-turns, and being flexible (31). The head and tail regions of SPRRs are essential for incorporation into the CE, the residues used by transglutaminases 1 and 3 both in vitro and in vivo for crosslinking being mainly localized in these domains. For example in SPRR3, only Gln4, Lys6, Gln17 and Gln25, and Glu158, Lys161 and Lys163 are used (32-34). SPRRs are mainly cross-linked to loricrin, the ratio of loricrin to SPRRs in the CE varying in the epidermis from one body site to another or from the epidermis to another epithelium. This ratio is suspected to be involved in the alteration of the physicochemical properties of the envelopes, including their mechanical properties (12, 34, 35 and references therein). On the other hand, it has been proposed that the central repetitive domain interacts with...
Figure 3. Alignment of human SPRR amino acid sequences. The NH2- and COOH-terminal domains of the various SPRRs are shown by double arrows. The numbers correspond to the amino acid number of each SPRR. The amino acids identical at a given position in at least 5 different SPRRs are in a gray box. The 8-9 amino acid repeats are indicated by a black rectangle on the sequences of SPRR1A, SPRR2A, SPRR3 and SPRR4. Gaps introduced to increase the identities are shown by dashes. UniProtKB/SwissProt accession numbers are the following: P35321/SPR1A, P22528/SPR1B, P35326/SPR2A, P35325/SPR2B, P22532/SPR2D, P22531/SPR2E, Q96RM1/SPR2F, Q9BYE4/SPR2G, Q9UBC9/SPR3 and Q96PI1/SPR4.

and coils around the helical solenoid formed by the repetitive central domain of involucrin (31).

The expression of SPRR genes has been carefully analyzed by semi-quantitative RT-PCR experiments (30). All of them except SPRR2F are induced during calcium-mediated in vitro keratinocyte differentiation. In the stratified epithelia tested (from skin, oesophagus, cervix, uterus and ovary), differential expression was observed. The expression of most of them is low or null in the non-squamous epithelia. All but SPRR4 are expressed in the oesophagus and cervix, and all but SPRR3 are expressed in the skin, in either the inter- or intra-follicular epidermis. SPRR4 expression has been essentially detected in the granular keratinocytes. In addition, individual SPRR genes have been shown to respond selectively to in vitro UV irradiation of keratinocytes: SPRR4 and SPRR2G, the most specific genes for epidermis, are also the only ones to be strongly induced by UV light (30). SPRR4 in particular is strongly induced in normal human epidermis after chronic UV exposure and incorporated into fragile CEs, concomitant with a thickening of the stratum corneum (36).

An overexpression of SPRR2 also occurs in both bronchial and intestinal epithelium during allergic inflammation (37). These data suggest that SPRRs are part of an adaptive tissue response to environmental stress. In favor of this hypothesis, SPRRs, both as isolated proteins and in purified CEs, seem to have a role in the quenching of reactive oxygen species, resulting in inter- and intramolecular S-S bond formation (38). Again in agreement, some SPRR genes are upregulated in psoriatic lesions (SPRR2B, SPRR2G and SPRR3) and in the epidermis of loricrin-deficient mice (Sprr2a and 2h) (39, 40). This result is evocative of a compensatory mechanism to maintain the skin barrier function.

