Direct Evidence That Involucrin Is a Major Early Isopeptide Cross-linked Component of the Keratinocyte Cornified Cell Envelope

(Received for publication, June 28, 1996, and in revised form, August 27, 1996)

Peter M. Steinert‡ and Lyuben N. Marekov

From the Laboratory of Skin Biology, NIAMS, National Institutes of Health, Bethesda, Maryland 20892-2752

Involucrin was the first protein to be identified as a likely constituent of the insoluble cornified cell envelope (CE) of stratified squamous epithelia. However, to date, direct isolation from CEs of involucrin cross-linked by way of the transglutaminase-induced isopeptide bond has not been reported. We have treated human foreskin CEs with methanol/KOH (saponification) to hydrolyze off much of the lipids. By immunogold electron microscopy, this exposed large amounts of involucrin epitopes as well as of desmoplakin, a desmosomal structural protein. About 20% of the total CE protein could be solubilized by proteolytic digestion after saponification, of which involucrin was the most abundant. Subsequent amino acid sequencing revealed many peptides involving involucrin cross-linked either to itself or to a variety of other known CE protein components, including cystatin a, desmoplakin, elafin, keratins, members of the small proline-rich superfamily, loricrin, and unknown proteins related to the desmoplakin family. Specific glutamines or lysines of involucrin were used to cross-link the different proteins, such as glutamines 495 and 496 to desmoplakin, glutamine 288 to keratins, and lysines 468, 485, and 508 and glutamines 465 and 489 for interchain involucrin cross-links. Many identical peptides were obtained from immature CEs isolated from the inner living cell layers of foreskin epidermis. The multiple cross-linked partners of involucrin provide experimental confirmation that involucrin is an important early scaffold protein in the CE. Further, these data suggest that there is significant redundancy in the structural organization of the CE.

The insolubility of the protein portion of the CE is due to extensive cross-linking of several constituent proteins by both disulfide bonds and the N\(^{-}\)γ-glycyllysine isopeptide cross-link introduced by the action of transglutaminases (1–5). Analysis of the protein composition of the CE has been hampered by the simple fact that the cross-link cannot be cleaved by reagents that do not also cleave peptide bonds. Nevertheless, many studies using biochemical and immunological techniques have identified several protein components of CEs of epidermal or other epithelia, including cystatin a (9, 10), formerly named keratolin (11), elafin (12–15), involucrin (4, 16–21), 23, 24), loricrin (25–30), members of the small proline-rich superfamily (Spr) (Spr1 and Spr2 in epidermis, and Spr3 in cultured keratinocytes) (9, 31–36), filaggrin (37, 38), keratin intermediate filaments (39–42), and possibly trichohyalin (41). Indeed, recent amino acid sequencing of peptides has demonstrated for the first time that the proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and Sprs are in fact isopeptide cross-linked components of human foreskin epidermal CEs (26, 42).

However, detailed information on the relative abundances, temporal orders of deposition, or the assembly mechanisms of these proteins onto the CE structure is incomplete. Nevertheless, two points are becoming more clear. First, the CEs of different epithelia are not the same. For example, loricrin seems to be unique to the CE of “dry” or orthokeratinizing epithelia such as the epidermis (as well as the stomach of rodents) (25–29). Spr amounts in CEs vary widely in epithelia, from very abundant in the periderm layer of the fetus (36) to absent in CEs of the relatively thin interfollicular postnatal epidermis (35, 36), yet very abundant in the thickened epidermis of the lip, footpad, and foreskin, and other epithelia such as vagina, penis, and hair follicle cells, which are subject to considerable mechanical stresses (35). Therefore the composition of the constituent proteins and thus the structure of the CE seems to vary in parallel with the function of different epithelia (36, 42). Second, a variety of data in toto suggest that involucrin may be a ubiquitous component of the CEs of most if not all epithelial tissues (reviewed in Refs. 3, 4, and 21). One hypothesis suggests that it may serve as an early or scaffold component of CE structure (3, 4, 20, 21, 44) onto which other proteins such as Sprs (36), loricrin (3, 28, 38, 42, 44), or sulfur-rich proteins (8, 45, 46) are later added to effect final stabilization. Interestingly, involucrin also has been shown to be a component of the primitive CE entity formed in liver apoptotic bodies (47).

We have utilized CEs isolated from human foreskin epidermis as a model system to study these structural and functional issues. Our recent work using limited proteolysis and sequencing of the released peptides has provided information on not only which proteins are cross-linked, but also which glutamines and lysines are used for cross-linking by the transglutaminase enzymes in vivo (42). The strategy was successful because, as suggested

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence and reprint requests should be addressed: Bldg. 6 Room 425, NIAMS, NIH, Bethesda, MD 20892-2752. Tel.: 301-496-1578; Fax: 301-402-2886; E-mail: pemast@helix.nih.gov.

