Journal of Cell Science 107, 693-700 (1994)
Printed in Great Britain © The Company of Biologists Limited 1994
JCS8484



# Protein composition of cornified cell envelopes of epidermal keratinocytes

**Alasdair C. Steven[1,*] and Peter M. Steinert[2]**

[1]Laboratory of Structural Biology and [2]Skin Biology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, MD 20892, USA

*Author for correspondence

## SUMMARY

**Terminally differentiated mammalian epidermal cells are lined with a 15 nm thick layer of proteins cross-linked by isodipeptide and disulfide bonds, called the cornified cell envelope (CE). A number of proteins, including involucrin, loricrin, cystatin A, filaggrin, a cysteine-rich protein (CRP) and the 'small proline-rich' proteins (SPRRs) have been reported to be components of this complex, but little information has been obtained as to their relative abundances because the acute insolubility of the CEs has precluded direct methods of analysis. To address this question, we have determined the amino acid compositions of isolated CEs, and then modelled them in terms of linear combinations of the candidate proteins. The results show that *stratum corneum* CEs have a loricrin content of 65-70% (w/w) in human, and 80-85% in mouse. In human epidermal CEs, the secondary contributors are filaggrin and CRP (each ~10%), with smaller amounts of involucrin, SPRR and cystatin A (2-5% each) also present. Mouse epidermal CEs have about the same amount of filaggrin and somewhat more SPRR, but only trace amounts of the other proteins. In marked contrast, the major constituents of the CEs of cultured keratinocytes induced to terminal differentiation in vitro are cystatin A, involucrin and CRP (each ~30%). No significant amount of loricrin was detected except in sloughed mouse cells, which represent a more advanced state of terminal differentiation than attached cells. These results demonstrate that the CEs of cultured cells are quite different from mature CEs from native epidermis, and further suggest that generation of the CE in vivo is a multi-stage process in which involucrin, SPRR, CRP and cystatin A are initially attached to the cell membrane, and are subsequently overlaid with a heavy deposition of loricrin together with some filaggrin. Our data suggest that cultured CEs approximate only the first stage(s) of this process.**

Key words: transglutaminase, cell envelope, keratinocyte, terminal differentiation, amino acid composition, apoptosis

## INTRODUCTION

One striking feature of terminal differentiation in mammalian epidermis is the deposition of a ~15 nm thick, insoluble, layer of protein on the cells' inner surface. This structure, termed the cornified cell envelope (CE), is thought both to serve a barrier function for the organism, and to help maintain the structural integrity of the epidermis (Hohl, 1990; Reichert et al., 1993). Early work established that either harsh chemical reagents (Crounse, 1966; Matoltsy and Matoltsy, 1966) or denaturants with reducing agents (Steinert and Idler, 1975; Sun and Green, 1976) are required to solubilize the proteins of *stratum corneum* tissue. Following exhaustive extraction with such solvents cell 'ghosts' remain, which retain the shape of the cell. These ghosts retain some soluble protein unless fragmented by sonication (Hohl et al., 1991a; Yaffe et al., 1993). Ultrastructurally, the resulting fragments resemble native CEs (Rice and Green, 1979; Steinert and Idler, 1979; Mehrel et al., 1990): operationally, we refer to them as isolated CEs. They constitute about 10% of the dry weight of normal human epidermal *stratum corneum* (Hohl, 1990). Their insolubility is due to the presence of Nε-(γ-glutamyl)lysine isodipeptide crosslinks formed by epidermal transglutaminases (Abernethy et al., 1977; Rice and Green, 1977), reinforced by inter-molecular disulfide bonding. Isodipeptide crosslink frequencies range from an average of one per 100 amino acid residues (bovine; Steinert and Idler, 1979) to one per 30 residues (human; Hohl et al., 1991a), sufficient to account for the remarkable insolubility of the CE, which no extraction procedures short of acid or alkaline hydrolysis are capable of disrupting (Folk, 1983).

Several proteins have been identified as potentially major constituents of the CE. Involucrin (~68 kDa; Rice and Green, 1979), cystatin A (14 kDa; Kartasova et al., 1987; Hawley-Nelson et al., 1988; Takahashi et al., 1992), the bovine protein previously called keratolinin (Lobitz and Buxman, 1982; Zettergren et al., 1984) which is the same as cystatin A (K. Wuepper, personal communication), the small proline-rich protein (SPRR) family (~14 kDa; Kartasova and van de Putte, 1988; Hohl and Backendorf, 1992; Gibbs et al., 1993, also called cornifin - Marvin et al., 1992), CRP (~16 kDa; Tezuka and Takahashi, 1987), and trichohyalin (248 kDa; Lee et al., 1993) are initially expressed as soluble proteins, which become insoluble at later stages of terminal differentiation. All have been localized by indirect immunofluorescence to the cell periphery, and can serve as substrates for epidermal transglutaminases in vitro. Moreover, a glycine-rich protein called

loricrin (about 35 kDa, mouse (Mehrel et al., 1990), or 26 kDa, human (Hohl et al., 1991a)) has been localized to the inner surface of the CE by indirect immunofluorescence (Mehrel et al., 1990; Hohl et al., 1991a) as well as by immunoelectron microscopy (Mehrel et al., 1990; Steven et al., 1990; Yoneda et al., 1992). Filaggrin, an interstitial protein of the keratin filament lattice in early *stratum corneum* cells, which is subsequently degraded to free amino acids (Scott and Harding, 1986), has also been reported to be a significant CE component (Richards et al., 1988). In addition, several other proteins, including keratins (Abernathy et al., 1977), have been proposed as candidates (see Discussion), and it has even been suggested that the CE becomes a repository for the cytoplasmic proteins of moribund keratinocytes (the 'dustbin' hypothesis; Michel et al., 1987).

Conditions for induction of terminal differentiation in cultured epidermal keratinocytes have now been established (e.g. see Watt, 1989; Yuspa et al., 1989; Hohl et al., 1991b). The primary indicators of terminal differentiation (Watt, 1984) are generally taken to be switching from expression of the keratins K5 and K14, which are typical of basal cell gene expression, to K1 and K10, and generation of a CE. Induction of involucrin expression has also been used as a marker for terminal differentiation (Rice and Green, 1979). However, it has not been established that the CEs generated in culture fully emulate native CEs (Hohl et al., 1991b; Reichert et al., 1993): for instance, in such basic respects as protein composition.

In order to clarify the mechanism of CE formation both in vivo and in cultured cells, it is essential to obtain quantitative data on the relative abundances of putative CE proteins. The cross-linked character of the CE precludes the use of quantitative SDS-PAGE for this purpose. Previously, it has been noted that the amino acid compositions of isolated *stratum corneum* CEs, which are unusually rich in Gly, Ser and Cys, bear a marked resemblance to that of loricrin (Mehrel et al., 1990), suggesting that it is a major component, and are quite different from those of involucrin and trichohyalin (which are unusually rich in Glu/Gln residues), SPRR, cystatin A or CRP. In this study, we have performed a more detailed analysis of the amino acid compositions of isolated CEs from several kinds of cells. Mathematical optimization techniques have been used to model these data, using methods similar to those employed by Kapp et al. (1990a,b) to investigate the stoichiometries of oligomeric proteins. The CEs were described as linear combinations of the various proteins thought to be present. The mass fractions that they contribute are estimated by the resulting coefficients.