It should be noted that the Sprr2 group contains four additional proteins in the mouse: Sprr2h-k (Sprr2j gene is probably a pseudogene; compared to other members of the family it presents a frameshift mutation after the codon encoding amino acid 20). Twelve of the fourteen mouse Sprr genes are upregulated in mice with a target ablation of the transcription factor Kruppel-like factor 4. The sequences upstream of the transcription start site of the Sprr2 genes contain common putative binding sites for Fos/Jun (AP-1), Oct, Ets and Kruppel-like zinc-finger transcription factors. These regulatory elements are conserved in human SPRR2 genes (41). The clustering of these genes and their pattern of expression suggest that they may be held together in a single locus to allow coordinated
Figure 4. Alignment and phylogenetic tree of human LCE protein amino acid sequences. (A): Sequence alignment. The numbers correspond to the amino acid number of each LCE protein. The amino acids identical at a given position in at least 7 different proteins are in bold. Gaps introduced to increase the identities are shown by dashes. (B) Phylogenetic tree obtained using UniProt tools. UniProtKB/SwissProt accession numbers of the various paralogous proteins are the following: Q9BYE3/LCE3D, Q5T5B0/LCE3E, Q5T5A8/LCE3C, Q5TA77/LCE3B, Q5TA76/LCE3A, Q5T752/LCE1D, Q5T7P3/LCE1B, Q5T7P2/LCE1A, Q5T754/LCE1F, Q5T753/LCE1E, Q5T751/LCE1C, Q5TA79/LCE2A, O14633/LCE2B, Q5TA81/LCE2C, Q5TA82/LCE2D, Q5TCM9/LCE5A, Q5TA78/LCE4A and A0A183/LCE6A.
regulation. Interestingly, a conserved non-coding sequence serves as a long-range strong enhancer of the Sprr genes in stress conditions only (42).

In addition, LELP1 (also referenced as late cornified envelope-like proline-rich protein 1) encoded by a gene located on the telomeric side of SPRR2G contains five repeats related to those of SPRR2s but an unrelated head domain and no tail domain (our unpublished observations). Its function is completely unknown.

4.2. LCE proteins

The eighteen genes of this family of proteins span over 320 kb within the EDC, where they are located between the involucrin gene and those encoding the SFTPs (43). The “late cornified envelope” (LCE) generic name resumes the various terms used in the past: LEP for “late envelope protein”, SPRL for “small proline-rich like” and xp for “skin specific protein”. The LCE genes can be divided into three groups designated LCE1 (seven different LCE1A-F genes), LCE2 (LCE2A-D) and LCE3 (LCE3A-E), based on chromosomal position and protein homology. The
**Epidermal differentiation complex**

LCE1, LCE2 and LCE3 clusters are interspersed with three individual LCE genes (LCE4A, LCE5A and LCE6A), four LCE pseudogenes (LCEP1-4) and several non-LCE genes (NICE-1 or Clorf42, xp33 or Clorf46, xp32 or LEP7 or Clorf68 and hKPRP or Clorf45). The LCE genes generally comprise three exons, the first one being non-coding. The LCE proteins, 80-118 amino acids long (Figure 4), are serine-, glycine- and cysteine-rich polypeptides and, unlike SPRRs, do not contain any tandemly repeated sequences. Most of the LCE genes are expressed late during keratinocyte differentiation, in the granular keratinocytes. They are among the last CE components to be cross-linked to the structure. Using real-time PCR, Jackson et al. (43) found that the LCE1 and LCE2 genes are expressed in the epidermis, particularly LCE1C, LCE2A and LCE2B, whereas the LCE3 genes show low-level, tissue-specific expression in stratified non-cornified epithelia, including those of the oesophagus and tongue. LCE6A expression is highly upregulated during keratinocyte differentiation (44). When the expression of most LCE genes was analyzed in a large panel of human tissues (45), most samples have demonstrated very low to undetectable mRNA levels. Moderate expression has been observed in some oral epithelia and trunk skin. High expression level, in particular of LCE6A, has been noted in plantar skin. The LCE genes responded in a group-specific manner to calcium and UV light, with calcium causing upregulation of LCE2, and UV light causing upregulation of LCE1 and LCE2 (43).

In the mouse, there are only three groups of LCE proteins: eleven Lce1 proteins (Lce1a1, Lce1a2, Lce1b-j, Lce1l and Lce1m), four Lce3 proteins (Lce3a-c and Lce3f) and a unique Lce6a. Lce and Sprr genes are differentially regulated in mice with a target ablation of the Grainyhead-like 3 transcription factor, e.g. Sprr2a, Sprr2d and Lce3c are up regulated whereas Sprr1a and Lce1m are down regulated (46).

