Trichohyalin Mechanically Strengthens the Hair Follicle

MULTIPLE CROSS-BRIDGING ROLES IN THE INNER ROOT SHEATH*

Received for publication, February 26, 2003, and in revised form, June 20, 2003
Published, JBC Papers in Press, July 9, 2003, DOI 10.1074/jbc.M302037200

Peter M. Steiner†‡, David A. D. Parry§, and Lyuben N. Marekov¶

From the ¶Laboratory of Skin Biology, NIAMS, National Institutes of Health, Bethesda, Maryland 20892-8023 and the §Institute of Fundamental Sciences, Massey University, Palmerston North, Private Bag 11-222, New Zealand

Trichohyalin is expressed in specialized epithelia that are unusually mechanically strong, such as the inner root sheath cells of the hair follicle. We have previously shown that trichohyalin is sequentially subjected to post-synthetic modifications by peptidylarginine deaminases, which convert many of its arginines to citrullines, and by transglutaminases, which introduce inter- and interprotein chain cross-links. Here we have characterized in detail the proteins to which it becomes cross-linked in vivo in the inner root sheath of the mouse hair follicle. We suggest that it has three principal roles. First, it serves as an interfilamentous matrix protein by becoming cross-linked both to itself and to the head and tail end domains of the inner root sheath keratin intermediate filament chains. A new antibody reveals that arginines of the tail domains of the keratins are modified to citrullines before cross-linking, which clarifies previous studies. Second, trichohyalin serves as a cross-bridging reinforcement protein of the cornified cell envelope of the inner root sheath cells by becoming cross-linked to several known or novel barrier proteins, including involucrin, small proline-rich proteins, repetitive, and epilakin. Third, it coordinates linkage between the keratin filaments and cell envelope to form a seamless continuum. Together, our new data document that trichohyalin is a multi-functional cross-bridging protein that functions in the inner root sheath and perhaps in other specialized epithelial tissues by conferring to and coordinating mechanical strength between their peripheral cell envelope barrier structures and their cytoplasmic keratin filament networks.

Native trichohyalin (THH) is a large highly charged α-helical and insoluble protein that is expressed in specialized mammalian epithelial cell types (1–3). The tissues in which it is most abundantly expressed include the inner root sheath (IRS) cells of the hair follicle (approximately one third of total protein) (4, 5) and the medulla, a central column of cells within many coarse hairs (most of total protein) (4–6). THH is also expressed in trace amounts in other tissues such as newborn human foreskin epidermis, the hard palate, and rodent forestomach alone (6–11), or colocalized with filaggrin in hybrid granules in the filiform ridges of the tongue, the nail bed, and hyperkeratotic epidermis in skin diseases (11). Interestingly, each of these tissues is especially hardened or toughened to withstand mechanical abrasion and wear-and-tear during normal use. Thus, the question has arisen as to whether and how THH might contribute greater mechanical strength to these tissues (12).

THH can be recovered intact from these tissues only before their terminal differentiation (1, 5). However, recovery from mature tissues requires proteolysis (13–15) because it is subjected to extensive postsynthetic modification. First, many of its arginine residues are modified to citrullines (14–16) by a group of enzymes termed peptidylarginine deiminases (PAD) (17–23). In the case of THH in vitro, THH can be recovered intact from these tissues only before their terminal differentiation (1, 5). However, recovery from mature tissues requires proteolysis (13–15) because it is subjected to extensive postsynthetic modification. First, many of its arginine residues are modified to citrullines (14–16) by a group of enzymes termed peptidylarginine deiminases (PAD) (17–23). In the case of THH in vitro, this reaction destroys its α-helical structure to a random coil, makes it more soluble in physiological buffers, and thereby apparently renders it more amenable to subsequent modifications (Fig. 1) (24). An important second modification is cross-linking by transglutaminase (TGase) enzymes (25, 26), which catalyze the formation of an isopeptide bond between peptide-bound glutamine and lysine residues; the net result is a stable insoluble protein polymer complex (27–31). In the case of the medulla cells, the amorphous THH protein is extensively cross-linked to itself. In the nearby IRS cells of the hair follicle, there is also a very high content of cross-link (26, 32). Earlier data suggested that these may link THH to the keratin intermediate filaments (KIF) characteristic of this tissue because KIF can be released from mature IRS cells only after a brief proteolytic digestion step (14) that clips off keratin head and/or tail domain sequences (32). In addition, we have identified a few cross-linked peptides involving THH linked to several barrier protein components of the cell envelope (CE) of mouse forestomach tissue (12).

The purpose of this study is to fully characterize the utilization of THH in the IRS tissue in which it is abundantly expressed. Our new protein sequence data provide robust support for the biomechanical role of THH in stabilization of cell structure by coordination of the CE with the KIF-THH complex within the IRS cells.