## MATERIALS AND METHODS

### Preparation of CEs

Mature intact CEs were obtained as a residue from newborn or adult mouse and human foreskin epidermis following exhaustive boiling in a solution of 2% SDS, 20 mM dithiothreitol, 0.1 M Tris-HCl, pH 8.5, and 5 mM EDTA. CEs were then purified on Ficoll gradients and recovered as sonicated fragments, free of soluble keratin/filaggrin proteins, after pelleting through Ficoll (Mehrel et al., 1990). Human foreskin epidermal cells (Clonetics Corp., CA) were grown to confluency under standard conditions in low $Ca^{2+}$ for 3 days and switched to 0.6 mM $Ca^{2+}$ for 2 days (Hohl et al., 1991b). Newborn mouse epidermal cells were grown for 6-24 hours in low $Ca^{2+}$ and switched to 1.2 mM $Ca^{2+}$ for 3 days. The sloughed cells and the remaining attached cells were harvested separately (Steinert and Yuspa, 1978). CEs were then recovered and purified from all three populations of cells as described above.

### Amino acid analyses

Samples of 5-50 mg of CEs from two to four different preparations were hydrolyzed in duplicate or triplicate using standard conditions and analyzed on either a Beckman 119CL Analyzer equipped with a fluorescence detection system (on which proline could not be quantitated) or a Beckman System 6300 Analyzer. Cysteine residues were converted to their *S*-carboxymethyl derivatives prior to hydrolysis. Tryptophan was measured following hydrolysis with 4 M methanesulfonic acid. In all cases, hydrolyses were controlled and appropriate corrections were made for hydrolytic losses of acid-labile amino acids.

### Mathematical analyses

These were carried out using the MLAB program (Knott, 1979) running on an IBM PS2/30. The modelled amino acid compositions were fitted to the experimental data by least squares. For fits in which the total amount of all components was set to 100%, this constraint was applied by Lagrangian Undetermined Multipliers (Boas, 1983).

The fits are stabilized by the consideration that the equations are strongly over-constrained, i.e. in each case we have 18 data points - the number of amino acids distinguished in the hydrolysate, and usually 6 or fewer variables. In general, 18 variables are required for a perfect solution, and somewhat fewer may be expected to provide a close solution. Thus, when a satisfactory fit is achieved with only 3-6 variables, the result is highly significant, in the sense of supporting the hypothesis that proteins in question account for all, or almost all, of the CEs' composition.

The goodness-of-fit was assessed in terms of residuals, i.e. the root-mean-square excursion per amino acid between the calculated and the experimental amino acid compositions, which is the quantity that the fit minimizes; and the median absolute value excursion per amino acid. The latter measure is less dominated by the contribution of a single outlier. Another important criterion of a good fit is that the coefficients should add up to 100% or thereabouts, and that all coefficients should be non-negative (Kapp et al., 1990a), at least, to within a few per cent. A solution with sizable negative components, while mathematically optimal, does not make sense biochemically. The robustness of the solutions was investigated by performing a large number of control calculations in which certain of these proteins were systematically omitted, and other components were included.

## RESULTS

### Amino acid composition data

Table 1 gives the amino acid compositions of the various CE preparations analyzed. Recently, it has been stressed by Yaffe et al. (1993) that rigorous procedures are required to dissociate uncrosslinked proteins from CE preparations. Consistent with this, our experience has been that flotation on, and subsequent pelleting through, Ficoll gradients, followed by fragmentation by sonication (Mehrel et al., 1990) is essential to ensure complete removal of non-cross-linked proteins. Our data for CEs derived from cultured human foreskin keratinocytes are in quite good agreement with those of Rice and Green (1979).

Table 2 summarizes the amino acid compositions of the proteins used in these calculations. Although murine equivalents of involucrin, SPRR and CRP have not yet been reported, we have allowed for the possible existence of such proteins by including the human versions of these proteins in modelling the mouse CE data.

**Table 1. Amino acid compositions (%) of cornified cell envelopes**

| Amino acid | CEs of mouse keratinocytes: Epidermis, Newborn | Adult | Cultured cells, Attached | Sloughed | CEs of human keratinocytes: Foreskin epidermis | Cultured cells |
|---|---|---|---|---|---|---|
| Asx | 0.5 | 0.5 | 6.4 | 5.1 | 1.8 | 6.2 |
| Thr | 1.4 | 1.6 | 5.9 | 5.3 | 2.2 | 6.7 |
| Ser | 23.4 | 23.8 | 7.5 | 10.1 | 20.2 | 7.2 |
| Glx | 4.1 | 4.6 | 21.4 | 16.8 | 9.2 | 22.3 |
| Pro | 3.7 | 4.2 | 5.7 | 5.2 | 0.0* | 5.5 |
| Gly | 45.1 | 43.1 | 9.0 | 16.0 | 35.3 | 9.1 |
| Ala | 0.8 | 1.0 | 5.5 | 5.0 | 3.6 | 6.2 |
| Val | 3.0 | 3.1 | 6.0 | 5.8 | 3.8 | 5.6 |
| Cys | 8.1 | 7.8 | 1.4 | 2.8 | 4.7 | 1.1 |
| Met | 0.3 | 0.4 | 1.1 | 0.8 | 0.3 | 1.2 |
| Ile | 0.3 | 0.3 | 2.3 | 2.9 | 2.0 | 2.8 |
| Leu | 0.7 | 1.0 | 10.5 | 8.5 | 1.6 | 11.1 |
| Phe | 0.2 | 0.3 | 2.0 | 2.1 | 2.5 | 1.9 |
| Tyr | 4.3 | 4.0 | 1.8 | 1.9 | 2.3 | 1.6 |
| Lys | 2.5 | 2.1 | 6.7 | 5.9 | 5.0 | 6.7 |
| His | 0.5 | 0.7 | 2.6 | 2.1 | 0.7 | 2.9 |
| Arg | 1.1 | 1.5 | 2.6 | 2.1 | 1.9 | 2.2 |
| Trp | 0.0* | 0.0* | 0.5 | 0.0* | 0.0* | 0.0* |

The amino acid compositions were determined in triplicate from 2-4 different batches of CEs.

*Not detected.

## Protein compositions of native epidermal and cultured CEs

The outcome of our modelling experiments is presented in Table 3. The results of some control experiments involving variations in the modelling procedure are given in Table 4. The calculated amino acid compositions are compared with the experimental data in Table 5.

### Native human CEs

Loricrin was found to account for 65-70% of the protein content of human epidermal CEs. This value proved to be very stable with respect to variations in the procedure (Table 4). Conversely, it proved impossible to obtain an even remotely satisfactory fit when loricrin was omitted (Tables 4c-h, 5). The secondary contributors are filaggrin and CRP at about 10%, followed by cystatin A, SPRR and involucrin, all in a range of 2.5-5%. The calculated filaggrin content was, as with loricrin, a very stable quantity (Table 4a-i), and inclusion of this protein was essential to obtaining a good fit (cf. Table 4a,b). The individual estimates for cystatin A and CRP were less stable than those for the other three proteins: however, their combined amount was quite stable, at 15-20% (Table 4a-i). The observed instability in their individual contributions arises from cystatin A and CRP having rather similar amino acid compositions (Table 2), which gives rise to a degeneracy in the fitting process.