Among the several psoriasis susceptibility loci identified in genome-wide search analyses, PSORS4 has been shown to comprise the EDC (47). More recently, a genome-wide scan for copy number variants showed that a deletion of LCE3B and LCE3C (LCE3C_LCE3B-del) was significantly associated with susceptibility to the disease in a cohort of 557 patients from Spain (48). This finding has been replicated in additional cohorts from several European and North American countries (a total of 2,831 cases and a combined p value of 1.38 x 10^-8), and also in a family-based study of 2,473 individuals (p=5.4 x 10^-4) (48). In addition, a single-nucleotide polymorphism (SNP) in strong linkage disequilibrium with LCE genes has been linked to psoriasis susceptibility in a Chinese cohort (49), suggesting that LCE gene deletion is a general risk factor independent of the ethnic origin. However, this deletion is not associated with AD susceptibility (50). Quantitative expression studies using RT-PCR have shown that the expression of LCE genes is differentially regulated in psoriatic epidermis: all members of the LCE1, 2, 5 and 6 groups are down-regulated whereas all genes of the LCE3 group are induced. After tape stripping, the LCE genes are similarly either down- or up-regulated (45). A reconstructed skin model has been used to show that a combination of the psoriasis-associated proinflammatory cytokines TNF-alpha, IL-1alpha and IL6 induce the expression of LCE3 genes (45). It has been speculated that LCE3 proteins may have a role in repairing a defective and stressed CE (45). In psoriasis, deletion of LCE3B and LCE3C could potentially affect the epidermal barrier repair in response to skin injury. In consequence, environmental antigens may penetrate through the altered stratum corneum, leading to activation of innate immune mechanisms in predisposed individuals.

5. FAMILY OF THE S100-FUSED TYPE PROTEINS

The genes encoding the SFTPs are clustered in a region spanning ~330 kb. They all display the same intron/exon organization, with a small 5’ non-coding exon, a small second exon containing the translation start site, and an unusually large third exon encoding the major part of the protein (Figure 5A). The SFTPs also share the same structural organization at the protein level. They all contain an N-terminal domain homologous to S100A proteins which is fused to a large central domain consisting of multiple tandem repeats of one or more sequences, and to a C-terminal region. The latter is specific to each SFTP (Figure 5B). Except for the S100 domain (Figure 5C), the amino acid sequences of the SFTPs are not really related. Like S100A proteins, the N-terminal region of SFTPs comprises two EF-hand calcium-binding domains. In vitro overlay assays with radioactive calcium have been performed to test for the functionality of profilaggrin, trichohyalin, repetin, and cornulin EF-hands sites. The results clearly demonstrate that they bind calcium in a reversible manner (51, 52). This property has also been shown for the S100 domain of mouse filaggrin-2 (53). The precise function of this N-terminal region is as yet unknown, although its involvement in the regulation of intracellular calcium levels and terminal differentiation program may be suspected. The central repetitive domain is believed to impart their specific functions to each of the proteins. However, the repetitive domain of hornerin and filaggrin-2 exhibit the highest sequence homology with filaggrin (41.5% and 42% respectively). In addition, the amino acid compositions of the three proteins are well conserved, with high levels of serine, glycine, arginine, histidine and glutamine, these 5 amino acids accounting for more than 70% of the global composition of the three proteins (Table 1). This suggests that their functions may also be conserved (see 4.1 to 4.3). Finally, the role if any of the C-terminal domain of SFTPs remains to be discovered. Another common feature of the SFTPs is their specific expression in the stratified cornified epithelia and/or the hair follicles.

5.1. Profilaggrin

Profilaggrin is the best characterized and the most studied of the SFTPs. Since many excellent reviews have already been published, e.g. see (11, 54-56), we only mention the major points concerning this protein here. Profilaggrin (400 kDa in human) consists of 10-12 tandem repeats of filaggrin joined by a short hydrophobic linker peptide (7 amino acids long), and flanked by two truncated repeats and by NH₂- and COOH-unique terminal domains (Figure 5B) (57). It is synthesized by granular
keratinocytes, where it is stored in the cytoplasmic keratohyalin granules as a phosphorylated precursor. Several kinases including casein kinase II are suspected to be involved in profilaggrin phosphorylation (54). In rodents, profilaggrin is formed by at least twenty identical filaggrin subunits (58). During cornification, profilaggrin is dephosphorylated and proteolytically processed into mature basic filaggrin monomers (324 amino acids; 37 kDa in human). The S100-related amino-terminal domain released is translocated to the nucleus of transitional cells, thanks to a functional bipartite nuclear localization signal (59). At the same time, filaggrin associates with intermediate filaments to promote the formation of macrofibrils and the fibrous matrix (61-63) and promotes degradation of the protein into derivatives. In addition, part of filaggrin is also incorporated in CEs, together with some molecules of keratins (69, 70).