EXPERIMENTAL PROCEDURES
Preparation of Mouse IRS Tissue

IRS tissue was harvested from the hair follicles of newborn albino mice (32). Hair follicles isolated from trunk skin dermis of newborn (<1 day old) albino mice (33) were extracted in a buffer of 25 mM Tris-HCl (pH 8.0) containing 5 mM urea. This buffer dissolves the lower portions of the hair follicles, thereby releasing the mature hardened IRS tissues as discrete conical, cylindrical multicellular structures. It does not appreciably dissolve the detached hair fiber material. The urea suspension was filtered through nylon gauze (pore size approximately 0.2 mm)
through which the IRS tissues pass but hair fibers do not. In our experience, approximately 60% of the mass of these hair follicles is IRS tissue; the hair fiber cortical tissue is still poorly developed, and there are no hair fibers. The highly enriched IRS tissue was pelleted by centrifugation at 100,000 × g and gently washed in phosphate-buffered saline to remove urea but not to dissociate the IRS. We found in preliminary sequencing experiments that mouse IRS tissue was contaminated by solubilized outer root sheath proteins, particularly Trx and K16. To obtain a fraction enriched in KIF proteins (34), newborn mouse hair follicles were homogenized by sonication in phosphate-buffered saline containing 0.5 M KCl, and a protease mixture kit (Serva, Heidelberg, Germany). This suspension was made to 5 mM dithiothreitol, and fractionated on a 95/5 H11003/mM dithiothreitol, and fractionated on a 95/5 H11003 grade trypsin (Sigma) (0.5 mg/ml) for 10 min at 23 °C. Trypsin digestion (32). Pellets of IRS were resuspended in 25 mM sodium phosphate (pH 8.0) (2 mg/ml) and digested with sequencing grade endopeptidase Asp-N (sequencing grade) (Roche Molecular Biochemicals, Mannheim, Germany). This procedure was used to solubilize cytoplasmic THH and KIF proteins without significantly proteolyzing CEs, the proteins of which generally contain few aspartic acid residues. The solubilized material was recovered by centrifugation at 15,000 × g, dissolved in gel sample buffer, and resolved on 8–16% gradient polyacrylamide gels (NoveX, Invitrogen).

Isolation of KIF and Three Peptide Fractions

The recovery of three protein fractions for subsequent sequencing analyses was as illustrated in Fig. 2. Fraction A Peptides—First, IRS KIF were harvested following limited trypsin digestion (32). Pellets of IRS were resuspended in 25 mM sodium phosphate (pH 8.0) (2 mg/ml) and digested with sequencing grade trypsin (Sigma) (0.5 mg/ml) for 10 min at 23 °C, and terminated by addition of a 1.5-fold molar excess of trypsin inhibitor (Sigma). This dissociated the IRS tissues into single cells that were pelleted at 500 × g, resuspended at the same concentration, chilled to 4 °C, and sonicated for 30 s (three 10-s bursts). Cellular debris was removed by centrifugation at 10,000 × g for 15 min. The KIF retained in the supernatant were pelleted at 100,000 × g in an Airfuge (Beckman Instruments). They were then dissociated in phosphate buffer containing 0.5% SDS and 1 mM dithiothreitol, and fractionated on a 95 × 2-cm column of Sephadex G-200 (32). Three protein peaks were recovered for further analyses. The broad low molecular weight fraction (V1) was collected and further digested to completion with trypsin (1% by weight for 16 h at 37 °C), and peptides were retained for HPLC separation. Fraction B Peptides—Alternatively, purified IRS tissue was digested to completion (1% by weight for 16 h at 37 °C) in a buffer of 0.1 M N-ethylmorpholine acetate (pH 8.0) with endopeptidase Asp-N (sequencing grade) (Roche Molecular Biochemicals, Mannheim, Germany). This procedure was used to solubilize cytoplasmic THH and KIF proteins without significantly proteolyzing CEs, the proteins of which generally contain few aspartic acid residues. The solubilized material was recovered by centrifugation at 15,000 × g for 15 min, digested further with trypsin (1% by weight for 3 h), and retained for HPLC separation.

Fraction C Peptides—The Asp-N pellet described above consisted almost entirely of intact or fragments of CEs, presumably mostly from the cuticle IRS cells. These were further digested in the same buffer with proteinase K (Invertogen, 3% by weight for 3 h), clarified by centrifugation at 15,000 × g for 15 min, and the solubilized material used for HPLC separation.

Protein Chemical and Sequencing Procedures

Amino acid analysis was routinely used to measure protein amounts. Amounts of citrulline were corrected for hydrolytic losses to ornithine as described previously (24). The amount of isodipeptide cross-link was measured following total enzymic digestion of desired fractions and then quantitated by amino acid analysis (35). Negative controls included peptides known not to contain the cross-link or samples of an unmodified expressed portion of human THH termed THH-8 (7, 24). Peptides suspected of containing cross-links recovered from the three fractions described above were resolved by HPLC on a 2.1 × 250-mm C18 column (Phenomenex, Torrance, CA) with a flow rate of 0.25 ml/min and an acetonitrile gradient of 5–60% (90 min) followed by 60–95% (15 min). We harvested those peptides eluted with >30% acetonitrile, covalently attached them to a solid support, and performed sequencing for up to 15 Edman degradation cycles as described previously (7, 12). Empirically we found that such peptide species contained >15 residues and collectively accounted for >90% of the total cross-link content of the samples. Additionally, on sequencing, these typically contained two or more peptide “branches” adjoined by one or more cross-links. Sequences could be assigned from database searches. The phenylthiohydantoin derivative of citrulline eluted at the same time as threonine in our system (Porton LF-3000 gas phase), and assignment was rarely in doubt because of the paucity of threonine in the proteins.

Generation of Polyclonal Antibodies and Indirect Immunofluorescence Procedures

Peptides corresponding the carboxyl terminus of mouse c29 protein (36) of sequence STKVRNTEQZIPS (IRSwt) and STKVRNTEQZIPS (IRScit) were synthesized, coupled to an 8-branched inert matrix, and used to generate polyclonal antibodies in guinea pigs, which were subsequently affinity-purified using the peptide antigen. Frozen serial sections of 5-day-old mouse skin (7 mm) were used for indirect immunofluorescence methods exactly as described previously (7). Primary antibodies used were: polyclonal rabbit anti-THH-8 (dilution 1:30) (7), polyclonal goat anti-human TCase 1 (dilution 1:50) (37), polyclonal guinea pig anti-IRSwt (dilution 1:100), polyclonal guinea pig anti-IRScit (dilution 1:200), and mouse monoclonal anti-K6 (Roche Molecular Biochemicals). Primary antibodies were then visualized using the appropriate affinity-purified secondary antibodies (Cappel Organon Technika Corp., Durham, NC) (7). Western blotting, dot blotting, and immunoprecipitation methods were used exactly as described previously (7).