The fits indicate that small but detectable amounts (a few per cent) of involucrin and SPRR are also present in native human CEs. Although the amounts involved are small, these coefficients were repeatedly obtained as positive, and were quantitatively stable through many variations in the procedure. Therefore, these fits support earlier evidence that these proteins are indeed incorporated into CEs, albeit as minor components. By the same criteria, we found no evidence for the presence of trichohyalin in detectable quantities.

### Native mouse CEs

In mouse epidermal CEs, both newborn and adult, the only proteins other than loricrin that were detected in significant amounts were filaggrin and SPRR. The filaggrin content is about the same as in native human CEs, and our estimate is consistent with the one obtained for the filaggrin content of rat

**Table 2. Amino acid compositions (%) of prospective CE components**

| Amino Acid | Mouse loricrin | Human loricrin | Human inv'crin | Human SPRR | Human Cyst.A | Mouse Cyst.A | Human CRP | Human fil'grin | Mouse fil'grin | H/M keratin | Human tr'hyalin | GenBank average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Asx | 0.0 | 0.3 | 2.8 | 0.0 | 10.3 | 11.5 | 8.3 | 7.2 | 5.2 | 8.6 | 3.3 | 9.3 |
| Thr | 1.5 | 2.2 | 1.6 | 2.4 | 7.2 | 8.3 | 5.4 | 4.4 | 0.0 | 3.6 | 0.7 | 6.0 |
| Ser | 22.3 | 22.8 | 1.6 | 2.4 | 2.1 | 1.0 | 9.8 | 24.8 | 19.8 | 13.9 | 1.4 | 7.3 |
| Glx | 3.2 | 4.4 | 45.8 | 28.0 | 14.4 | 17.7 | 12.7 | 15.1 | 21.4 | 12.7 | 45.7 | 10.4 |
| Pro | 1.7 | 2.9 | 5.7 | 31.2 | 5.2 | 5.2 | 0.0 | 0.6 | 2.8 | 0.8 | 1.4 | 5.3 |
| Gly | 55.1 | 46.8 | 6.7 | 0.0 | 8.3 | 9.4 | 18.0 | 15.1 | 15.7 | 23.8 | 1.5 | 7.1 |
| Ala | 0.4 | 1.0 | 1.5 | 0.0 | 5.2 | 7.3 | 5.3 | 6.6 | 8.9 | 3.3 | 1.5 | 7.6 |
| Val | 1.0 | 3.5 | 3.7 | 9.6 | 9.3 | 5.2 | 5.3 | 1.3 | 3.6 | 3.2 | 0.8 | 6.4 |
| Cys | 7.1 | 6.0 | 0.3 | 11.2 | 0.0 | 0.0 | 4.3 | 0.0 | 0.0 | 1.0 | 0.3 | 1.9 |
| Met | 0.0 | 0.0 | 0.9 | 0.0 | 2.1 | 1.0 | 0.0 | 0.0 | 0.0 | 1.5 | 0.1 | 2.3 |
| Ile | 0.0 | 1.6 | 0.4 | 0.0 | 4.1 | 7.3 | 3.7 | 0.6 | 0.0 | 3.7 | 0.4 | 5.4 |
| Leu | 0.0 | 0.0 | 14.6 | 1.6 | 8.3 | 4.2 | 6.8 | 0.6 | 0.4 | 6.9 | 9.9 | 9.3 |
| Phe | 0.0 | 2.9 | 0.6 | 0.0 | 2.1 | 5.2 | 2.9 | 0.6 | 0.8 | 3.6 | 2.0 | 3.9 |
| Tyr | 5.2 | 2.5 | 0.8 | 0.0 | 6.2 | 4.2 | 0.0 | 0.6 | 0.0 | 4.1 | 0.9 | 3.2 |
| Lys | 1.9 | 2.2 | 7.4 | 12.8 | 13.4 | 8.3 | 9.0 | 0.0 | 0.0 | 4.8 | 5.5 | 5.7 |
| His | 0.4 | 0.3 | 4.7 | 0.8 | 1.0 | 3.1 | 2.2 | 12.0 | 8.9 | 0.6 | 1.2 | 2.3 |
| Arg | 0.0 | 0.0 | 0.7 | 0.0 | 1.0 | 1.0 | 6.3 | 9.8 | 12.5 | 4.7 | 22.4 | 5.4 |
| Trp | 0.2 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.6 | 0.0 | 0.2 | 1.0 | 1.4 |

The amino acid compositions of mouse loricrin (Mehrel et al., 1990), human loricrin (Hohl et al., 1991a), human involucrin (Eckert and Green, 1986), human epidermal SPRR (Marvin et al., 1992), human (Kartasova and van de Putte, 1988) and mouse cystatin A (Hawley-Nelson et al., 1988), human (McKinley-Grant et al., 1989) and mouse filaggrin (Rothnagel and Steinert, 1990); and keratins, human K1 (Zhou et al., 1988) and mouse K10 (Steinert et al., 1983); and human trichohyalin (Lee et al., 1993), are based on protein sequences deduced from cDNA sequences. That of the cysteine-rich protein (Tezuka and Takahashi, 1987) is from acid hydrolysates. The GenBank average is from Lupas et al. (1991). For the proteins whose compositions are known in both systems (loricrin, cystatin A, filaggrin), the human and mouse versions differ by rather similar margins (2.21-2.60) and (1.24-1.92) in terms of the residuals used in Table 3.

**Table 3. Modelled mass fractions (%) of cell envelope proteins**

| Type of cell envelope | Loricrin | Involucrin | SPRR | Cyst.A | CRP | F'grin | Residuals | Sum |
|---|---|---|---|---|---|---|---|---|
| Man | | | | | | | | |
| (a) Foreskin epidermis | 66.1 | 2.4 | 3.0 | 4.9 | 14.2 | 10.6 | 0.69/0.53 | 100.4 |
| (b) Cultured cells | * | 30.0 | 2.0 | 26.9 | 35.8 | * | 1.42/1.03 | 94.7 |
| Mouse | | | | | | | | |
| (c) Newborn epidermis | 82.1 | * | 5.6 | * | * | 8.4 | 1.11/0.70 | 96.1 |
| (d) Adult epidermis | 77.8 | * | 6.0 | * | * | 12.2 | 1.27/0.83 | 96.0 |
| (e) Cultured cells (attached) | * | 25.7 | 5.9 | 21.6 | 39.7 | * | 1.36/1.09 | 92.9 |
| (f) Cultured cells (sloughed) | 10.7 | 14.5 | 7.5 | 13.3 | 47.9 | * | 1.12/0.96 | 93.9 |

The first residual is the root-mean-square discrepancy; the second residual is the median discrepancy (absolute value) in units of %/amino acid residue.
*Protein deemed not to be present in significant amounts, i.e. less than about 1-2%.