### Table 1. Amino acid composition (%) of the repeated domains of the seven S100-fused type proteins

<table>
<thead>
<tr>
<th></th>
<th>FLG2</th>
<th>HRNR</th>
<th>FLG2</th>
<th>trichohyalin</th>
<th>repetin</th>
<th>cornulin</th>
<th>trichohyalin-like 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>25.3</td>
<td>38.3</td>
<td>23.4</td>
<td>1.2</td>
<td>12.8</td>
<td>11.7</td>
<td>9.7</td>
</tr>
<tr>
<td>Gly</td>
<td>13.6</td>
<td>25.9</td>
<td>22.0</td>
<td>13.8</td>
<td>11.1</td>
<td>13.1</td>
<td>8.0</td>
</tr>
<tr>
<td>Arg</td>
<td>11.2</td>
<td>5.9</td>
<td>23.9</td>
<td>8.8</td>
<td>7.1</td>
<td>4.7</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>10.7</td>
<td>10.0</td>
<td>11.7</td>
<td>0.9</td>
<td>8.9</td>
<td>3.9</td>
<td>2.0</td>
</tr>
<tr>
<td>Glu</td>
<td>9.6</td>
<td>9.5</td>
<td>11.2</td>
<td>19.2</td>
<td>21.3</td>
<td>16.1</td>
<td>9.7</td>
</tr>
<tr>
<td>Ala</td>
<td>6.5</td>
<td>5.3</td>
<td>4.3</td>
<td>28.6</td>
<td>5.5</td>
<td>8.5</td>
<td>13.6</td>
</tr>
<tr>
<td>Thr</td>
<td>3.5</td>
<td>9.0</td>
<td>9.8</td>
<td>9.6</td>
<td>7.6</td>
<td>Thr: 12.9</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>6.0</td>
<td>4.3</td>
<td>4.3</td>
<td>9.6</td>
<td>5.5</td>
<td>8.5</td>
<td>13.6</td>
</tr>
</tbody>
</table>

The most abundant amino acids (in % of the total repeated domain composition) are indicated in bold. UniProtKB/SwissProt accession numbers of the various proteins are the following: P20930/filaggrin (FLG), Q86YZ3/hornerin (HRNR), Q5D8G2/filaggrin-2 (FLG2), Q07283/trichohyalin, Q6XPR3/repetin, Q9UBG3/cornulin and Q5QJ38/trichohyalin-like 1.

5.2. Hornerin

Human HRNR gene encodes hornerin (also known as S100A18), a protein of 2850 amino acids with a predicted molecular mass of 282 kDa and a predicted isoelectric point ≈10. The hornerin large central domain consists of six tandem repeated basic subunits (468±2 amino acids), each being divided into 3 segments A, B and C (Figure 5B) (76).

Hornerin has been first identified in the mouse where it was shown to be expressed in the epidermis and the cornified epithelia, including those of the tongue, forestomach and esophagus (77, 78). In humans, although conflicting results were first published, it is now established that HRNR gene is expressed in healthy epidermis. Hornerin has been detected in the granular and cornified layers of the epidermis in skin biopsies from different body sites, including head, trunk, leg, hands and feet (79, 80). In the stratum granulosum, hornerin is present in the cytoplasm of the upper keratinocytes at the periphery of keratohyalin granules, surrounding filaggrin molecules uniformly distributed within these aggregates (78-80). This suggests that hornerin is synthesized later than profilaggrin during keratinocyte differentiation. In agreement with this, the lowest granular keratinocytes have been shown to contain profilaggrin but not hornerin. In the stratum corneum, while filaggrin is mainly detected in the intracorneocyte fibrous matrix of only the lower corneocytes, hornerin is detected at the periphery of all corneocytes, from the lowest to the most superficial ones (80).

