High-Molecular-Weight Precursor of Epidermal Filaggrin and Hypothesis for Its Tandem Repeating Structure†

John D. Lonsdale-Eccles,‡ Katheryn A. Resing, Rick L. Meek, and Beverly A. Dale*

ABSTRACT: Filaggrin is a histidine-rich protein that is intimately involved in mammalian epidermal keratinization. Using a combination of immunologic and in vivo pulse–chase studies with radiolabeled histidine and phosphate, we show that the phosphorylated precursor of both rat and mouse filaggrin has an apparent molecular weight much higher than previously realized (6×10^5 and 3.9×10^5, respectively). These high-molecular-weight filaggrin precursors can be rapidly labeled with histidine and extracted from the epidermis under denaturing conditions. More than half of the label incorporated in the precursor at 2 h is found in filaggrin at 24 h after injection, even though filaggrin is less than 10% of the size of the precursor. Limited proteolytic digestion of the precursor in vitro results in the formation of an oligomeric series of peptides based on a phosphorylated fragment slightly larger than filaggrin itself. More extensive digestion of this fragment shows that it is composed of filaggrin with few or no additional unrelated peptides, suggesting that the major part of the high-molecular-weight filaggrin precursor must be composed of repeated domains of filaggrin. Because the primary translation product of filaggrin mRNA is large, we propose that these domains are repeated in tandem. In addition, from molecular weight computations and peptide map analyses, we suggest that the filaggrins are themselves composed of multiple repeating units of an unidentified peptide of approximately M_r 8600. This value is derived from the molecular weights of filaggrin from several mammalian species that differ by integral multiples of 8600. A model for the structure of the high-molecular-weight precursor of filaggrin is presented. It has two types of repeating untis: those that make up the filaggrin molecule itself and the tandem repeated copies of filaggrin.

Filaggrin is a protein isolated from the stratum corneum of skin (Dale, 1977; Ball et al., 1978; Steinert et al., 1981). It aggregates with epidermal keratin filaments and apparently functions as the keratin matrix in the cornified cells (Dale et al., 1978; Steinert et al., 1981; Lynley & Dale, 1983). Pulse–chase studies have shown that filaggrin is derived from a precursor located in extracts of keratohyalin granules (Dale & Ling, 1979), but unlike filaggrin, the precursor is highly phosphorylated and does not aggregate with keratin filaments (Lonsdale-Eccles et al., 1980, 1982). Filaggrin and its related keratohyalin-derived protein contain 5–10 times more histidine than most other proteins and consequently have been given the generic name "histidine-rich" proteins (HRPs) (Hoober & Bernstein, 1966; Ugel, 1969; Balmain et al., 1977; Murozuka et al., 1979). The name filaggrin is based on the function of the protein (Steinert et al., 1981).

Nomenclature notwithstanding, the epidermal histidine-rich proteins have proved to be an interesting class of proteins that are associated with the differentiation of epidermis, in particular with the presence of keratohyalin granules (Ugel, 1971; Sibrack et al., 1974; Fukuyama & Epstein, 1975; Balmain, 1977; Ball et al., 1978). The amino acid compositions of these proteins from several species are similar in their overall character; they are very polar molecules and have little hydrophobic nature. The molecular weight of these proteins differs between species (Lonsdale-Eccles et al., 1980; Steinert et al., 1981; Scott & Harding, 1981; Lynley & Dale, 1983). However, the values reported for the molecular weight of the precursor protein differ from tens of thousands to several hundred thousands, even within a single species (Ball et al., 1978; Dale & Ling, 1979; Scott & Harding, 1981; Lonsdale-Eccles et al., 1982). This has caused problems in the analysis of the filaggrin pathway and in the interpretation of posttranslational modification. The variation in the size of the precursor from different laboratories could be due to aggregation of the putative 54-kDa phosphorylated precursor (rat; Dale & Ling, 1979; Lonsdale-Eccles et al., 1980, 1982), or it may be the result of instability of a larger precursor as suggested by Scott & Harding (1981). Recent evidence on the size of the messenger RNA supports the latter interpretation (Meek et al., 1983). R. L. Meet et al. have isolated very large mRNA (30–40 S) from rat and mouse epidermis that codes for filaggrin-related proteins with $M_r > 200\,000$ when translated in a cell-free protein synthesis system.

Filaggrin and its precursor can be identified by their reaction with specific antiserum, as well as by their incorporation in vivo of radiolabeled histidine (and phosphate, in the case of the precursor). In this paper, we have used each of these methods, combined with simple and rapid extraction procedures, to demonstrate the presence of precursors of rat and mouse filaggrin(s) that have $M_r \geq 300\,000$. Furthermore, by peptide mapping studies, we show that these high-molecular-weight precursors are composed of multiple units of filaggrin-like protein. We hypothesize that the epidermal histidine-rich proteins are composed of tandem repeating domains.