**Table 4. Modelled mass fractions (%) of cell envelope proteins - controls**

| CEs | Fit | Loric. | Invol. | SPRR | Cyst.A | CRP | F'grin | Keratin | GenBank | Residuals | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Human foreskin epidermis | (a) | 67.1 | 4.4 | x | 8.0 | 9.7 | 11.1 | x | x | 0.69/0.52 | 100.4 |
| | (b) | 55.8 | 3.5 | x | −10.9 | 51.1 | x | x | x | 1.11/0.89 | 99.5 |
| | (c) | 69.1 | 4.3 | x | 13.9 | x | 13.2 | x | x | 0.72/0.57 | 100.5 |
| | (d) | 65.3 | 5.1 | x | x | 20.0 | 8.9 | x | x | 0.72/0.57 | 99.3 |
| | (e) | 65.6 | x | x | 11.9 | 11.7 | 12.5 | x | x | 0.76/0.58 | 101.7 |
| | (f) | x | −9.1 | x | −60.0 | 162.8 | 11.7 | x | x | 4.29/3.33 | 105.4 |
| | (g) | x | −6.5 | x | −49.3 | 26.6 | −6.9 | 137.0 | x | 2.96/2.29 | 100.8 |
| | (h) | 68.5 | 4.4 | x | 9.3 | 12.9 | 11.7 | −6.0 | x | 0.68/0.51 | 100.7 |
| | (i) | 66.9 | 4.2 | x | 7.9 | 10.5 | 10.8 | x | 0.3 | 0.69/0.51 | 100.6 |
| Human cultured cells | (j) | x | 31.1 | x | 28.3 | 34.8 | x | x | x | 1.42/1.04 | 94.3 |
| | (k) | −1.8 | 30.7 | x | 26.5 | 39.0 | x | x | x | 1.25/1.00 | 94.4 |
| | (l) | 10.0 | 33.4 | x | 48.4 | x | x | x | x | 1.68/1.29 | 91.8 |
| | (m) | −8.2 | 33.5 | x | x | 65.5 | x | x | x | 1.64/1.34 | 90.7 |
| | (n) | −11.5 | x | x | 47.9 | 64.8 | x | x | x | 2.84/2.06 | 101.2 |
| Newborn mouse epidermis | (o) | 81.1 | x | x | x | x | 12.4 | x | x | 1.23/0.92 | 93.5 |
| | | 82.9 | x | x | x | x | 17.1 | x | x | 1.38/1.05 | 100.0* |
| | (p) | 85.1 | x | 8.2 | x | x | x | x | x | 1.23/0.76 | 93.3 |
| | (q) | 82.2 | x | 5.6 | x | x | 8.7 | −5.1 | x | 1.11/0.71 | 91.4 |
| | (r) | x | −32.8 | 33.3 | −95.0 | −12.0 | x | 215.5 | x | 3.48/2.63 | 108.9 |
| | (s) | 85.8 | x | x | x | x | x | x | x | 1.52/1.07 | 85.8 |
| | | 100.0 | x | x | x | x | x | x | x | 2.52/1.24 | 100.0* |

The first residual is the root-mean-square discrepancy; the second residual is the median discrepancy (absolute value) in units of %/amino acid residue. x, Protein omitted from fit. In (o) and (s), the second fits (*) were constrained to add up to 100%.

epidermal CEs (~10%) by an entirely different approach (Richards et al., 1988). The amount of SPRR appears to be somewhat greater than in human CEs. No significant amounts of involucrin, CRP or cystatin A were detected. On the other hand, there is a larger amount of loricrin than in native human CEs, and the fits (Tables 3c,d; 4o-t) consistently indicate that the proteins under consideration account for 90-95% of the total composition, implying that the remaining 5-10% is contributed by other, as yet unknown, component(s). This discrepancy may reflect, in part, the use of some human proteins (involucrin, SPRR and CRP) in modeling the mouse data.

## Cultured keratinocyte CEs

Our analyses of three kinds of cultured CEs indicates that their protein compositions are markedly different from those of native CEs, as is already suggested by their respective amino acid compositions (Table 1). The human cultured CEs were estimated to contain approximately equal amounts (~30% each) of involucrin, cystatin A and CRP, and essentially no loricrin. Unconstrained fits initially suggested that there is also 5-15% of filaggrin in these CEs (data not shown): however, since measurements of the amount of filaggrin mRNA in these cells indicated levels that are about 100-fold lower than in the epidermis (Markova et al., 1993), the fits were constrained accordingly (Table 3). The origins of this apparently exaggerated estimate of filaggrin in cultured CEs are discussed further below.

In estimated protein composition, the cultured mouse keratinocyte (attached) CEs closely resemble the corresponding human material. This observation implies that there are indeed mouse CE proteins whose amino acid compositions resemble those of human involucrin and CRP. These proteins are only expressed in low amounts in native epidermis, but are major constituents of cultured CEs. SPRR is detected as a minor but significant component of all three kinds of cultured cell CEs. In both human and mouse, it is the only protein that makes up about the same fraction of the cultured CEs as it does of native epidermal CEs.

No detectable content of loricrin was present in attached mouse cell CEs, but for sloughed cell CEs a substantial component of loricrin (>10%) was detected. The data also suggest that CRP expression is enhanced relative to cystatin A in the attached → sloughed transition, but in view of the observed instability in their relative amounts (see above), this