The abbreviations used are: CE, cornified cell envelope; PTH, phénylthiohydantoin; HPLC, high pressure liquid chromatography; Spr, small proline-rich family of proteins, consisting of several members such as Spr1, etc.
by immunogold decoration studies (44), the enzyme proteinase K could penetrate the cytoplasmic face but not the lipid face of this CE. We showed that the outer one-third of this CE structure consists almost entirely (>90%) of loricrin (i.e. inter- and intrachain cross-linked loricrin molecules) and was admixed with smaller amounts of Spr1 and Spr2 which seem to function as cross-bridging molecules among the loricrin. A middle third of the CE structure was also ~85% loricrin, admixed with Spsr, and the elastase inhibitor elafin. Trace amounts of keratins and filaggrin were also cross-linked. However, a similar analysis of the innermost protein portion of this CE, which was enriched in involucrin and which is perhaps common to the CEs of many other types of stratified squamous epithelia, was not possible. Thus, no data exist yet as to how involucrin is cross-linked with other proteins and to the loricrin-rich phase in the epidermal CE or in any other CE structure.

In this study, we have developed methods to circumvent these technical problems. We have characterized peptides from two sources: (i) mature stratum corneum CEs from which much of the covalently attached lipids have been removed by alkaline hydrolysis (saponification) and (ii) less mature CEs from the inner living cell layers of the epidermis. Our new data identify the many proteins to which involucrin is cross-linked in vivo and confirm that it is indeed a major early cross-linked component of the CE.

MATERIALS AND METHODS
Preparation of CEs—The epidermis of human foreskins was extracted in a buffer of 8 M urea, 50 mM Tris-HCl (pH 7.6), 1 mM EDTA. In the absence of a reducing reagent, only the inner living cell layers are dissociated (48, 49). This extract was then filtered through nylon gauze (mesh, 0.1 mm). The retentate of stratum corneum sheets and a filtrate of living cell material were separately pelleted, washed in phosphate-buffered saline, and then used to prepare mature and immature CEs, respectively, by exhaustive boiling in SDS buffer (38, 42, 44). The resultant CE fragments were pelleted through 20% Ficoll in phosphate-buffered saline to remove most adherent (i.e. not cross-linked) solubilized keratins (38, 44). Mature CEs were extracted in 1 x KOH, 95% methanol at 45° for 1 h (saponification). This reaction hydrolyzes the ester linkages by which the ceramide-rich lipid envelope is attached to the CE protein (6, 50, 51).

Protein Chemistry Procedures—Amino acid analysis of hydrolyzed samples (5.7 N HCl at 110° for 22 h in vacuo) was used routinely to measure protein amounts. The isopeptide cross-link was measured by amino acid analyses of complete enzymatic digestions of CE samples (26, 42). To generate peptides suitable for microsequencing, aliquots of CEs (1–3 mg of CE protein) before or after saponification were resuspended in 0.1 x N-ethylmorpholine acetate, pH 8.3. They were digested at 37 °C with either trypsin (Sigma; sequencing grade, 1% by weight) for 1–6 h or with proteinase K (Life Technologies, Inc.; 3% by weight for 15 residues long and eluted from the Bio-Rad HPLC column, whereas peptides derived from non-cross-linked portions of constituent CE proteins generally were ≥15 residues long and eluted <40 min. In the case of peptides from saponified CEs, since the initial tryptic peptides were >50 residues long and poorly resolved, they were subjected to a second digestion with proteinase K (0.5% by weight for 30 min) before fractionation. Selected peptides were sequenced as described before (42).

Immunogold Electron Microscopy—Several affinity-purified antibodies were used to decorate isolated CE fragments: (i) a polyclonal antibody generated in rabbits using as immunogen a synthetic peptide of the sequence PVCSPOGGQIVEWNNQSLNQLSNLVEIDPERKQVRSRE corresponding to the H1 region of the human keratin 1 chain (52, 53) (Western blotting methods on epidermal extracts demonstrated specificity for the epidermal type II keratins 1, 2e, and 5 (and 6 in cultured cells) data not shown); (ii) a rabbit polyclonal anti-human keratin 10 antibody (53); (iii) a goat polyclonal anti-human loricrin SAF-3 antibody (26); (iv) a mouse monoclonal anti-human involucrin antibody (Biomedical Technologies Inc., Stoughton, MA); (v) a rabbit polyclonal anti-human desmoplakin antibody (gift of Dr. R. D. Goldman); (vi) a polyclonal rabbit anti-rat plectin antibody (Sigma); (vii) a mouse monoclonal anti-human integrin α3 (ATCC, Rockville, MD); and (viii) a rabbit anti-human BPAG1 antibody (gift of Dr. J.C.R. Jones). Pellets of CEs were subjected to pre-embedding, labeled as described (25, 44, 54) and using protein A-gold with a diameter of either 5 or 10 nm.

RESULTS

Our earlier attempt to resolve the structure of the foreskin epidermal stratum corneum (“mature”) CE by controlled proteolysis was complicated by the presence on one side of the lipid envelope, which precluded direct access by proteases to the inner layers, and by the dense layer of cross-linked loricrin on the cytoplasmic side (42). Therefore, in this study, we have utilized mature CEs in which the ceramide-rich lipid envelope layer has been removed by alkaline hydrolysis (saponification) or CEs that contain much less loricrin and that have not yet assembled the lipid envelope (from “immature” epidermal cells).