RESULTS AND DISCUSSION

The purpose of this study is to more fully characterize the utilization and function of THH in a tissue in which it is abundantly expressed: the IRS of the mouse hair follicle. We generated three different fractions of peptides (Fig. 2) designed to separately explore the possible multiple functions of THH. The amounts of cross-link and citrulline were measured. Numerous peptides containing cross-links from each fraction were sequenced. Table I lists and sorts the protein partners. Each will be discussed in detail below.
Comparison of THH Sequences—The sequences of human (1897 residues; Ref. 2), rabbit (1407 residues; accession no. P37709), mouse (1439 residues, accession no. XP_177952), and sheep (1549 residues; Ref. 3) THH proteins are available. All display marked similarities, yet notable differences; they consist of varying numbers of predominantly α-helical, highly charged quasi-repeating peptide motifs that are poorly conserved although recognizably similar between species. Human THH has been divided into nine domains, based on regions of either high or low sequence regularity among adjacent repeats (2). By use of the University of Wisconsin Genomics Computer Group software, domain 1 of mouse THH consists of two well defined EF-hand repeats (residues 1–95), domain 2 (96–243), and domains 3 and 4 (244–373), each similar to human THH.

Next, its domain 5 (residues 374–595) consists of irregular sequence repeats of lower overall predicted α-helical content, roughly corresponding to human domains 5 and 7. Residues 596–697 define a region here termed domain 5* of weakly defined repeats unique to mouse THH. Residues 698–880 and 881–1406 of mouse THH define sets of well ordered sequences of different repeat motifs that are equivalent to human domains 6 and 8, respectively. Finally, domain 9 (residues 1407–1439) is homologous to human.

Direct Evidence That THH Functions as an Interfilamentous KIF Cross-bridging Protein in the IRS—KIF were harvested from mature IRS by use of a preliminary limited trypsin digestion (14, 32), which generated a KIF yield of approximately 40% of total IRS protein mass. Some of the cellular debris after brief trypsinization consisted of clumps of KIF apparently still cross-linked together. Unlike all other types of IF we have investigated, these IRS KIF are rigid straight rods that have a central densely staining core (Fig. 3A). This is reminiscent of the core observed in the center of the distinctly different trichocyte KIF of the adjacent hair fiber cortical cells (38). The KIF were pelleted and dissolved in SDS buffer. Upon fractionation by column chromatography, three peaks were recovered (Fig. 3B). The first and second peaks contained only traces of cross-link (<1 residue/1000 residues) and citrulline (ratio < 0.01), and presumably contained somewhat pruned IF chains, as judged by their high α-helix contents (data not shown). However, the peak eluted at Vc contained >90% of the citrulline and isopeptide cross-link of the isolated KIF.

This Vc material, termed fraction A, was further digested to completion with trypsin and the peptides resolved by HPLC (Fig. 4A). From the resultant reproducible profile, we recovered 38 well resolved peaks that reflect quantitatively major species. Sequencing revealed that 35 possessed one or more cross-links; 28 peptides had one cross-link, 4 had two cross-links, and 3 had three cross-links, so that a total of 44 peptide partners was found (Tables I and III). Together, these are likely to be highly representative, as they accounted for 69% of the total cross-link and 73% of the total citrulline content of the Vc fraction. It remains likely, however, that many other minor peptide/protein species have been overlooked. Another 20 peptides involving KIF chains, having a total of 30 partners, were recovered from fraction B (Fig. 4C and Table III).

### Table I

| Partners | Fraction A | Fraction B | Fraction C |
|----------|------------|------------|------------|
| THH-THH  | 77         | 3          |            |
| THH-keratin | 45b       | 30b        | 6b         |
| THH-CE proteins | 10 | 34 |            |
| CE-CE** | 41         | 3          |            |
| CE-keratinb | 1         | 3          |            |

* Listed in Table III.  
** Listed in Table IV.

Analyses of the combined data revealed two types of sequences: those from THH, and those from type I and type II keratin chains. For THH, all cross-links involved sequences of the highly regular domain 6 and 8 regions and the consensus THH peptide sequences involved were DZK(F/I)(Z/R)(Z/R) for the lysine residues and (E/Z)/E(Q/E)/Z/L/Z for the glutamine residues. Based on ongoing studies of murine, ovine, and human data, the IRS uniquely expresses a set of least three (ovine) type I keratin chains (39–42). One mouse sequence (KRT1-c29) has been reported (39), which has highest homology to the sheep type I IRSa3 proteins (36). Our new sequencing data reveal that tail domain sequences of c29 were commonly found to participate in cross-linking, and that its arginines were usually modified to citrullines (Tables II and III). Further, three other homologous sequences were found (Table II) that may represent polymorphisms of this protein, or other mouse type I IRS keratins that perhaps correspond to the additional known sheep proteins for which only limited data have been reported. Several other sequences matched a known type II keratin chain expressed in IRS tissue (41, 42), but we cannot exclude the possibility that other K6-like type II IRS chains might be present in minor amounts, or might possess identical sequences around cross-linking sites. All of the keratin sequences were from end domains, with most from the tail. This may be because there are few Gln and Lys residues in the head domain sequences and because the tails may be more readily accessible for cross-linking. No links were obtained from central rod domain/linker Gln and Lys residues, but we cannot exclude that such linkages occur in minor amounts. This may be because the rod domain Gln and Lys residues are relatively inaccessible because of intra- and/or interchain molecular interactions.