**Table 5. Comparison between experimental and calculated amino acid compositions of CEs**

| | Human foreskin epidermis | | | Cultured human cells | | Newborn mouse epidermis | | | Adult mouse epidermis | | Cultured mouse cells (attached) | | Cultured mouse cells (sloughed) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Residue | Exptl | Calc. | (K) | Exptl | Calc. | Exptl | Calc. | (K) | Exptl | Calc. | Exptl | Calc. | Exptl | Calc. |
| AsX | 1.8 | 2.7 | 8.2 | 6.2 | 6.3 | 0.5 | 0.4 | 5.9 | 0.5 | 0.6 | 6.4 | 6.3 | 5.1 | 5.8 |
| Thr | 2.2 | 3.2 | 2.4 | 6.7 | 4.3 | 1.4 | 1.4 | –0.5 | 1.4 | 1.3 | 5.9 | 4.4 | 5.3 | 4.2 |
| Ser | 20.2 | 19.3 | 18.8 | 7.2 | 4.5 | 23.4 | 20.1 | 28.2 | 23.8 | 19.9 | 7.5 | 4.6 | 10.1 | 7.6 |
| GlX | 9.2 | 9.0 | 9.6 | 22.3 | 22.6 | 4.1 | 6.0 | 3.3 | 4.1 | 6.8 | 21.4 | 22.2 | 16.8 | 17.5 |
| Pro | 2.9 | 3.3 | –2.0 | 5.5 | 4.3 | 3.7 | 3.4 | –4.7 | 3.7 | 3.5 | 5.7 | 4.9 | 5.2 | 4.3 |
| Gly | 35.3 | 35.7 | 32.0 | 9.1 | 10.8 | 45.1 | 46.6 | 38.0 | 45.1 | 44.8 | 9.0 | 11.0 | 16.0 | 16.9 |
| Ala | 3.6 | 2.4 | 2.8 | 6.2 | 3.6 | 0.8 | 1.1 | –0.9 | 0.8 | 1.4 | 5.5 | 4.0 | 5.0 | 3.7 |
| Val | 3.8 | 4.0 | 0.8 | 5.6 | 5.8 | 3.0 | 1.7 | 3.1 | 3.0 | 1.8 | 6.0 | 4.8 | 5.8 | 4.7 |
| Cys | 4.7 | 4.9 | 2.5 | 1.1 | 1.9 | 8.1 | 6.5 | 5.2 | 8.1 | 6.2 | 1.4 | 2.4 | 2.8 | 3.7 |
| Met | 0.3 | 0.1 | 1.0 | 1.2 | 0.9 | 0.3 | 0.0 | 2.3 | 0.3 | 0.0 | 1.1 | 2.9 | 0.8 | 0.3 |
| Ile | 2.0 | 1.9 | 4.0 | 2.8 | 2.5 | 0.3 | 0.0 | 0.5 | 0.3 | 0.0 | 2.3 | 3.1 | 2.9 | 2.8 |
| Leu | 1.6 | 1.9 | 6.1 | 11.1 | 9.2 | 0.7 | 0.1 | 5.5 | 0.7 | 0.1 | 10.5 | 7.6 | 8.5 | 6.1 |
| Phe | 2.5 | 2.5 | 4.6 | 1.9 | 1.7 | 0.2 | 0.1 | 2.4 | 0.2 | 0.1 | 2.0 | 2.3 | 2.1 | 2.1 |
| Tyr | 2.3 | 2.0 | 2.4 | 1.5 | 1.8 | 4.3 | 4.3 | 4.7 | 4.3 | 4.0 | 1.8 | 1.0 | 1.9 | 1.2 |
| Lys | 5.0 | 4.0 | 1.9 | 6.7 | 9.4 | 2.5 | 2.3 | 3.1 | 2.5 | 2.2 | 6.7 | 8.1 | 5.9 | 7.7 |
| His | 0.7 | 2.0 | –0.2 | 2.9 | 2.6 | 0.5 | 1.1 | –3.6 | 0.5 | 1.4 | 2.6 | 0.5 | 2.1 | 2.3 |
| Arg | 1.9 | 2.0 | 6.8 | 2.2 | 2.7 | 1.1 | 1.1 | 8.1 | 1.1 | 1.5 | 2.6 | 2.9 | 2.1 | 3.2 |
| Trp | 0.0 | 0.3 | 0.2 | 0.0 | 0.1 | 0.0 | 0.2 | –0.3 | 0.0 | 0.2 | 0.5 | 0.1 | 0.4 | 0.1 |

The calculated compositions correspond to the fractions of the various proteins involved, according to Table 3. For both newborn and adult mouse epidermis, the 3-component fits (loricrin, SPRR, filaggrin) from Table 3 are given. The third column (K) for human and newborn mouse epidermal data correspond to the fits given in Table 4(g) and (r), and illustrate the inadequacies of the solutions obtained when loricrin is omitted from the analysis, and keratins K1 and K10 are substituted as the glycine-rich component.

apparent alteration is not necessarily significant. Thus, the difference between attached and sloughed mouse keratinocytes in culture correlates with the incorporation of loricrin into their CEs, together with reduced levels of involucrin and cystatin A/CRP, as these cells detach and slough into the medium.

### Numerical control experiments

The analysis of the human CEs and the newborn mouse epidermal CEs has been pursued in greater detail (Table 4). Addressing the possibility that the conspicuously high glycine contents of the epidermal CEs may be attributed, at least in part, to keratins K1 and K10, which also have unusually high glycine contents (Steinert et al., 1983, 1985; Yuspa et al., 1989), these molecules have been included in the fits (Table 4g,h and q,r). When keratin is substituted for loricrin, striking examples of misfits are obtained, with unacceptably high residuals, and unrealistic coefficients (Table 4g,r). The reason for this is that the glycine content of these keratins, at ~25%, is much lower than that of the CEs: to generate enough glycine, the least-squares solution conjures up a keratin content that exceeds the maximum possible amount, and offsets this ‘overshoot’ with an (equally unphysical) large negative component of cystatin A. Even so, the resulting solution represents a poor fit, with prohibitive mismatches in Asx, Pro, Leu and Arg (Table 5). When keratin is included in addition to loricrin, the least-squares residual is not significantly improved (cf. Tables 3a, 4a,h), i.e. no benefit is obtained from the availability of this additional free parameter. Moreover, the estimated contents of the other proteins are largely unaffected by the inclusion of keratin (provided that loricrin is also included). We conclude that there is no reason to suppose that keratins K1 and K10 become incorporated into the cross-linked CE in significant amounts, although they are expressed abundantly in terminally differentiated keratinocytes.

To address the possibility that many different cytoplasmic and/or organellar proteins may become attached to the cornifying envelope by the transglutaminases (Michel et al., 1987), the average composition of GenBank proteins (Lupas et al., 1991) was also included in the fit. However, the residuals and the estimated amounts of the other proteins were largely unaffected (e.g. Table 4i).

For the mouse epidermal CE data, which are dominated by their loricrin content, the model was progressively simplified until it consisted of loricrin alone (Table 4o,s). In fact, the least-squares residual does not deteriorate much, reflecting the fact that the amino acid composition of loricrin itself is already a good first approximation to that of entire CEs, and the dynamics of the fitting process are dominated by the requirement of reducing the glycine content in the fit, from that of loricrin, closer to that of the CE. However, the sum of the components, which should be 100%, drops to 86% on eliminating all non-loricrin components. Accordingly, the latter fits were repeated, imposing the constraint that the components should add up to 100%. These results more clearly document the reduction in the residual that accompanies inclusion of more components. Moreover, they corroborate the conclusion that loricrin accounts for >80% of the mass of native mouse CEs. On the other hand, the tendency for the unconstrained fit to yield a solution whose components add up to less than 100% most likely reflects the presence, in small amounts, of other yet-to-be-identified protein(s).

## DISCUSSION

The generation of a covalently cross-linked amalgam of proteins through the action of transglutaminases inside terminally differentiating cells is not restricted to the epidermis, but is fairly widespread among other terminally differentiating epithelia, such as the palate, esophagus, etc. It is also observed in apoptosis or ‘programmed cell death’ of other, non-epithelial, tissues (Folk, 1983; Harsfalvi et al., 1991; Fesus et al.,

1991; Tarcsa et al., 1992). However, a distinctive feature of its manifestation in epithelia is the confinement of this complex to the cells' inner surface, i.e. formation of a CE. The covalently cross-linked character of the CE, which is the basis of its physical and chemical resilience, has thwarted previous attempts to quantitate its protein composition.

### Scope of mathematical modelling of amino acid composition data

Provided that certain conditions are met (see below), the technique of mathematical modelling on the basis of amino acid composition data provides a powerful analytical tool. The technique can be used either at the supramolecular level to probe the subunit composition of oligomeric complexes (Kapp et al., 1990a,b), or at the submolecular level to locate a peptide fragment within an intact polypeptide chain (Conway and Steven, 1991). It can also be used, as here, to analyze insoluble protein complexes. The conditions necessary for successful analyses of the latter kind are: (1) the availability of accurate amino acid composition data; (2) the complex should have relatively few components; (3) preferably, its constituent proteins should have amino acid compositions that are individually distinctive. However, the latter requirement is not absolute, and we note that Kapp et al. (1990b) obtained quite good results in their analysis of the stoichiometries of several oligomeric proteins whose subunits have rather uniform amino acid compositions. In the present study, we have encountered some degeneracy between cystatin A and CRP, but not between involucrin and trichohyalin, although the overall protein compositions of the latter two proteins are also rather alike: presumably the high Arg content of trichohyalin (>22%; Table 2) is sufficient to resolve this potential degeneracy. It is also desirable - although not essential - that all major constituents should be known: the implications of the omission of a significant component from the fit are discussed further below. All of the above conditions appear to be met in the present context.