Saponification of Mature CEs Reveals Buried Epitopes for Involucrin and Desmoplakin—We probed mature foreskin epidermal CEs for the presence of epitopes of several proteins that are known to be present, including involucrin and loricrin (42, 44) (Fig. 1, B and D, first part). The linear distributions of gold particles over >50 μm of CE fragments were summed to obtain more quantitative information (Table I). The distributions were reproducible between multiple experiments with each antibody, but there was wide variation between antibodies probably due to differences in epitope accessibility or abundance. Of three available keratin 1 antibodies, only the new one elicited against the H1 subdomain labeled fragments reliably (Fig. 1C).

We also tested for other keratinocyte cell peripheral antigens including α3 integrin (a marker for cell junctions in terminally differentiating keratinocytes (55)) and desmoplakin (56) (a marker of desmosomal junctions) as well as two markers for hemidesmosomal junctions, plectin (57) and BPAG1 (58). All of these were negative (Fig. LA for desmoplakin).

Following digestion with trypsin, epitopes for involucrin and K1 were lost, but epitopes for loricrin were retained (Fig. LA, second part), as described previously (44).

Next, mature CEs were treated with methanol/KOH (saponification), washed, and probed for the several epitopes. As a control, we found that there was little change in the amount or distribution of loricrin labeling between intact and saponified CEs (Fig. 1D, compare first and third parts; Table I; i.e. saponification had not altered the accessibility of loricrin epitopes on the cytoplasmic side. However, epitopes for desmoplakin, involucrin, and keratin 1 (but none of the other cell peripheral antigens tested) became exposed in substantial amounts (Fig. 1, A–C, third part; Table I). By use of double labeling experiments, we found that desmoplakin (Fig. 2A) and involucrin (Fig. 2B) labeling occurred on the side opposite to that of loricrin (Table I). This means that epitopes for desmoplakin and involucrin had become exposed after saponification. Similarly, epitopes of keratin 1 became exposed on both sides after saponification (Table I). Based on previous immunogold decoration studies, most of the keratin labeling on the cytoplasmic side was due to contaminating protein (44); thus, keratin 1 epitopes also became exposed on the delipidized inner surface.

Trypsin digestion after saponification removed ~25% of the total CE protein and resulted in loss of most epitopes, except only for those of loricrin, which became exposed on both sides (Fig. 1D, fourth part; Table I). Mathematical modeling (38, 44) of the amino acid compositions of the solubilized trypsin peptides showed marked enrichment for involucrin, whereas the 75% protein in the insoluble remnant was estimated to be >90% loricrin.

Together, these data suggest that the lipid layer of mature CEs had masked epitopes for certain inner CE protein com-
ments, which on exposure could be easily removed, leaving a remnant consisting almost entirely of polyloricrin.

**CEs from Immature Foreskin Epidermis**—Immature CEs obtained from the inner living layers of foreskin epidermis constituted about 1% of the protein mass of the epidermis, which is about one-tenth of that for mature CEs. By light microscopy, they consisted of mixtures, from “fragile” translucent to more rigid structures as reported previously (3) (data not shown). They are expected to contain only traces of lipids (3, 51). These mixed CEs were estimated to contained about 30% loricrin. Preliminary immunogold analyses revealed similar distributions of gold particles for the same antigens shown in Fig. 1 (data not shown). In contrast to the mature CEs, however, digestion with trypsin solubilized ~85% of the total protein. The undigested remnant was estimated to be highly enriched for loricrin.

**Characterization of Tryptic Peptides From CEs**—The tryptic peptides from mature CEs were recovered for sequencing analysis in the following way. First, the CEs were digested to completion with trypsin to remove ~5% of contaminating adherent non-cross-linked protein, mostly consisting of keratins and filaggrin (42, 44). This fraction contained ~0.5% of the total CE cross-link. The remnant was subjected to saponification and then redigested to completion with trypsin, which solubilized 19.5% of total CE protein mass, which contained 12.7 of 89 nmol of cross-link/mg of total CE protein. Following a brief digestion with proteinase K, the peptides were fractionated by HPLC (Fig. 3). In this way, 187 peptides ~15–50 residues long were recovered and sequenced, of which 157 contained one or more cross-link, so that they contained two or more “peptide branches.” In almost all cases, the structures of the peptides were solved in the sense that (i) in the two or more branches, the exact identity of the protein and the location within the protein was identified and (ii) we could unambiguously assign which glutamine(s) and lysine(s) were linked by the isopeptide cross-link. Table II illustrates examples of how sequence information was assigned. Together, the recovered and sequenced peptides included 402 peptide branches and accounted for 89% of the total amount of cross-link in the unfractionated tryptic peptide preparation (Table III); the remaining cross-link was present as short unresolvable peptides that eluted very early on the HPLC column (Fig. 3).