Human epidermal hornerin appears not to be very soluble, since it is extracted only in denaturing and/or reducing conditions. It has been detected as numerous immunoreactive bands corresponding to the full-length protein and to smaller fragments down to 45 kDa, indicating that hornerin, like filaggrin, is produced as a large precursor and then processed by proteolysis during cornification (80).
Epidermal differentiation complex

Based on its localization at the periphery of corneocytes, as highlighted by immunoelectron microscopy, and on its high content in glutamine residues (9.5%), hornerin has been suspected to be a component of CE (80). Immunoelectron microscopy analysis of purified plantar envelopes confirmed this hypothesis and revealed that hornerin was located on the internal face of the structures. The presence of hornerin in CEs has been further evidenced by western blot analysis of the purified envelopes digested with protease V8. Among the three different transglutaminases involved in CE formation, in vitro experiments have suggested that transglutaminase 3 is the isoform responsible for hornerin cross-linking. It has also been shown that hornerin and involucrin colocalize at the periphery of in vitro differentiated primary keratinocytes, although hornerin accumulates there several days after involucrin, suggesting that hornerin is cross-linked to the protein scaffold at a late step in CE formation (80). Hornerin function is probably to reinforce the envelopes and to contribute to the mechanical resistance of the stratum corneum. A role in antimicrobial defense has also been proposed for some hornerin-derived peptides (81). Additional experiments are needed to accurately assess the role of hornerin in the epidermal barrier functions.

Immunofluorescence experiments have revealed a decreased level of hornerin in the cornified layer of involved and uninvolved epidermis of AD patients (n = 5). A reduced expression of hornerin has also been underlined by western blot analysis of epidermal extracts of the same patients (80). Interestingly, a SNP located 6.5 kb downstream of the hornerin-encoding gene has been reported as a susceptibility variant in AD (82). Whether this polymorphism could modify the transcription or stability of HRNR mRNA has not been tested yet. Although these results need to be confirmed using a large cohort of patients and controls, they suggest that abnormalities in hornerin expression could contribute to alterations in the CE properties and therefore to the epidermal barrier defects associated with AD. This hypothesis deserves to be experimentally tested.

In a study published in 2006, HRNR was identified as the candidate gene involved in a case of acute myeloid leukemia with t(1;2) (q21;q37) chromosome translocation. Inappropiate expression of hornerin was suspected to play a role in the development of human malignant diseases (83).

5.3. Filaggrin-2

The human FLG2 gene encodes filaggrin-2, a protein of 2391 amino acids (with a predicted molecular mass of 248 kDa and a predicted pl of 8.45). Filaggrin-2 large central domain contains two types of tandem repeats, all of them being 75-77 amino acids long (Figure 5B). The nine A-type repeats are highly homologous to hornerin subunits (50-77% identity), while the 14 B-type repeats are closer to filaggrin monomers (28-39% identity) (84). FLG2 is mainly expressed in skin from different body sites (44, 84). Similarly to FLG, FLG2 gene expression is probably regulated at the transcriptional level since the level of the corresponding mRNA is dramatically (800 fold) increased in granular compared to basal keratinocytes (44). In agreement, FLG2 expression at the mRNA level is highly increased (~200 fold) in response to calcium-induced keratinocyte differentiation after four days of culture, as well as FLG and HRNR (84). Filaggrin-2 is mainly detected in the granular and lower cornified layers of the epidermis (84, 85). Immunofluorescence double staining has shown that filaggrin-2 is expressed slightly later than profilaggrin during keratinocyte differentiation but nevertheless colocalizes with profilaggrin in keratohyalin granules of the upper granular keratinocytes. In the stratum corneum, filaggrin-2 and filaggrin are detected in the intracorneocyte fibrous matrix of the lower corneocytes, both proteins concomitantly disappearing in the upper cornified layer (84, 85).