The data suggest that cross-linking of the IRS KIF is man-
Multiple Roles of Trichohyalin

aged differently from the epidermis and in other epithelia studied heretofore. In the case of normal epidermis for example, essentially only the “KSISIS” Lys residue in the head domain of the type II keratin chains is used for KIF cross-linking and at only approximately 10% efficiency (43); only trace amounts of cross-links were found with other head or tail domain Lys and Gln residues (43). However, the KSISIS motif is not present on the IRS K6 protein (42). In case of the IRS KIF in the present study, multiple Lys and Gln residues were used in both the type I and type II chains (Tables II and III). Moreover, there is total approximately 1.3 mol of cross-link/mol for all c29 keratin sequences, arginines were commonly modified to citrullines, and that each Gln and Lys residue was variably used for cross-linking. Thus, a question arises as to the temporal order of these two modification events. Previously, based on in vitro assays with THH-8, representing the terminal 40% of human THH, we suggested that its arginine residues should be modified first, to render it more soluble and accessible for subsequent cross-linking (7). However, it is unknown how and when the KIF end domains are modified, as multiple PAD enzymes and TGases co-exist with the KIF in the IRS cells (7, 22, 45–48).

To further explore this question, we raised in guinea pigs two polyclonal antibodies to the wild-type and citrulline form of the terminal 13 residues of the mouse c29 type I IRS KIF protein, and used them for studies on mouse hair follicle proteins. Both antibodies cross-reacted with the same select bands of monomeric KIF bands extracted from whole mouse hair follicle tissue (Fig. 5). Thus, some of the c29 protein had been modified by PAD enzymes. To confirm this, we performed immunoprecipitation reactions, and then measured citrulline contents of the precipitates. Only the products of the IRScit antibody contained citrulline. In addition, the IRScit antibody also recognized higher molecular weight bands (Fig. 5, lane 3).

By use of indirect immunofluorescence, we found that the wild-type antibody stained the entire IRS (Fig. 6, B and C), beginning near the base of the follicle with the Henle cell layer, but within a few cell lengths, the Huxley and then cuticle cell columns as well. However, the IRScit antibody decoration started many cell lengths from the bottom of IRS KIF expression; stained the Henle, Huxley, and cuticle layers simultaneously; and then ceased a few cell lengths later (Figs. 6 (E and F), 7 (D and F), and 8 (A, C, D, and F)). Neither antibody stained the medulla (Figs. 6–8) or the companion layer (note gap between K6a in Fig. 7F), as predicted based on the expected absence of these type I IRS keratins. Thus, some but not all of the type I c29 protein is modified by the PAD enzymes. To confirm this, by use of dot-blotting methods, the IRScit antibody readily cross-reacted with its IRScit peptide antigen, that is, in which the Arg had been substituted and in which the Lys and Gln residues had not been modified by cross-linking. However, it cross-reacted only weakly with the IRSwt peptide, consistent with the immunofluorescence data. This result also specifies that the IRScit antibody reacts with a subset of the protein of Fig. 5. However, the protein in which the arginines are modified becomes no longer detectable within a few cell lengths along the IRS. To confirm that this loss of epitope was the result of TGase cross-linking, we performed two additional experiments. First, the IRScit antibody did not react on blots with peptides 30, 32, 33, or 55 from Table III in which the arginine residue had been modified to citrullines and nearby Lys and/or Gln residues participated in cross-links. As a control the antibody still recognized the peptide STKVRKTEQZIPS.
Multiple Roles of Trichohyalin

Table II

Identities and similarities of mouse IRS type I keratin tail sequences recovered in this work

Mouse c29 388-SCVKANGQGR PGNQTKDPSK TAIVKTVVEE LDPGRKVLSR VHTLLEKSTK VNNTQUEQIPS

Matched data

TAIVKTVVEE LDPGRKVLSR VHTLLEKSTK VNNTQUEQIPS

Related c29-1

TVEE IDQGRKVLSR IHSIEEKSTK AEQZIPF

Related c29-2

TVLEE VDQSKILTTZ LHSELEK

Related c29-3

ILSTZ LHSELEKSQS N

Second, we explored TGase expression. In confirmation of earlier data (7), the TGase 3 enzyme is expressed only in that part of the IRS above the reactive zone of the IRScit antibody (Fig. 8C). However, the TGase 1 enzyme is present throughout the IRS (Fig. 8E) and co-localizes with the zone of IRScit antibody reactivity (Fig. 8F). These data suggest that cross-linking by the TGase 3 enzyme may account for the loss of the epitopes for the IRScit antibody.

Direct Evidence That THH Functions as a Major Reinforcement Protein of IRS CEs—Mouse IRS tissue was digested to completion with Asp-N peptidase to generate soluble fraction B peptides consisting of the bulk of the KIF/THH cytoplasmic constituents (Fig. 2) (see below). The insoluble residue (recovered by centrifugation at 15,000 × g for 15 min) consisted principally of translucent rigid cellular shells or fragments thereof that appear identical to CEs of cornified epidermal cells (data not shown). However, the precise cell type of origin of these structures is not clear. Based on ultrastructural morphological studies, the IRS cuticle cells do form a barrier structure but it is less clear whether the Henle and Huxley cells do as well (4, 5). However, by earlier immunohistological criteria, all IRS cell layers appear to express involucrin and SPR proteins (49), suggesting that all IRS cell types may build a physical barrier entity. The total amino acid content revealed little (49), suggesting that all IRS cell types may build a physical barrier entity. The total amino acid content revealed little of the IRS above the reactive zone of the IRScit antibody (Fig. 8C). However, the TGase 1 enzyme is present throughout the IRS (Fig. 8E) and co-localizes with the zone of IRScit antibody reactivity (Fig. 8F). These data suggest that cross-linking by the TGase 3 enzyme may account for the loss of the epitopes for the IRScit antibody.

Most of the insoluble residue was digested to completion with proteinase K. 85–95% of the protein material was solubilized, and the peptides were resolved into many peaks by C18 reverse phase HPLC (Fig. 4B). In preliminary bulk analyses, the 35% of peptide mass eluted by <30% acetonitrile contained <10% of the total isopeptide cross-link. Accordingly, we collected 71 peaks resolved by >30% acetonitrile for sequencing. Of these, 63 generated identifiable sequence information, for which 48 contained one cross-link, 15 contained two cross-links, and 4 contained three cross-links, and altogether, presented 157 peptide partners (Tables IV and V). These accounted for 78% of the cross-link content of the fraction C CE fragment pool. Of the total, 34 involved links between THH and various known or novel CE proteins and 41 involved CE-CE links (Tables I, IV, and V). Listed also are an additional seven peptides (10 partner pairs) from fraction B.