The solubilized tryptic peptides from the immature CEs were...
resolved (data not shown) and characterized similarly. In this case, the peptides contained 26.7 nmol of cross-link/mg of CE protein, of which 21.1 nmol (79%) could be accounted for in 131 peptides having 380 peptide branches. However, in three four-branched peptides, there was no unique solution as to which glutamine was cross-linked to which lysine (data not shown).

Analysis of Complete List of Sequences—Table III lists the yields and proteins of origin of the peptide branches from both sources of CEs, as well as the data for proteinase K digestion of mature CEs obtained previously (42). More than 83% (molar basis) of these sequences exactly matched known or suspected CE structural proteins (26, 42), of which by far the most abundant were involucrin (256 sequences; 36 or 14% molar basis) and loricrin (169 new sequences, total of 458). Interestingly, the third most abundant were a group of three closely related peptides (total of 93 times, 17 or 9% molar basis) of unknown identity, but they are homologous to human desmoplakin (59), BPAG1 (60), and plectin (61) (Table IV). The new cross-linked peptides involving loricrin discovered here were found to use the same glutamines and lysines of loricrin as seen before, so that the molar usage of these residues remained unchanged (42). All of the cross-links involving Spr1 and Spr2 proteins and elafin used amino- or carboxyl-terminal sequences as seen before, confirming the idea that these proteins serve as crossbridges between CE proteins, usually loricrin (42). A fourth major group of sequences involved the type II keratins 1, 2ε, and 5, which will be described in detail elsewhere. Table V lists the frequency of cross-linking between various protein partner pairs.

Analysis of the Involucrin Cross-links—Sequences involving involucrin were the second most abundant (Table VI). Most notably, it was evident that of a total of 150 glutamines and 45 lysines in involucrin (62), only a limited selection of them was used for cross-linking to specific proteins. In all 27 occurrences, the “unknown” protein was cross-linked only by way of involucrin lysine 485. In all 18 occurrences, involucrin glutamine 495 or 496 was used only to cross-link to desmoplakin. The type II keratin chains 1, 2ε, or 5 usually used involucrin glutamine 288. In other cases, there was somewhat less sequence specificity. Involucrin-involucrin cross-links involved glutamines 465 and 489 and lysines 468, 485, or 508; loricrin was cross-linked by way of several glutamines (308, 309, 368, 369, 425, 426, 455, or 456); and in involucrin-cystatin a and involucrin-elafin cross-links, as many as six glutamines each were used, although the exact residues were uncertain because of peptide repeats in involucrin. All of these residue positions are located in the center of involucrin, encompassing its modern sequences (62, 63).

Likewise, there was considerable conservation in the glutamine or lysine residues used for cross-linking within the other proteins, including elafin, cystatin a, and the keratins. In the case of desmoplakin, only two lysines in the entire sequence were used, located on its carboxyl tail at the end the C domain (59). Similarly, only two glutamines were used in homologous sequences of the unknown protein.

Thus, involucrin participated in many interchain cross-links with multiple different partner proteins (Table V). In some
Examples of interchain cross-linked peptides

Table II

| Interchain involucrin cross-link | Involucrin Gln<sup>465</sup> | Involucrin Lys<sup>466</sup> |
|----------------------------------|-----------------------------|-----------------------------|
| LEQEEKQLE | Involucrin Gln<sup>465</sup> | Involucrin Lys<sup>466</sup> |
| LEQEEKQLE | Involucrin Gln<sup>465</sup> | Involucrin Lys<sup>466</sup> |