Western blot analysis of epidermal protein extracts indicates that filaggrin-2 is synthesized as a large precursor with an apparent molecular weight of 250,000 in SDS-gels. This form probably accumulates in keratohyalin granules since it requires either urea or SDS to be extracted. Filaggrin-2 is then proteolytically processed. Fragments of 130, 45 and 15 kDa have been detected. They correspond to the entire B-type repeat domain, part of this domain, and to part of the N-terminus spacer, respectively (84, 85).

The colocalization of filaggrin and filaggrin-2 in the epidermis, their simultaneous disappearance in the upper stratum corneum, and their similar physicochemical properties and proteolytic processing suggest that filagrin-2 may be modified by one or several of the enzymes involved in filaggrin catabolism, e.g. PADs, caspase 14 and calpain 1. We have demonstrated that B-type repeats of filaggrin-2 are proteolyzed in vitro by calpain 1 into several peptides. We have further shown that deimination of filaggrin-2 promotes its proteolysis by calpain 1 into numerous small peptides and probably into free amino acids. Calpain 1 cleavage sites have been identified in all B-type repeats of filaggrin-2 but not in the A-type repeats, suggesting a different metabolism for the two domains (85).

The role of filaggrin-2 is still unclear. Related molecular structure and amino acid composition of filaggrin, hornerin and filaggrin-2 (Table 1) indicate that these proteins have similar or complementary functions in the epidermis. First, filaggrin-2 B-type domain is rich in histidine (15%) and glutamine (8%) two essential components of the natural moisturizing factor, and second it may be completely proteolyzed into free amino acids in the upper stratum corneum. Thus filaggrin-2 could contribute, along with filaggrin, to stratum corneum hydration and photoprotection properties. As stated above, another role of filagrin in the lower cornified layer is the aggregation of intermediate filaments to form the intracellular corneocyte matrix. This is performed through ionic interactions between positive charges of filagrin and negative charges distributed along the rod domain of keratins (86). The high percentage of basic amino acids (histidine + arginine = 23.4%) in the filaggrin-2 B-type repeat domain suggests this protein is able to associate with
intermediate filaments. In agreement with this hypothesis, a recombinant fragment of mouse filaggrin-2 repeats have been shown, in vitro, to interact with and bundle keratin filaments isolated from bovine muzzle epidermis (53). The high homology between the A-type repeats of filaggrin-2 and hornerin subunits and their high content in glutamine (14.9%) suggest they may be component of CEs. Consistently with this hypothesis, two peptides derived from filaggrin-2 spacer-1 have been identified in purified CEs (38). Finally, whether the N-terminal S100-like domain of filaggrin-2 is translocated to the nucleus remains to be tested. Interestingly, this domain contains a potential monopartite nuclear localization signal (residues 100-104).

Comparative proteomic profiling of superficial extracts of the epidermis of atopic patients has revealed that filaggrin-2 is expressed at significantly lower levels in lesional versus non lesional skin, suggesting exacerbation of barrier defects and water loss (87). Since disruption of lipid rafts in keratinocytes evokes typical features of AD, it is interesting in this context to note that FLG2 has been identified as one of the mostly downregulated genes (fold change ~17; p < 2x10^-4) in a transcription profiling analysis after cholesterol depletion (88).

5.4. Trichohyalin

Human trichohyalin (220 kDa) comprises a highly charged central domain that consists of tandem repeated sequences of 23 amino acids. Barely detected in sparse keratinocytes in the granular and cornified layers of the epidermis, and in the filiform papillae of the tongue epithelium, trichohyalin is most abundantly expressed in the inner root sheath cells and in the medulla of the hair follicle (89-91).