Several observations are apparent.

1) The nature of the cross-linked peptides strongly implies that IRS CEs are built the same way as all other CEs we have studied (12, 50–52). They consist of a scaffold of “early” CE proteins such as involucrin, desmoplakin, envoplakin, and SPRs. This in turn is overlaid by specialized reinforcement proteins, which in this case is predominantly THH mixed with a variety of other proteins such as SPRs, repetin, LEP proteins, and epiplakin (below).

2) THH was cross-linked to all other CE proteins, including predicted early CE “scaffold” proteins such as involucrin, envoplakin, and desmoplakin (Table V). The most frequently recovered cross-links were with involucrin, keratin, repetin, SPRs, and LEP. The THH peptides were derived from its entire sequence, except for the predicted EF-hand motifs at the amino terminus or domain 9 at the carboxyl terminus (and see Table VI). Many were from the less-ordered domains 5 and 6 of THH, but few were from the domains 6 or 8 to which KIF were exclusively cross-linked. Notably, peptides 46, 50, 57, 58, 69, and 70 demonstrate that THH serves as a cross-bridging protein between a variety of CE proteins. Moreover, peptides 52, 59, and 62 show that THH serves as a cross-bridging protein between the scaffold/reinforcement aspects of the CE and the KIF/THH cytoskeleton of the IRS cells.

3) Small proline-rich (SPR) protein 1 and 2 families constituted the second principal set of IRS CE proteins, and one or the other was cross-linked to all other CE proteins. Because most members of the SPR2 family share the same amino and carboxyl termini involved in cross-linking (53), we cannot ascertain which specific SPR2 members are expressed in the IRS tissue. However, as found before, none of its central repeating peptide repeats was utilized for cross-linking.

4) LEP proteins (54, 55) could be unambiguously assigned on 10 occasions, because of characteristic Cys residues, and this report is the first direct documentation of their role in a CE structure. However, as the short peptide fragments recovered after proteolysis are either identical or almost identical to those of SPR1 proteins, it is possible some were misassigned, as was likely in previous studies (12, 50). Nevertheless, based on their cross-linking partners (Tables IV and V), THH, LEP (see, for example, peptides 48, 50, 53 and 61), and SPR (many peptides) all appear to function as cross-bridging proteins, either among themselves and/or between the several other CE proteins. As LEP and SPR proteins are abundantly expressed in IRS tissue (50–53), we thus predict that they together with THH serve coincident and perhaps complementary roles as reinforcement proteins in IRS CEs.

5) Repetin sequences were recovered in 11 cross-links, in an apparent cross-bridging role also (see peptides 51, 56, 60, and 63 of Table IV), and so this study confirms for the first time the
| Peptides from Fraction A | Type I c29 | Type I c29 |
|-------------------------|------------|------------|
| 1. (k) CQSN             | Type I c29-3 (Q-3) | Type I c29-3 (Q-3)* |
|                         | THNH K1141 | THNH K977  |
| 2. (k) GSSLTPSKX        | Type II K6IR5 K420 | Type II K6IR5 K420 |
|                         | THNH Q1714 | THNH Q1714 |
| 3. SQPTCK               | Type II K6IR5 Q4 | Type II K6IR5 Q4 |
|                         | THNH K1161 | THNH K1047 |
| 4. SQPTCK               | Type II K6IR5 Q4 | Type II K6IR5 Q4 |
|                         | THNH K1161 | THNH K1047 |
| 5. (k) STKVKY            | Type I c29 K438 | Type I c29 K438 |
|                         | THNH Q1277 | THNH Q1277 |
| 6. (k) VNTQZQP5          | Type I c29 K441 | Type I c29 K441 |
|                         | THNH Q852 | THNH Q852 |
| 7. (k) ACRQYF            | Type I c29-1 (Q-5) | Type I c29-1 (Q-5) |
|                         | THNH K1171 | THNH K1171 |
| 8. (k) STKVKY            | Type I c29 K438 | Type I c29 K438 |
|                         | THNH Q1038 | THNH Q1038 |
| 9. (k) VNTQZQP5          | Type I c29 K441 | Type I c29 K441 |
|                         | THNH Q907 | THNH Q907 |
| 10. (k) STKVKY           | Type I c29 K438 | Type I c29 K438 |
|                         | THNH Q1028 | THNH Q1028 |
| 11. (k) TQZQP5           | Type I c29 Q444 | Type I c29 Q444 |
|                         | THNH X345 | THNH X345 |
| 12. (k) TQZQP5           | Type I c29 K441 | Type I c29 K441 |
|                         | THNH Q1206 | THNH Q1206 |
| 13. (k) TQZQP5           | Type I c29 Q444 | Type I c29 Q444 |
|                         | THNH K979 | THNH K979 |
| 14. (k) GSSLTPSKX        | Type II K6IR5 K420 | Type II K6IR5 K420 |
|                         | THNH Q916 | THNH Q916 |
| 15. SLQFENGSHVNL       | Type II K6IR5 K420 | Type II K6IR5 K420 |
|                         | THNH K1003 | THNH K1003 |
| 16. (k) GSSLTPSKX        | Type II K6IR5 K420 | Type II K6IR5 K420 |
|                         | THNH Q916 | THNH Q916 |
| 17. SQPTCKEANR           | Type II K6IR5 Q4 | Type II K6IR5 Q4 |
|                         | THNH K972 | THNH K972 |
| 18. (k) TQZQP5           | Type I c29 Q444 | Type I c29 Q444 |
|                         | THNH K1263 | THNH K1263 |
| 19. (k) VNTQZQP5         | Type I c29-1 (Q-26) | Type I c29-1 (Q-26) |
|                         | THNH K1117 | THNH K1117 |

* In the cases of the type I c29 variants, the Gln/Lys residue number is identified from the carboxyl terminus shown in Table II.
involvement of this protein in a CE barrier structure (56). The new data document that repetin likely serves as a complete substrate; multiple adjacent different glutamines and lysines were used for cross-linking in vivo. Repetin, like involucrin, also consists of a tandem array of peptide repeats that are unique to it, but are more conserved than those of THH. However, unlike those of SPRs but more similar to those of involucrin, repetin repeats were extensively used for cross-linking. Thus, even though repetin, THH, and SPRs all seem to function as cross-bridging proteins, there are important differences in the way in which they may function as cross-bridgers.