(1) The amino acid compositions of all but one of the proteins were determined from cDNA sequences, and that of the sole exception, CRP, from amino acid analyses of the purified protein.

(2) When the number of unknowns (the various proteins involved) is much less than the number of data points (the 18 distinct species of amino acid distinguished in the hydrolysates), the equations to be solved are heavily over-constrained, which exerts a stabilizing effect on the fit, in the presence of experimental error. In this study, the inclusion of only six or fewer proteins produced excellent fits, whose residual error was close to that typical of the reproducibility of the CE amino acid composition data (on average, ~0.5% per residue).

(3) When two or more proteins with similar amino acid compositions are included in a fit, a certain ambiguity arises as to their relative amounts, although their combined amount remains a fairly stable quantity. The severity of such problems will, in general, depend on how closely the proteins resemble each other, how many proteins in all are involved, the level of experimental uncertainties in the data, and on the specimen's amino acid composition per se. In our analyses, the only occurrence that we encountered of such an ambiguity was between cystatin A and CRP in native CEs. The cultured cell data, on the other hand, were able to distinguish relatively stably between these two proteins, presumably on the basis of the substantial difference between their respective Gly and Ser contents (Table 2). With native CEs, the Gly and Ser components in the fit were always dominated by the loricrin contribution, whereas in cultured CEs, which are essentially devoid of loricrin, the respective contributions of cystatin A and CRP could be distinguished on this basis.

The stability of these fits, and the success of this analytical approach as applied to the problem at hand, was assisted by the unusual amino acid compositions of the proteins involved, which, for the most part, are markedly divergent from the GenBank average (cf. Table 2).

### Have any major CE components been overlooked?

One feature of our fits to the human CE data that inspires confidence is the fact that its coefficients sum up to almost exactly 100% without being specifically constrained to do so. This suggests that the proteins included account for most, if not all, of the mass of these CEs. The coefficients of the mouse CE fits differ somewhat in that they consistently add up to ~95%, and slightly less, for the cultured CEs (Tables 3, 4). In principle, this shortfall might arise from differences between the human proteins used in these fits (involucrin, CRP) and their supposed mouse counterparts. However, the success of these human proteins in fitting the cultured mouse keratinocyte CE data argues against this explanation, and points to the alternative that the remainder is contributed by other, as yet unidentified, protein(s).

Our analysis has included all CE candidate proteins for which amino acid composition data are currently available. Only a few other proteins have so far been proposed as CE components: a 195 kDa protein (Simon and Green, 1985; Ma and Sun, 1986); a group of small, soluble, basic proteins ('pancornulins') reported by Baden et al. (1987); and the proteins mentioned in the review of Reichert et al. (1993). While our analysis does not exclude the possibility of these or other proteins also being present in CEs, it does indicate that either: (i) they are minor components, amounting cumulatively to no more than a few per cent in man, and about 5% in mouse; or (ii) they should have amino acid compositions that closely resemble those of established CE constituents - in which case, they will share the mass fractions currently assigned to those proteins. In this context, the pancornulins are thought to resemble the SPRR proteins (Baden and Kvedar, 1993), and therefore may contribute part of the CE fractions presently assigned to SPRR. More generally, we note that CE proteins tend to have unusual and distinctive amino acid compositions (Table 2), very much different from the GenBank average. This property implies that, a priori, the occurrence of other proteins with the same compositions, other than additional members of the same family, is relatively unlikely.

It may be that other proteins are associated with the CE in vivo by disulfide or other non-covalent bonds (Yaffe et al., 1993), and could thereby contribute to its physical and chemical properties. However, such proteins would be extracted by our isolation procedure, and thus would not be detected in our compositional analysis of the covalently cross-linked CE.

Our analyses are based on the assumption that CE proteins are incorporated intact, not as proteolytically degraded fragments. While we have not been able to justify this assump-

tion rigorously, the excellent fits obtained with just a few (intact) proteins suggest that this assumption is largely sound.

The one anomaly is the seemingly excessive amounts of filaggrin attributed to cultured CEs in both human (10-15%) and mouse (5-7%) by the compositional analyses. This may well reflect the relatively enhanced expression of transglutaminase 1 in culture (Reichert et al., 1993), so that, although filaggrin is expressed at lower levels than in the epidermis, a much larger fraction of it enters the CE. Alternatively, we note that the fitting procedure assigns these fractions of filaggrin primarily in order to reduce the Lys content and to increase the Ser content of the modelled composition, so as to better match the experimental data (Table 5). (Filaggrin has no Lys, and 25% (man) or 20% (mouse) of Ser.) Thus there may be another, as yet unidentified, protein that is rich in Ser and poor in Lys, for which filaggrin is aberrantly substituting.

### The 'dustbin' (trash-can) hypothesis

Noting that the disappearance of organelles from the cytoplasm of epidermal keratinocytes late in the terminal differentiation pathway correlates approximately with the appearance of a morphologically defined CE, Michel and coworkers (1987) have put forward the hypothesis that the proteins of these organelles become recycled as CE components. Qualitatively, we do not reject the idea that small amounts of some such proteins may enter the CE, but quantitatively our analysis implies that they do not amount to a substantial mass fraction of the cross-linked CE. The fact that these proteins have to be disposed of somehow, upon breakdown of the organelles, does not necessarily mean that they enter the CE. Alternative possible fates are proteolytic digestion (like the majority of filaggrin), or that they somehow become incorporated (or trapped) in the keratin filament-matrix that eventually occupies the interior of stratum corneum cells. Keratins K1 and K10 provide a case in point. They are expressed in large amounts in the epidermis, and have a substantial content of lysine (and thus, are potential substrates for transglutaminase) but, as our analysis shows, they do not become incorporated into the CE in significant amounts. Abnormal keratinization in the epidermis, such as occurs in psoriasis, may represent an exception (Hanigan and Goldsmith, 1978).

### CE synthesis in vivo and in vitro

Perhaps the most striking result of our analysis is that it shows the molecular compositions of the CEs of cultured keratinocytes to be very different from those of native CEs. This observation may not apply to all systems in which terminal differentiation of keratinocytes is induced in vitro. It does, however, make the point that generation of a covalently cross-linked CE does not necessarily mean that the native CE has been reproduced. To address this question for a given cell culture system, compositional analysis based on amino acid content as done in this study affords a simple and sensitive assay. The same approach may also be useful in investigating the molecular compositions of the CEs or cross-linked matrices of other apoptotic cells, where the functions and fates of these structures may differ from those of the epidermis (e.g. see Tarcsa et al., 1992).