Like profilaggrin, trichohyalin initially accumulates as large cytoplasmic electron-dense granules termed trichohyalin granules. Upon terminal differentiation, trichohyalin is deaminated by PADS, which induces granule solubilization (92, 93). After deamination, the fate of trichohyalin appears to be site-dependent. In the medulla, it forms unorganized amorphous deposits. In the inner root sheath, it acts as an intermediate filament-associated protein (92, 94). Its central region, predicted to form a flexible single-stranded α-helical domain, associates with keratin filaments in regular arrays, to produce a rigid structure allowing the reinforcement of the inner root sheath which supports the hair shaft growth (95). More accessible to transglutaminases, the deaminated trichohyalin contributes to CE reinforcement, either by direct homooligomerization, or by cross-linking with other envelope precursors, or as an intermediate protein necessary for keratin filament binding to this structure (92, 94). Finally, it is interesting to note that unlike the three SFTPs described above, trichohyalin is not proteolyzed during keratinocyte differentiation (95, 96).

5.5. Trichohyalin-like 1

Human trichohyalin-like 1 (904 amino acids, 99 kDa predicted mass) is the most recently identified member of the SFTP family, and is not much characterized. It is predominantly expressed in the inner root sheath of the hair follicle (98). In vitro assays have shown that trichohyalin-like 1 is a substrate of transglutaminase 1, suggesting its incorporation into CEs. This is consistent with its high glutamine residue content (9.1%). Because a recombinant fragment of trichohyalin-like 1 appeared to display in vitro killing activity against E. coli, it has also been proposed that trichohyalin-like 1 contributes to the antimicrobial properties of the cornified layer (99).

5.6. Repetin

Human repetin is a protein of 784 amino acids with a predicted molecular weight of 91,000. Its central domain contains 28 repeats of 12 amino acids and is very rich in glutamine residues (23.8%) (51). Its murine ortholog is a 1130 amino acid protein with a calculated molecular mass of 130 kDa and pl of 7.7. The central segment of mouse repetin consists of 49 tandem repeats of 12 amino acids (99). Repetin is expressed at very low levels in the interfollicular epidermis, where it is detected closely associated with keratohyalin granules of granular keratinocytes and diffusely distributed in the cytoplasm of transitional cells, but is not detected in the upper cornocytes. Repetin is also detected in the inner root sheath of hair follicles and in the filiform papillae of the tongue. By immunoblot analysis of foreskin extracts, repetin is detected as a double band with an apparent molecular mass of 100 kDa, suggesting that repetin most likely undergoes unknown post-translational modifications but no proteolytic processing during epidermal differentiation (51, 99).

Some biochemical data have demonstrated that repetin is cross-linked in vivo to CE precursors, including trichohyalin, involucrin, SPRR1 and SPRR2, in the inner root sheath of mouse hair follicles (96). Cross-links between repetin and loricrin, SPRR1 and SPRR2 have also been detected in human foreskin (51). This evidences a role for repetin probably as a cross-bridging protein involved in CE formation and in hair follicle integrity. These data are corroborated by its late expression, its high glutamine residue content, and its overexpression in the epidermis of loricrin-deficient mice (39). A strong induction of epidermal repetin expression has been reported in krupell-like factor 4-null mice (100). This indicates that repetin can be strongly upregulated upon alteration of the epidermal barrier, and suggests that repetin overexpression is involved, like LCE3 and SPRR proteins, in the compensatory mechanisms allowing the organism to repair an altered epidermal barrier.

5.7. Cornulin

Human cornulin gene encodes a protein (also called SEP53) of 495 amino acids, whose central domain consists of 2 repeats of 60 amino acids. Initially described as an oesophagus-specific and cancer-associated protein (101), cornulin expression has also been found in normal scalp and foreskin epidermis. It has been detected in the granular layer of the inter-follicular epidermis and in the inner root sheath of the hair follicles (102). At the ultrastructural level, cornulin is located at the periphery of granular keratinocytes and lower corneocytes, suggesting its involvement in CE formation although its function is
Epidermal differentiation complex

still poorly understood. This involvement is supported by the fact that it is a glutamine-rich (14.1%) protein. Interestingly it may be involved in the protection from apoptosis since it attenuates deoxycholic acid-induced cell death implicated in oesophageal reflux disease. It is up-regulated by heat-shock and other stresses. When overexpressed in oral squamous carcinoma cell lines, cornulin negatively regulates cell proliferation by the induction of G1 arrest (103, 104).