6) Desmoplakin and envoplakin cross-links were abundant. Only two Lys and one Gln residue in desmoplakin and one Gln residue in envoplakin were used, as seen previously. These residues are located late in the “C plakin” domain region and are thereby suggestive of the high degree of specificity with which these proteins are used in the early stages of CE assembly (50).

7) Epiplakin, a new member of the “plakin” family of cytolinkers (57), was found in seven cross-links (Table V). These will be discussed in more detail below.

8) Finally, cross-links involving other known CE proteins including members of the S100 family of calcium-binding proteins, elafin, etc., were not found; this could be because they are not abundant components of IRS CEs.

**THH Is Extensively Cross-linked to Itself in IRS Tissue**—Fraction B peptides were not well resolved by several HPLC methods used because of the presence of a very large number of different species (Fig. 4). Because there is no evidence for polymorphism of mouse THH, the most likely explanation is that the PAD and TGase modifications of THH are complex, leading to a bewildering number of possible peptides following cleavage. Nevertheless, we were able to recover 77 peptides that contained one (62 peptides), two (11 peptides), or three (4 peptides) cross-links that contained only THH sequences linked together (Table I). Altogether in one experiment, these represented approximately 40% of the total cross-link and citrulline content of Asp-N-soluble IRS peptides. The large unresolved region late in the chromatogram contained an additional 53% of cross-link (one per 19 residues) and 48% of citrulline (ratio of 0.8).

Table VI summarises the domain origins of all THH sequences recovered in this work. The data reveal domain-specific partner linkages, which suggests modification preferences.
Antibodies used were as follows: A, THH; B, K6; C, both; D, IRScit; E, K6; F, both. Bar = 50 μm.

First, no links were found with the domain 1 EF-hand portion of THH. It is possible these motifs are removed during terminal differentiation, as occurs for the related protein, profilaggrin (58). All THH-keratin links involved only domain 6 or 8 sequences. THH-THH and THH-CE protein links were distributed among domains 2–5, but were rare in domains 6 and 8. Most intra-THH cross-links occurred in the least organized domain 5 region at a 3.5-fold higher frequency (Table VI).

A Possible Explanation as to Why KIF Cross-link Preferentially Only to Domains 6 and 8—As for human THH, domains 6 and 8 represent the most orderly almost perfectly α-helical portion of mouse THH. In addition to regular sequence repeats, both domains exhibit a regular quasi-repeat of 20 residues in the distribution of their charged residues, which is equivalent to an axial rise of −2.97 nm. The coiled-coil α-helical 1B and 2 rod domain segments of the c29 and K6IRS KIF chains, like all types I/II KIF chains (44), display a 9.8-residue repeat of charged residues, which corresponds to 1.46 nm of axial rise along the KIF. It is therefore likely the rod domains of the KIF and domains 6 and 8 of THH will engage in favorable ionic interactions, as the two periodicities will frequently “beat” together. This will not occur in a regular way with any other domain of THH. These favorable interactions will impact modification events by the PAD and TGase enzymes. First, as the tail (and possibly beginning of the head) domains of the KIF chains project laterally out from the wall of the KIF (44), they will be more accessible for modification events. Second, because of their close proximity, we predict the KIF end domains will be preferentially cross-linked to domains 6 and 8 to THH than any to other domain of THH. Third, cross-linking of THH to KIF rod domains is unlikely to occur because they are tightly packed together. Fourth, we predict that prior cross-linking of the KIF end domains to THH domains 6 and 8 will preclude their subsequent utilization for cross-linking to THH itself or other CE proteins.

Cross-links Involving the Novel “Plakin” Epiplakin—Five cross-links involving epiplakin in the fraction C peptides and two more in fraction B were recovered (Table IV). Human epiplakin consists largely of 13 highly conserved B or B-like plakin domains interspersed by linker regions (57). Our new sequence data reveal that the human and mouse species have been highly conserved at least around cross-linking sites. Three neighboring Gln residues were found that are repeated in its several B domains. All epiplakin links involve lysine residues from either THH or a keratin chain or both. As expected these peptides were derived from the intracellular milieu of the IRS tissue. Notably, peptide 71 suggests that epiplakin can act as a cross-bridge between the KIF proteins. Together, the data suggest that the role of epiplakin role in CE barrier formation and/or the cytoskeleton is different than for desmoplakin and envoplakin.