Our observations imply that the generation of CEs in vivo is a multi-stage process. Initially, involucrin, cystatin A, CRP, SPRR and, possibly, other minor components are attached to the cell envelope, either concurrently or in some unknown sequence; subsequently, they are overlaid with a heavy deposition of loricrin, together with some filaggrin. The cultured cells analyzed here appear to emulate only the first stage(s) of this process. In support of this hypothesis, we note that cystatin A, CRP and involucrin are expressed in approximately the same relative amounts, both in culture and in vivo. The apparent dearth of the murine counterparts of these proteins in native mouse CEs may indicate that, in mouse, the first phase is shorter and/or the second, loricrin-dominated, phase is longer than in human epidermis. The fact that we detect incipient loricrin incorporation into cultured mouse cell CEs only for sloughed cells, which represent a more advanced state of terminal differentiation than attached cells, further supports this scenario.

This account of CE biosynthesis in native epidermis and in cultured cells is in general agreement with that of Reichert et al. (1993), who also drew attention to the difference between the mechanical properties of fragile and robust CEs (Michel et al., 1988), which they equated with immature and mature stages of CE assembly. In this context, it has long been known that involucrin is expressed quite early in the epidermal pathway of terminal differentiation (Rice and Green, 1979), compared with loricrin or filaggrin, which are not detected prior to the granular layer stage (Mehrel et al., 1990; Steven et al., 1990). However, to clarify our respective views of the order of events (see Reichert et al., 1993), our data imply that cystatin A (keratolinin) and CRP are probably incorporated together with involucrin early in CE cross-linking (see above), and certainly earlier than loricrin, whose time-course of incorporation appears to be matched only by that of filaggrin.

We greatly appreciate insightful conversations with Dr Kirk Wuepper and his clarification of the cystatin A/keratolinin story. We also thank Drs Benes Trus, James Conway, Joseph Carroll, Stuart Yuspa, Tonya Kartasova and Nelli Markova, and Mr Mark Vivino, for valuable input of various kinds.

## REFERENCES

**Abernethy, J. L., Hill, R. L. and Goldsmith, L. A.** (1977). Epsilon-(gamma-glutamyl)lysine cross-links in human *stratum corneum*. *J. Biol. Chem.* **252**, 1837-1839.

**Baden, H. P., Kubilus, J., Philips, S. B., Kvedar, J. C. and Tahan, S. R.** (1987). A new class of soluble basic protein precursors of the cornified envelope of mammalian epidermis. *Biochim. Biophys. Acta* **925**, 63-73.

**Baden, H. P. and Kvedar, J. C.** (1993). Epithelial cornified envelope precursors are in the hair follicle and nail. *J. Invest. Dermatol.* **101**, 72S-74S.

**Boas, M. L.** (1983). *Mathematical Methods in the Physical Sciences*, 2nd edn. Wiley, New York.

**Conway, J. F. and Steven, A. C.** (1991). Localization of the chymotryptic sheet fragment in the primary sequence of bacteriophage T4 major capsid protein, gp23. *J. Struct. Biol.* **106**, 234-236.

**Crounse, R. G.** (1966). Epidermal keratin and prekeratin. *Nature* **211**, 1301-1302.

**Eckert, R. L. and Green, H.** (1986). Structure and evolution of the human involucrin gene. *Cell* **46**, 583-589.

**Fesus, L., Davies, P. J. A. and Piacentini, M.** (1991). Apoptosis: molecular mechanisms in programmed cell death. *Eur. J. Cell Biol.* **56**, 170-177.

**Folk, J. E.** (1983). Mechanism and basis for specificity of transglutaminase-catalyzed e(g-glutamyl)lysine bond formation. *Advan. Enzymol.* **54**, 1-54.

**Gibbs, S., Fijneman, R., Wiegant, J., Geurts van Kessel, A., van de Putte, P. and Backendorf, C.** (1993). Molecular characterization and evolution of the SPRR family of keratinocyte differentiation markers encoding small proline-rich proteins. *Genomics* **16**, 630-637.

**Hanigan, H. and Goldsmith, L. A.** (1978). Endogenous substrates for epidermal transglutaminase. *Biochim. Biophys. Acta* **522**, 589-601.

**Harsfalvi, J., Fesus, L., Tarcsa, E., Laczko, J. and Loewy, A. G.** (1991). The presence of a covalently cross-linked matrix in human platelets. *Biochim. Biophys. Acta* **1073**, 268-274.

**Hawley-Nelson, P., Roop, D. R., Cheng, C. K., Krieg, T. M. and Yuspa, S. H.** (1988). Molecular cloning of mouse epidermal cystatin A and detection of regulated expression in differentiation and tumorigenesis. *Mol. Carcinogen.* **1**, 202-211.

**Hohl, D.** (1990). Cornified cell envelope. *Dermatologica* **180**, 201-211.

**Hohl, D., Mehrel, T., Lichti, U., Turner, M. L., Roop, D. R. and Steinert, P. M.** (1991a). Characterization of human loricrin. Structure and function of a new class of epidermal cell envelope proteins. *J. Biol. Chem.* **266**, 6626-6636.

**Hohl, D., Lichti, U., Breitkreutz, D., Steinert, P. M. and Roop, D. R.** (1991b). Transcription of the human loricrin gene in vitro is induced by calcium and cell density and suppressed by retinoic acid. *J. Invest. Dermatol.* **96**, 414-418.

**Hohl, D. and Backendorf, C.** (1992). A common origin for cornified envelope proteins. *Nature Genetics* **2**, 91.

**Kapp, O. H., Qabar, A. N. and Vinogradov, S. N.** (1990a). Calculation of subunit stoichiometry of large, multisubunit proteins from amino acid compositions. *Anal. Biochem.* **184**, 74-82.

**Kapp, O. H., Qabar, A. N., Bonner, M. C., Stern, M. S., Walz, D. A., Schmuck, M., Pilz, I., Wall, J. S. and Vinogradov, S. N.** (1990b). Quaternary structure of the giant extracellular hemoglobin of the leech Macrobdella decora. *J. Mol. Biol.* **213**, 141-158.

**Kartasova, T., Cormelissen, B. J. C., Belt, P. and van de Putte, P.** (1987). The effects of UV, 4NQO and TPA on gene expression in cultured human keratinocytes. *Nucl. Acids Res.* **15**, 5945-5962.

**Kartasova, T. and van de Putte, P.** (1988). Isolation, characterization, and UV-stimulated expression of two families of genes encoding polypeptides of related structure in human epidermal keratinocytes. *Mol. Cell. Biol.* **8**, 2195-2203.

**Knott, G. D.** (1979). MLAB - A mathematical modeling tool. *Comp. Progs. Biomed*. **10**, 271-280.

**Lee, S-C., Kim, I.-G., Marekov, L., O'Keefe, E. J., Parry, D. A. D. and Steinert, P. M.** (1993). The structure of human trichohyalin. *J. Biol. Chem.* **268**, 12164-12176.

**Lobitz, C. J. and Buxman, M. M.** (1982). Characterization and localization of bovine epidermal transglutaminase substrate. *J. Invest. Dermatol*. **78**, 150-154.

**Lupas, A., Van Dyke, M. and Stock, J.** (1991). Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164.

**Ma, A. S.-P. and Sun, T.-T.** (1986). Differentiation-dependent changes in the solubility of a 195-kD protein in human epidermal keratinocytes. *J. Cell Biol.* **103**, 41-48.