Western blot analysis of protein extracts from differentiated primary keratinocytes has revealed an aberrant migration of cornulin in SDS-gels with an apparent molecular mass of 70 kDa (instead of a predicted 54 kDa), suggesting it undergoes post-translational modifications, e.g. N-glycosylation (102). This is in agreement with the in silico prediction of N-glycosylation sites. To date, there is no evidence of proteolytic processing of cornulin in vivo.

In addition, cornulin mRNA level in the skin is decreased in eczema, both in a Der-p2 induced mouse model and in AD patients (105). However, further studies are needed to test whether this contributes to the disease.

6. CONCLUSIONS

Among the 60 different genes of the EDC, many encode components of the CEs: involucrin, loricin, SPRRs, LCE proteins and some of the S100 family but also most of the SFTPs, i.e. hornerin, trichohyalin, repentin, cornulin and, to a lesser extend, filaggrin. Involucrin and loricin are the major components and are involved in the initiation of CE assembly and their maturation, respectively. The function of the other proteins is probably to reinforce the envelopes. Their specific expression depends on the tissue type and environmental stimuli. In particular many of them seem to be induced when the epidermal barrier is impaired. Coordinate expression of the EDC genes during embryonic skin development and adult epidermal differentiation suggests a genomic mechanism and common regulatory elements to control their transcription. Non-coding sequences conserved during the evolution and shown to display keratinocyte- and differentiation-specific enhancer activities have been identified throughout the locus (106). Recently, many studies have provided evidence that proteins encoded by genes of the EDC are associated with several common skin disorders. In particular, a deletion of the genes LCE3B and LCE3C has been associated with psoriasis, and loss-of-function mutations in FGL have been shown to cause ichthyosis vulgaris and to predispose to AD, and very recently an extra octapeptide repeat of SPRR3 was identified as a risk factor for eczema (107). A possible pathophysiological implication in skin diseases including ichthyosis of the other proteins encoded by the EDC genes remains to be studied.

7. ACKNOWLEDGMENTS

We would like to apologize to our colleagues for omitting some of their work in this review because of lack of space. The authors were supported in part by the CNRS, the Toulouse III University, the INSERM, the “Société de Recherche Dermatologique”, the French Society for Dermatology (“Société Française de Dermatologie”), and Pierre Fabre Dermo-Cosmétique. We acknowledge the support of the COST BM0903 action (skin barrier in atopic dermatitis, SkinBAD).

8. REFERENCES


13. N. Jonca, E. A. Leclerc, C. Caubet, M. Simon, M. Guerri and G. Serre: Corneodesmosomes and corneodesmosin: from the stratum corneum cohesion to the
Epidermal differentiation complex


Epidermal differentiation complex


Epidermal differentiation complex


Epidermal differentiation complex


83. L. Wang, Y. Y. Wang, Q. Cao, Z. Chen and S. J. Chen: Hornerin gene was involved in a case of acute myeloid leukemia transformed from myelodysplastic syndrome with t(1;2)(q21;q37). Leukemia 20, 2184-2187 (2006)


89. G. E. Rogers, H. W. Harding and I. J. Llewellyn-Smith: The origin of citrulline-containing proteins in the hair follicle and the chemical nature of trichohyalin, an intracellular precursor. Biochim Biophys Acta 495, 159-175 (1977)


Epidermal differentiation complex


Abbreviations: AD, atopic dermatitis; CE, cornified cell envelope; EDC, epidermal differentiation complex; LCE, late cornified envelope protein; SFTP, S100-fused type proteins; SNP, single nucleotide polymorphism; SPRR, small proline rich proteins; UV, ultraviolet

Key Words: Keratinocytes, Differentiation, Filaggrin, Atopic Dermatitis, Psoriasis, Epidermis, Cornified Cell Envelopes, Review

Send correspondence to: Michel Simon, UMR5165 CNRS, Toulouse, France, Tel: 335-6115-8427, Fax: 335-6149-9036, E-mail: michel.simon@udear.cnrs.fr

http://www.bioscience.org/current/vol17.htm