Example 1<sup>a</sup>

Example 2<sup>b</sup>

Solved interchain multiprotein cross-link

E K Q E A Q L

A Q L Q D E S S Y E K D L

E K Q E A Q L E L P E Q Q V

I L T C P R T K

Involucrin Lys<sup>468</sup>

Unknown Gln<sup>3</sup>–Gln<sup>6</sup>

Involucrin Lys<sup>468</sup>–Gln<sup>495</sup>

Desmplakin Lys<sup>1559</sup>

Example 3<sup>c</sup>

Solved four-branched interchain involucrin cross-linked peptide

N L E Q E K Q L E L

E Q E K Q L E L

E Q E K Q L E L

N L E Q E K Q L E L

Involucrin Lys<sup>468</sup>

Involucrin Gln<sup>465</sup>–Lys<sup>468</sup>

Involucrin Gln<sup>465</sup>–Lys<sup>468</sup>

Involucrin Gln<sup>465</sup>

*See Table VI, peptide 10, first item. Cycle 1: L (295); cycle 2: E (145); cycle 3: Q (135); cycle 4: E (130); cycle 5: E (95); cycle 6: K (145), X (120); cycle 7: Q (230); cycle 8: L (230); cycle 9: E (60); cycle 10: L (20). (i) The sequence(s) are human involucrin. (ii) The molar amount of PTH-Gln released at cycle 3 is ~0.5. (iii) Both PTH-X and PTH-Lys (~0.5 mol each) are released at cycle 6. Therefore, two identical peptide sequences are present, cross-linked through Gln<sup>465</sup> of one and Ly<sup>468</sup> of the other. *See Table VI, peptide 17. Cycle 1: E (10), A (30), I (30); cycle 2: X (30), L (30); cycle 3: Q (55), T (30), L (30); cycle 4: E (10), L (30); cycle 5: A (55), P (25), X (25); cycle 6: Q (50), D (5); cycle 7: T (25), E (10), L (45); cycle 8: S (15), E (10), K (5); cycle 9: S (5), L (20); cycle 10: P (20), Y (15); cycle 11: E (5); cycle 12: K (10), X (15); cycle 13: D (~2), Q (10); cycle 14: V (~2) L (~2). (i) The data show that four different peptide sequences were present, joined by three cross-links. The sequences were determined empirically and then confirmed by use of the Swiss Protein Database. The strong clues were that these four sequences were commonly identified in the large body of cross-linked peptides analyzed in this study (summarized in Tables III and VI), and that two involucrin sequences, one longer, were present. (ii) In the involucrin sequences, the Ly<sup>468</sup> residues were not seen; and in the longer, one Gln<sup>495</sup> was not seen. (iii) Therefore, the long involucrin and the unknown sequences served as cross-bridges for the desmplakin and short involucrin sequence. (v) Because one X was released at cycle 2, it is most likely that Ly<sup>468</sup> of the short involucrin sequence was cross-linked to Gln<sup>3</sup> of the unknown sequence and that Ly<sup>1559</sup> of desmplakin was cross-linked to Gln<sup>465</sup> of the longer involucrin sequence. *See Table VI, peptide 18. Cycle 1: N (105), E (20); cycle 2: L (200); cycle 3: E (30); cycle 4: Q (100), E (20); cycle 5: E (30), X (195); cycle 6: Q (195), E (20); cycle 7: K (100), X (95), L (190); cycle 8: Q (185), E (15); cycle 9: L (140), cycle 10: E (10); cycle 11: L (5). (i) The strong clue here is that at cycle 7 four equimolar derivatives were released, suggesting that the peptide contains four branches adjoined by three cross-links. (ii) Based on the times of types of residues released, the sequences are almost certainly involucrin. (iii) A second strong clue is that Asn residues are rare in involucrin. (iv) From cycles 1 and 2 and cycles 7–11, using the derivatives released at ~100 pmol/cycle, the sequences can be identified unambiguously as lying between involucrin residues 462 and 472. (v) With this information, it becomes clear there are two pairs of sequences: NLEQEEKQLEL and EQEEKQLEL each pair offset by two residues. Note: the low yields of E and L at cycles 10 and 11 imply their presence on only one sequence, shown at the top. (vi) In the first pair, Gln<sup>465</sup> is seen once, but not both times; in the second pair, both Gln<sup>465</sup> and Lys<sup>468</sup> are not seen but are released as X. (vii) Thus, the second pair of sequences serve as cross-bridges. (viii) Therefore Lys<sup>468</sup> on one branch is cross-linked to a Gln<sup>465</sup> on a second branch; its neighboring Lys<sup>468</sup> is cross-linked to Gln<sup>465</sup> on a third branch; its neighboring Lys<sup>468</sup> is cross-linked to Gln<sup>465</sup> on a fourth branch.

DISCUSSION

The Importance of Involucrin—Historically, involucrin was the first protein to be identified as a constituent of the CE formed in epithelial cells (16). Many studies have since characterized in detail its expression, in vitro transglutaminase cross-linking, biochemical properties, and structural properties (reviewed in Refs. 3, 4, and 21). The data from all of these studies are consistent with the view that involucrin is a covalently attached "early" component of the CE. One extant hypothesis holds that it may serve as a scaffold for the later attachment of other CE structural proteins (3, 4, 20, 21, 44, 65). Two recent studies have shown unequivocally that the same involucrin-immunoreactive fragments can be released from CEs of foreskin epidermis and cultured keratinocytes by use of CNBr methods (21, 66), indicating that involucrin is covalently attached to the CEs. A 68-kDa fragment was released, which indicates that it had been covalently linked by way of the amino-terminal half of the intact protein (66). However, to date, sequences of involucrin joined together by way of the isopeptide cross-link to itself or another protein have not been isolated and characterized, as has been the case for several other "later" CE proteins (26, 42). The present study reports...
the identification and characterization of a large number of peptides containing one or more cross-links that adjoin involucrin itself and/or to other proteins. These data, together with our immunogold work, provide important information on the likely multiple roles of involucrin in the CE. Our data are predicated on the operational definition of the CE as that which is insoluble after exhaustive extraction with powerful protein solvents that break disulfide bonds but not peptide bonds; i.e., the CEs used in these studies are an insoluble protein complex cross-linked by isopeptide bonds (38, 42–44). However, it is also possible that involucrin and other structural proteins and enzymes are associated with or even covalently attached to the CE but do not become cross-linked by transglutaminases. In this case, they may be lost by our method of isolation of CEs and thus not recognized in this study.