**Markova, N. G., Marekov, L. N., Chipev, C. C., Gan, S.-Q., Idler, W. W. and Steinert, P. M.** (1993). Profilaggrin is a major epidermal calcium-binding protein. *Mol. Cell Biol*. **13**, 613-625.

**Marvin, K. W., George, M. D., Fujimoto, W., Saunders, N. A., Bernacki, S. H. and Jetten, A. M.** (1992). Cornifin, a cross-linked envelope precursor in keratinocytes that is down-regulated by retinoids. *Proc. Nat. Acad. Sci. USA* **89**, 11026-11030.

**Matoltsy, A. G. and Matoltsy, M. N.** (1966). The membrane protein of horny cells. *J. Invest. Dermatol.* **46**, 127-129.

**McKinley-Grant, L. J., Idler, W. W., Bernstein, I. A., Parry, D. A., Cannizzaro, L., Croce, C. M., Huebner, K., Lessin, S. R. and Steinert, P. M.** (1989). Characterization of a cDNA clone encoding human filaggrin and localization of the gene to chromosome region 1q21. *Proc. Nat. Acad. Sci. USA* **86**, 4848-4852.

**Mehrel, T., Hohl, D., Rothnagel, J. A., Longley, M. A., Bundman, D., Cheng, C., Lichti, U., Bisher, M. E., Steven, A. C. and Steinert, P. M.** (1990). Identification of a major keratinocyte cell envelope protein, loricrin. *Cell* **61**, 1103-1112.

**Michel, S., Schmidt, R., Robinson, S. M., Shroot, B. and Reichert, U.** (1987). Identification and subcellular distribution of cornified envelope precursor proteins in the transformed human keratinocyte line SV-K14. *J. Invest. Dermatol*. **88**, 301-305.

**Michel, S., Schmidt, R., Shroot, B. and Reichert, U.** (1988). Morphological and biochemical characterization of the cornified envelopes from human epidermal keratinocytes of different origin. *J. Invest. Dermatol.* **91**, 11-15.

**Reichert, U., Michel, S. and Schmidt, R.** (1993). The cornified envelope: A key structure of terminally differentiating keratinocytes. In *Molecular Biology of the Skin: The Keratinocyte* (ed. M. Darmon and M. Blumenberg), pp. 107-150. Academic Press, San Diego.

**Rice, R. H. and Green, H.** (1977). The cornified envelope of terminally differentiated human epidermal keratinocytes consists of cross-linked protein. *Cell* **11**, 417-422.

**Rice, R. H. and Green, H.** (1979). Presence in human epidermal cells of a soluble protein precursor of the cross-linked envelope: Activation of the cross-linking by calcium ions. *Cell* **18**, 681-694.

**Richards, S., Scott, I. R., Harding, C. R., Liddell, J. E., Powell, G. M. and Crtis, C. G.** (1988). Evidence for filaggrin as a component of the cell envelope of the newborn rat. *Biochem. J.* **253**, 153-160.

**Rothnagel, J. A. and Steinert, P. M.** (1990). The structure of the gene for mouse filaggrin and a comparison of the repeating units. *J. Biol. Chem.* **265**, 1862-1865.

**Scott, I. R. and Harding, C. R.** (1993). Filaggrin breakdown to water binding compounds during development of the rat *stratum corneum* is controlled by the water activity of the environment. *Dev. Biol.* **115**, 84-92.

**Simon, M. and Green, H.** (1985). Enzymatic cross-linking of involucrin and other proteins by keratinocyte particulates in vitro. *Cell* **40**, 677-683.

**Steinert, P. M. and Idler, W. W.** (1975). The polypeptide composition of bovine epidermal α-keratin. *Biochem. J.* **151**, 603-614.

**Steinert, P. M. and Idler, W. W.** (1979). Postsynthetic modifications of mammalian epidermal α-keratin. *Biochemistry* **18**, 5664-5669.

**Steinert, P. M. and Yuspa, S. H.** (1978). Biochemical evidence for keratinization by mouse epidermal cells in culture. *Science* **200**, 1491-1493.

**Steinert, P. M., Rice, R.,H., Roop, D. R., Trus, B. L. and Steven, A. C.** (1983). Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments. *Nature* **302**, 794-800.

**Steinert, P. M., Parry, D. A. D., Idler, W. W., Johnson, L. D., Steven, A. C. and Roop, D. R.** (1985). Amino acid sequences of mouse and human epidermal type II keratins of Mr 67,000 provide a systematic basis for the structural and functional diversity of the end domains of keratin intermediate filament subunits. *J. Biol. Chem*. **260**, 7142-7149.

**Steven, A. C., Bisher, M. E., Roop, D. R. and Steinert, P. M.** (1990). Biosynthetic pathways of filaggrin and loricrin - two major proteins expressed by terminally differentiated epidermal keratinocytes. *J. Struct. Biol*. **104**, 150-162.

**Sun, T. and Green, H.** (1976). Differentiation of the epidermal keratinocyte in cell culture: formation of the cornified envelope. *Cell* **9**, 511-521.

**Takahashi, M., Tezuka, T. and Katunuma, N.** (1992). Phosphorylated cystatin alpha is a natural substrate of epidermal transglutaminase for formation of skin cornified envelope. *FEBS Lett*. **308**, 79-82.

**Tarcsa, E., Kedei, N., Thomazy, V. and Fesus, L.** (1992). An involucrin-like protein in hepatocytes serves as a substrate for tissue transglutaminase during apoptosis. *J. Biol. Chem*. **267**, 25648-25651.

**Tezuka, T. and Takahashi, M.** (1987). The cystine-rich envelope protein from human epidermal *stratum corneum* cells. *J. Invest. Dermatol.* **88**, 47-51.

**Watt, F. M.** (1984). Selective migration of terminally differentiating cells from the basal layer of cultured human epidermis. *J. Cell Biol.* **98**, 16-21.

**Watt, F. M.** (1989). Terminal differentiation of epidermal keratinocytes. *Curr. Opin. Cell Biol.* **1**, 1107-1115.

**Yaffe, M. B., Murthy, S. and Eckert, R. L.** (1993). Evidence that involucrin is a covalently linked constituent of highly purified cultured keratinocyte cornified envelopes. *J. Invest. Dermatol.* **100**, 3-9.

**Yoneda, K., Hohl, D., McBride, O. W., Wang, M., Cehrs, K. U., Idler, W. W. and Steinert, P. M.** (1992). The human loricrin gene. *J. Biol. Chem.* **267**, 18060-18066.

**Yuspa, S. H., Kilkenny, A. E., Steinert, P. M. and Roop, D. R.** (1989). Expression of murine epidermal differentiation markers is tightly regulated by restricted extracellular calcium concentrations in vitro. *J. Cell Biol.* **109**, 1207-1217.

**Zettergren, J. G., Peterson, L. L. and Wuepper, K. D.** (1984). Keratolinin: The soluble substrate of epidermal transglutaminase from human and bovine tissue. *Proc. Nat. Acad. Sci. USA* **81**, 238-242.

**Zhou, X.-M., Idler, W. W., Steven, A. C., Roop, D. R. and Steinert, P. M.** (1988). The complete sequence of the human intermediate filament chain keratin 10. *J. Biol. Chem.* **263**, 15584-15589.

(Received 30 September 1993 - Accepted 5 November 1993)