Summary—Our new data support the following timeline of events in the IRS. First, THH and KIF are coincidentally expressed from the lowermost cell layers of the IRS in the hair follicle (Fig. 6, A–C; Refs. 1, 4, and 5). Second, TGase 1 is expressed in all layers of the IRS starting also from the lowermost cell layers coincident with THH and KIF expression (Fig. 8, E and F), but the TGase 3 enzyme is expressed only in higher cell layers in the follicle (Fig. 8, B and C). Third, compilation of existing data suggests that the PAD enzymes (45) are expressed well before the TGase 3 enzyme (Fig. 8, B and C). Fourth, it is therefore likely that modification of the arginines of the KIF chain end domains to citrullines occurs before a significant degree of cross-linking (Fig. 5). The reason why only some of the KIF chains are modified may be the result of accessibility limits imposed once some cross-linking has occurred. Fifth, very shortly later, only the domains 6 and 8 of THH are used for cross-linking to KIF end domains (Tables II and VI), possibly because of coincident regular distributions of ionic charges efficiently juxtapose the two proteins. Sixth, this
Peptides from fraction C
1. THH domain 5 several K/desmoplakin* (Q)
2. Involucrin Q197/involucrin K383
3. THH domain 5 several/SPR2 (K-3)*
4. Type I e 29 Q444/SPR2 (K end)
5. THH domain 5 K470/involucrin Q424
6. Epiplakin** several Q (TGQQI) /involucrin K253
7. THH domain 5 Q973/type I c29–1 (K-8)
8. THH domain 5 K470/LEP Q9
9. Involucrin (Q22)/SPR2 (K-3)
10. SPR2 (Q-8)/SPR2 (K-3)
11. THH domain 2 K209/involucrin Q295
12. THH domain 2 K138/SPR2 Q6
13. Desmoplakin (Q)/SPR1 (end K-3)
14. Enoplakin Q197/involucrin K383
15. Involucrin Q106/SPR2 (K end)
16. THH domains 3 + 4 several Q/SPR2 (K end)
17. THH domain 5* K658/LEP Q6
18. THH domain 3 several/THH domain 8 K1067
19. THH domain 5 K396/repetin several K (QGQKGRQ)
20. Involucrin Q22/desmoplakin (KTG)
21. Enoplakin Q197/SPR1 (K end)
22. Involucrin Q560/SPR2 (K end)
23. SPR2 Q6/SPR2 (K end)
24. THH domain 8 K1311/type I c29 Q433
25. THH domain 5 K548/epiplakin several Q (EAQIA)
26. Involucrin Q347/desmoplakin (KTK)
27. SPR1 Q6/SPR1 (K end)
28. THH domain 5 K396/LEP Q5
29. Involucrin Q398/SPR2 K8
30. Enoplakin Q197/type I c29 K438
31. Enoplakin Q197/type I c29 K441
32. SPR2 Q6/involucrin K206
33. SPR2 Q7/SPR1 (K end)
34. Type II K6IRS K7/epiplakin several Q (TGQQI)
35. Involucrin K206/LEP Q9
36. SPR1 Q17/SPR1 (K-3)
37. Repetin Q900/SPR2 (K-3)
38. SPR1 Q19/SPR2 (K-3)
39. SPR1 R21/epiplakin several Q (EAQAA)
40. THH domain 5 Q441/THH domain 2 K180
41. Repetin Q453/SPR1 (K-3)
42. SPR2 Q6/SPR1 (K-3)
43. THH domain 3 several Q/type I c29–1 (K-23)
44. Involucrin Q85/repetin several K (QGQKGRQ)
45. Involucrin K234/THH domain 5 Q550, K548/SPR1 (Q-1)
46. THH domain 2 Q170/THH domain 5 K548, Q550/repetin several K (QGQKGRQ)
47. THH domain 2 K138/SPR1 (K end, Q6) /involucrin Q250
48. THH domain 5 K548/LEP Q9, Q5/involucrin K320
49. THH domain 5 Q several/involucrin K255, Q198/involucrin K393
50. SPR2 (K end)/LEP Q5, Q9/THH domain 5 K555, K548/SPR1 (Q-1)
51. THH domain 2 K187/repetin K612, Q609/SPR1 Q6
52. Enoplakin Q197/THH domain 8 K1061, Q1066/type I c29–3 (K-4)
53. THH domain 5 K548/LEP Q9, Q5/THH domain 5 K396
54. THH domain 2 K187/involucrin Q398, K393/epiplakin several Q (EAQAA)
55. Enoplakin Q197/involucrin K186, Q190/repetin several K (QGQKGRQ)
56. Involucrin Q347/THH domain 5 Q550, K548/SPR1 (Q-1)
57. SPR2 Q6/THH domain 2 K209, Q213/SPR1 (K-3)
58. SPR2 (K end)/THH domain 5 Q550, K548/SPR1 (Q-1)
59. Type II K6IRS K7/THH domain 5 Q1008, K1002/involucrin K393
60. THH domain 2 Q126/repetin K370, Q373/involucrin Q106
61. THH domain 5 K548/LEP Q5, Q9/SPR1 (K-3, Q-1)/involucrin K205
62. Involucrin Q250/SPR1 (K end, Q19)/THH domain 8 K972, Q977/type I c29–3 (K-4)
63. THH domain 3 Q several/repetin K several, Q several (QGQKGRQDQSSH) /SPR1 (K end, Q6)/SPR1 (K-3)

Peptides from fraction B
64. THH domain 2 K138/epiplakin several Q (TGQQI)
65. THH domain 5* Q622/involucrin K214
66. THH domain 5 K514/LEP Q9
67. THH domain 3 several/SPR2 (K end)
68. THH domain 5 Q383/SPR2 (K end, Q-8)/THH domain 5 K555
69. SPR1 (K end)/THH domain 3 Q several, Q several/repetin K612
70. Involucrin K104/THH domain 3 Q several, Q several/involucrin K234
71. Type I e 29 K441/epiplakin Q (TGQQI), epiplakin Q/type I c29–1 K-23

** Sequences for mouse desmoplakin and epiplakin are not known; the assignments here are based on almost identical matches with human sequences. For desmoplakin, all sites involved the same Q (LQDTSSY) or nearby K (ILTCPKTKLK) residues late in its C plakin domain. For epiplakin, three different Q residues were employed (EAQAA, TGQQI, and EAQIA) in its B plakin domains.

* Both SPR1 and SPR2 mouse proteins consist of multiple members of near identical amino and carboxyl termini but with varying numbers of internal repeats; thus, the terminal residues were numbered from the common end.