Cross-linked Involucrin Is Indeed a Major Component of CEs

Formed in Vivo—In our previous studies on mature foreskin CEs, we were unable to find cross-linked peptides involving involucrin (42), in part because the exhaustive proteolysis procedures from the cytoplasmic side rendered the peptides too small for sequencing, although they were predicted to be enriched involucrin, and in part because the lipid envelope seemed to have precluded proteolytic access to the inner surface of the CE structure. However, we show here that following alkaline hydrolysis to remove lipids from the mature CEs formed in vivo, epitopes for several proteins become exposed, including involucrin, keratin 1, and desmoplakin. After saponification, trypsin could release about 20% of the protein mass of the CE. Sequencing of these peptides revealed prominent amounts of interchain cross-linked species involving these proteins, of which involucrin was in fact the most abundant (40%; Table III). Many peptides involved interchain cross-links be-

### Table III

**Occurrences of sequences of known proteins in cross-linked peptides**

| Protein          | Immature CEs | Mature CEs | 3 h trypsin/0.5 h proteinase K after saponification | 9 h proteinase K before saponification |
|------------------|-------------|------------|---------------------------------------------------|---------------------------------------|
|                  | Number | Yield | Number | Yield | Number | Yield |
| Cystatin         | 15     | 1690  | 21     | 1010  | 1      | 20   |
| Desmoplakin      | 18     | 1250  | 25     | 1270  | 20     | 71   |
| Elafin           | 21     | 1520  | 16     | 250   | 18     | 7050 |
| Filaggrin        | 95     | 3610  | 161    | 4510  | 9      | 70   |
| Involucrin       | 37     | 850   | 11     | 340   | 6      | 70   |
| Keratin 1        | 2      | 50    | 4      | 60    | 1      | 2    |
| Keratin 2        | 43     | 1100  | 12     | 390   | 1      | 10   |
| Keratin 5        | 2      | 50    | 4      | 60    | 1      | 2    |
| Keratin 10       | 43     | 1100  | 12     | 390   | 1      | 10   |
| Loricrin         | 110    | 7540  | 59     | 1150  | 299    | 54530|
| Spr1 or Spr3     | 12     | 590   | 7      | 420   | 8      | 1100 |
| Keratin 10       | 1      | 20    | 1      | 10    | 1      | 10   |
| Loricrin         | 15     | 1940  | 7      | 1850  | 4      | 120  |
| Unidentified     | 131    | 157   | 155    | 155   |
| Number of peptides | 430   | 380   | 402    | 356   |
| Number of peptide "branches" | 26.7 nmol/mg | 12.7 of 89 nmol/mg | 70.5 of 89 nmol/mg |
| Total amount of cross-link in CEs | 21.1 of 26.7 (79%) | 11.3 of 12.7 (89%) | 64.9 of 70.5 (92%) |

*The number in parenthesis is the total number of occurrences of the protein in the three experiments.

**Table IV**

**Unknown peptide sequence is related but not identical to human bullous pemphigoid antigen, desmoplakin, and plectin**

Sequences are from desmoplakin (position 1640) (59), bullous pemphigoid antigen (position 2034) (60), and plectin (position 4517) (61). Amino acids shown in boldface type participate in identified cross-links.

**Unknown variants**

- 46 times: A Q L L Q D E S S F E K D L
- 34 times: A Q L L Q D E S S Y E K D L
- 13 times: A Q L Q D A S S F E K V L L

**Human plectin**

- RTAQKLRDVGAYSKYLTCPKTK

**Human desmoplakin**

- RAARQLDTSYAKYLTCPKTK

**Human BPAG1**

- LIATKLKDQKSYVRNICAL

**Table V**

**Frequency of cross-linking between identified protein partners**

| Cystatin | Desmoplakin | Elafin | Filaggrin | Unknown | Involucrin | Keratin | Loricrin | Sprs |
|----------|-------------|--------|-----------|---------|------------|---------|----------|------|
| 0        | 6           | 3      | 0         | 0       | 0          | 0       | 0        | 0    |
| 6        | 0           | 4      | 0         | 0       | 0          | 0       | 0        | 0    |
| 1        | 0           | 0      | 0         | 0       | 0          | 0       | 0        | 0    |
| 3        | 15          | 4      | 0         | 0       | 0          | 0       | 0        | 0    |
| 9        | 18          | 5      | 0         | 27      | 162        | 0       | 0        | 0    |
| 4        | 0           | 13     | 0         | 31      | 8          | 0       | 0        | 0    |
| 11       | 0           | 26     | 0         | 10      | 25         | 53      | 289      | 0    |
| 1        | 2           | 0      | 0         | 3       | 2          | 6       | 40       | 0    |
Table VI
Cross-links involving involucrin

The data are generated from a total of 109 peptides from the two experiments involving one or more involucrin branches, for a total of 256 involucrin sequences. In experiments where the same sequence appeared several times, the peptides were separable by HPLC due to varying lengths of the branches.

| Proteins | Number of peptides | Immature | Mature saponified | Site(s) on involucrin | Site(s) on other protein |
|----------|--------------------|----------|-------------------|----------------------|-------------------------|
| **Dipeptides** | | | | | |
| 1. Involucrin-cystatin α | 9 | 2 | 7 | Gln172/Gln202 | Lys46 |
| | | | | Gln202/Gln342 | |
| | | | | Gln342/Gln402 | |
| 2. Involucrin-desmoplakin | 3 | 1 | 2 | Gln95 | Lys1359 |
| | | | | Gln96 | Lys1361 |
| | | | | Gln96 | |
| 3. Involucrin-elafin | 5 | 2 | 3 | Gln158/Gln178 | Lys6 |
| | | | | Gln158/Gln196 | |
| | | | | Gln196/Gln218 | |
| 4. Involucrin-unknown | 10 | 2 | 8 | Lys385 | Glu3 |
| | | | | Lys385 | Glu6 |
| 5. Involucrin-keratin 1 | 1 | 0 | 1 | Gln105/Gln128 | Lys83 |
| 6. Involucrin-keratin 2e | 2 | 1 | 1 | Gln65 | Lys69 |
| 7. Involucrin-keratin 5 | 5 | 1 | 4 | Lys65 | Lys77 |
| 8. Involucrin-Spr1 | 2 | 1 | 1 | Gln65 | Lys78 |
| 9. Involucrin-locrin | 2 | 1 | 1 | Gln65 | Lys9 |
| | | | | Lys105/Gln128 | |
| 10. Involucrin-Involucrin | 14 | 5 | 9 | Gln105/Lys128 | Gln65 |
| | | | | Gln105/Lys128 | |
| 11. Involucrin-desmoplakin-involucrin | 2 | 1 | 1 | Glu105 | Lys105/Lys128 |
| | | | | Glu105/Lys128 | |
| 12. Involucrin-unknown-involucrin | 6 | 3 | 3 | Lys385 | Glu3 |
| 13. Involucrin-unknown-cystatin α | 1 | 1 | 0 | Lys385 | Glu3 |
| 14. Involucrin-involucrin-desmoplakin | 2 | 1 | 1 | Gln105 | Lys105/Lys128 |
| 15. Involucrin-locrin-involucrin | 1 | 1 | 0 | Gln125 | Lys107 |
| 16. Involucrin-involucrin-involucrin | 8 | 4 | 4 | Glu385 | |
| 17. Involucrin-unknown-involucrin-desmoplakin | 1 | 1 | | Lys385 | Glu3 |
| 18. Involucrin-involucrin-involucrin-involucrin | 5 | 5 | | Gln105 | Lys1359 |
| 19. Involucrin-involucrin-involucrin-involucrin | 2 | 1 | 1 | Gln385 | Lys1359 |
| | | | | Gln385 | |
| | | | | Gln385 | |
| | | | | Gln385 | |

*In sequences denoted by a slash, the exact residue position is uncertain due to involucrin sequence repeats.*
Involucrin in the Cornified Cell Envelope

First, based on the yields of proteins released by proteolysis with different enzymes both before (late proteins) and after (early proteins) saponification (Table III), we can make estimates more accurately than heretofore possible of their total amounts in intact mature CEs (Fig. 4). As foreshadowed in earlier predictive analyses (3, 4, 21), these data illustrate quantitatively the inverse relationship between the amounts of early proteins involucrin and cystatin α (and the newly encountered desmoplakin and unknown proteins), and the late proteins elafin, Sprs, and loricin. The cross-link data from immature CEs, presumably obtained for CEs of an early degree of maturation, seem to afford an intermediary stage in this progression. These analyses provide the best evidence to date for the likelihood of an orderly temporal accumulation of proteins as the CE is assembled (3, 44). We calculate that mature foreskin epidermal CEs contain about 5.5% involucrin and 2–3% each of desmoplakin and the unknown proteins but more than 70% loricin. This value for involucrin is about twice that estimated by mathematical modeling (38, 44) but is within the range of accuracy of the method. Furthermore, since the CE itself is 10% of the mass of cornified keratinocyte, of which 90% is protein (3, 4), this means that 0.4–0.5% of the protein mass of the epidermis is involucrin cross-linked in the CE. Since involucrin constitutes about 1% of total epidermal keratinocyte cell protein (20, 73), this suggests that only part of it in fact becomes tightly cross-linked to the CE. This raises the intriguing possibility that a considerable amount of involucrin is associated with the CE in other ways: e.g. (i) is covalently attached by some other methods (including disulfide bonds, polyamines, or lipids); (ii) is partially cross-linked through glutamine and lysine residues other than those reported here (66); (iii) remains soluble; or (iv) may be utilized for some other purpose in the keratinocyte.