\* Both epiplakin and repetin consist of many near-identical peptide repeats. Thus, it was not possible to ascertain precisely which residues were used in the cross-links.
specific cross-linking coincides with the expression of the TGase 3 enzyme (Fig. 8, B and C), which suggests TGase 3, rather than TGase 1, is responsible for THH-KIF cross-linking. Indeed, the TGase 3 enzyme exhibits marked preference for THH rather than TGase 1, is responsible for THH-KIF cross-linking. This in turn is essential for appropriate morphogenesis of hair follicle cortical cells internal to the IRS as a more favorable reaction in living cells than disulfide bond formation. Further, proteins made insoluble by isopeptide cross-linking are more easily degraded than are disulfide-bonded proteins, as indeed is required for the IRS, which is degraded within the follicle canal (4, 5). As suggested by Tobin et al. (60), "it is likely that [a lysosomal cysteine protease, cathepsin L] CTSL may . . . use trichohyalin as a protein substrate." Finally, we submit that THH serves a similar biomechanical role in other toughened epithelial tissues such as the hard palate, filiform ridges of the tongue, and rodent forestomach.

REFERENCES

1. Rothnagel, J. A. & Rogers, G. E. (1986) J. Cell Biol. 102, 1419–1429
2. Lee, S.-C., Kim, I.-G., Marekov, L. N., O’Keefe, E. J., Parry, D. A. D. & Steinert, P. M. (1990) J. Biol. Chem. 265, 12164–12172
3. Feitz, M. J., McLaughlin, C. J., Campbell, M. T. & Rogers, G. E. (1993) J. Cell Biol. 121, 855–865
4. Fornier, H. (1995) Dermal Derm. Z. (Berlin) 10, 357–376
5. Auber, L. (1995) Trans. R. Soc. Edinb. 62, 191–254
6. Hamilton, E. H., Payne, J., Richard, E. & O’Keefe, E. J. (1992) J. Invest. Dermatol. 96, 666–672
7. Tarcea, E., Marekov, L. N., Andreoli, J., Idler, W. W., Candi, E., Chung, S.-I. & Steinert, P. M. (1997) J. Biol. Chem. 272, 700–705
8. O’Guin, M. W., Sun, T.-T. & Manabe, M. (1992–J. Invest. Dermatol. 98, 24–32
9. O’Keefe, E. J., Keefe, E. J., Hamilton, E., Lee, S.-C. & Steinert, P. M. (1993) J. Invest. Dermatol. 100, 655–719
10. Manabe, M. & O’Guin, M. W. (1994) Differentiation 58, 65–75
11. Steinitz, P. M., Kartasova, T. & Marekov, L. N. (1998) J. Biol. Chem. 273, 11758–11769
12. Rogers, G. E. (1962) Nature 194, 1149
13. Steinitz, P. M., Dyer, P. Y. & Rogers, G. E. (1969) J. Invest. Dermatol. 56, 49–54
14. Harding, H. W. & Rogers, G. E. (1976) Biochem. Biophys. Acta 427, 315–324
15. Rogers, G. E. (1963) J. Histochem. Cytochem. 11, 791–795
16. Takahara, H., Oikawa, Y. & Sugawara, K. (1983) J. Biochem. (Tokyo) 94, 1945–1953
17. Takahara, H., Tsuchida, M., Kuwabata, M., Akutsu, K., Tagami, S. & Sugawara, K. (1989) J. Biol. Chem. 264, 13361–13366
18. Watanabe, K., Akiyama, K., Hikibe, K., Ohsuka, R., Okuyama, A. & Senshu, T. (1988) Biochem. Biophys. Acta 966, 375–383
19. Kanno, T., Kawada, A., Yamanouchi, J., Yosida-Noro, C., Yoshiki, A., Shigeta, T., Satoh, M. & Takahara, H. (1991) J. Biochem. (Tokyo) 110, 661–666

Multiple Roles of Trichohyalin

TABLE V

Frequency of cross-links between protein partners in mouse IRS CEs

| No. of partners | Desmop | Envop | Epip | Invol | Keratin | Repetin | SPR1 | SPR2 | LEP | THH |
|----------------|--------|-------|------|-------|---------|---------|------|------|-----|-----|
| Desmop (4)     | 0      |       |      |       |         |         |      |      |     |     |
| Envop (6)      | 0      | 0     |      |       |         |         |      |      |     |     |
| Epip (7)       | 33     | 2     | 2    | 3     | 2       | 0       | 0    | 0    | 0   |     |
| Invol (10)     | 0      | 1     | 2    | 0     | 0       |         |      |      |     |     |
| Keratin (15)   | 0      | 0     | 0    | 4     | 0       |         |      |      |     |     |
| Repetin (24)   | 1      | 2     | 0    | 4     | 0       | 2       | 3    |      |     |     |
| SPR1 (23)      | 0      | 0     | 0    | 5     | 1       | 2       | 3    | 2    |     |     |
| SPR2 (11)      | 0      | 0     | 0    | 2     | 0       | 1       | 1    | 0    |     |     |
| LEP (53)       | 1      | 1     | 2    | 9     | 6       | 7       | 8    | 9    | 7   | 3   |

* Desmop, desmoplakin; envop, envoplakin; epip, epiplakin; invol, involucrin.

TABLE VI

Sequences involved in cross-links demonstrate domain-specific preferences of partners

| Domain and residues/molecule (%) | 1 | 2 | 3 + 4 | 5 | 5a | 6 | 8 | 9 |
|----------------------------------|---|---|------|---|----|---|---|---|
| THH-THH (179 partners)          | 0 | 22| 17   | 94| 37 | 2 | 7 | 0 |
| THH-keratin (81)                | 0 | 0 | 0    | 0 | 12 | 6 | 2 | 0 |
| THH-CE (44)                     | 0 | 8 | 9    | 22| 22 | 0 | 4 | 0 |