Third, we are intrigued by the abundance of cross-links between involucrin and carboxyl-terminal sequences of desmoplakin. Desmoplakin is a major structural protein of desmosomes. While part is perhaps anchored at the cell junction, a central rod domain projects into the cytoplasm, and a series of peptide repeating domains at the carboxyl-terminal end are believed to interact directly or indirectly with cytoskeletal intermediate filaments (56, 59). Our pre-embedded electron microscopy images show that in isolated CEs the desmosomes have lost their structural integrity, but desmosomal remnants could be recognized by the fact that desmoplakin antigens were exposed and located in thickened zones along the CE fragments (Fig. 1A, third part). The isolation of cross-links involving desmoplakin carboxyl-terminal sequences indicates that at least this portion had become attached to the CE by the action of transglutaminases. Perhaps more substantial portions of desmoplakin also form part of the CE, since epitopes thought to be located within its rod domain were recognized by the polyclonal antibody (Fig. 1A, third part). Moreover, we found remarkable specificity in cross-linking, since only two glutamates (residues 495 and 496) of involucrin were used to cross-link to only three lysines of desmoplakin (residues 1659, 1661, and 1667) (Table VI). Notably, in vitro cross-linking experiments with the model amine donor putrescine have documented previously that glutamates 495 and 496 are the most highly reactive residues in involucrin (19). An earlier study reported immunogold localization of two monoclonal antibodies to desmosomal remnants in cornifying epidermal cells (74), which may be related to desmoplakin or the unknown proteins described here. Therefore, further experiments now will be required to test the interesting possibility that the cross-linking of involucrin onto desmosomal proteins may be a very early step in CE assembly.

The Multiple Scaffold Roles of Involucrin—Previous structural analyses have suggested that involucrin may function as a scaffold during the assembly of the CE (4, 21). The data of Tables V and VI provide robust support for this concept. First, involucrin was most commonly interchain cross-linked to itself by way of neighboring lysines and glutamines (Table VI, peptides 10, 14, 16, 18, and 19). Second, >20% of involucrin cross-links involved other early CE components such as desmo-
plakin, cystatin α, and the unknown protein (peptides 1, 2, 4, 11–14, and 17). However, another 15% of involucrin cross-links involved the late CE proteins elafin, loricrin, and Sprs (peptides 3, 5–9, and 15). Taken together, these observations indicate that an intermolecularly cross-linked polymeric layer involucrin not only seems to form an early part of the CE but also serves as a platform for the addition of the late CE proteins. Moreover, some proteins had preferred cross-linked partners such as desmoplakin with involucrin or the unknown protein with involucrin or keratins; and Sprs with loricrin, etc. More significantly, while most proteins had no ill effect (28). Conversely, it could be predicted that diminished levels of involucrin may not cause a seriously negative phenomenon.
Involucrin in the Cornified Cell Envelope

(1991) J. Biol. Chem. 266, 12555–12559
61. Lu, C.-G., Maercker, C., Castaño, M. J., Hauptmann, R. & Wiche, G. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 4278–4283
62. Eckert, R. L. & Green, H. (1988) Cell 46, 583–589
63. Green, H. & Dijan, P. (1992) Mol. Biol. Evol. 9, 977–1017
64. Folk, J. E. (1980) Annu. Rev. Biochem. 49, 517–531
65. Ishida-Yamamoto, A., Ikeda, R. A., Watt, F. M., Hoop, D. R., Hohl, D. and Iizuka, H. (1996) J. Histochem. Cytochem. 44, 167–175
66. Robinson, N. A., LaCelle, P. T. & Eckert, R. L. (1996) J. Invest. Dermatol. 107, 101–107
67. Green, K. J., Virata, M. L., Elgart, G. W., Stanley, J. R. & Parry, D. A. D. (1992) Int. J. Biol. Macromol. 14, 145–153
68. Lioiska, N., Yang, H.-Y. & Goldman, R. D. (1985) J. Cell Biol. 100, 629–633
69. Skalli, O., Jones, J. C. R., Gaginsek, R. & Goldman, R. D. (1994) J. Cell Biol. 125, 159–170
70. Sonnenberg, A., Lindler, C. J. T., Daams, J. H. & Kennel, S. J. (1990) J. Cell Sci. 96, 201–217
71. Fujiwara, S., Kohno, K., Iwantatsu, A., Naito, I. & Shinkai, H. (1996) J. Invest. Dermatol. 106, 1125–1130
72. Ma, A. S. F. & Sun, T.-T. (1986) J. Cell Biol. 103, 41–48
73. Etoh, Y., Simon, M. & Green, H. (1986) Biochem. Biophys. Res. Commun. 136, 51–56
74. Serre, G., Milis, V., Haftek, M., Vincent, C., Croute, F., Réano, A., Ouahayoun, J.-P., Bettinger, S. & Soleilhacou, J.-P. (1991) J. Invest. Dermatol. 97, 1061–1072
75. Lazo, N. D., Meine, J. G. & Downing, D. T. (1995) J. Invest. Dermatol. 105, 296–300
76. Watt, F. M. & Green, H. (1982) Nature 295, 434–436