Materials and Methods

[^{3}H]Histidine was injected subcutaneously into newborn (0–24 h) Sprague-Dawley rats (20 μCi/animal) and newborn Balb/c mice (10 μCi/animal). Animals to be labeled for more than 2.5 h were given a chase injection containing a 100-fold excess of unlabeled histidine. At timed intervals after the injection of radioactivity, animals were killed by cervical dislocation, and the skin of each animal, except that covering

† From the Departments of Periodontics (J.D.L.-E. and B.A.D.), Biochemistry (K.A.R. and R.L.M.), Oral Biology (B.A.D.), and Medicine (B.A.D.), University of Washington, Seattle, Washington 98195. *Received February 28, 1983.* This study was supported by USPHS Grants DE04660 and DE02600 (Center for Research in Oral Biology) to B.A.D., GM07317 to R.L.M., and GM15751 to Kenneth Walsh supporting K.A.R.

‡ Present address: ILRAD, P.O. Box 30709, Nairobi, Kenya.

0006-2960/84/0423-1239$01.50/0

the limbs, tail, and head, was removed by dissection. The dermis and epidermis were separated after incubation of the skin in phosphate-buffered saline (PBS) containing 10 mM ethylenediaminetetraacetic acid (EDTA) for 3 min at 50 °C as previously described (Lynley & Dale, 1983). Epidermal proteins (from three skins per time point) were extracted with 8 M urea containing 50 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) (pH 7.6) and 20 μg/mL phenylmethanesulfonyl fluoride (PMSF) (1.5 mL/rat epidermis; 0.75 mL/mouse epidermis). The extracts were centrifuged at 27000g (4 °C; 40 min) to remove particulate matter, and the supernatants were stored at −20 °C. All procedures were performed as rapidly as possible in order to minimize proteolytic breakdown of the proteins. For the study of the incorporation of phosphate into epidermal proteins, sodium [^{32}P]phosphate in 0.9% (w/v) NaCl was injected into newborn animals. The procedure was essentially the same as above, except that no chase was used and the labeling period was 6 h; 50 μCi was injected into each mouse and 200 μCi into each rat.

Epidermal extracts were analyzed by electrophoresis on sodium dodecyl sulfate ($NaDodSO_4$)–polyacrylamide gradient slab gels (5–10% polyacrylamide) with the discontinuous system described by Laemmli (1970). Proteins were identified by staining with Coomassie brilliant blue R250 [0.05% w/v in 2-propanol–acetic acid–water (25:10:65)]. Fluorography using En^{3}hance (New England Nuclear, Boston, MA) was employed to identify radiolabeled proteins. To identify proteins that cross-react with antibody to rat filaggrin, duplicate samples of the epidermal extracts were subjected to electrophoresis. Half of the gel was stained as usual. The proteins on the other half were transferred electrophoretically to a nitrocellulose membrane (overnight at 60 V in 0.025 M Tris–0.325 M glycine buffer with 20% methanol). The nitrocellulose blots were incubated for 2 h in 3% bovine serum albumin (BSA) in buffer to block additional protein binding sites, then in rabbit antiserum to rat filaggrin (1/200 to 1/2000 dilution), then with goat anti-rabbit IgG, and then with rabbit peroxidase–antiperoxidase complex made in the rabbit (Miles-Yeda, Inc., Elkhart, IN; 1/80 dilution). Incubations were done on a rocker platform at room temperature with buffer washes between each step. Color was developed by incubation in freshly prepared substrate solution containing 50 mM Tris-HCl, pH 7.6, 3,3′-diaminobenzidine hydrochloride (0.05 mg/mL), and 0.01% hydrogen peroxide at room temperature for 30 s to 3 min. In order to get similar quantitative results, the mouse extracts had to be incubated with 10-fold greater amounts of antibody than the rat extracts.

Molecular weights of the various proteins were determined from $NaDodSO_4$–polyacrylamide gradient slab gels (5–10% and 5–12%). Values for standard proteins were plotted on log linear graphs. Molecular weight markers used were (1) aprotinin (6500), (2) cytochrome *c* (12300), (3) β-lactoglobulin (18400), (4) chymotrypsin(ogen) (25500), (5) DNase I (31000), (6) ovalbumin (43000), (7) pyruvate kinase (57000), (8) bovine serum albumin (68000), (9) phosphorylase (93000), (10) β-galactosidase (116250), (11) fibronectin (220000), and (12) unreduced cross-linked fibrinogen (340000). In some experiments, high-M_r standard proteins from Pharmacia Fine Chemicals, Piscataway, NJ, were used.

Peptide maps of filaggrin and the high-M_r precursor were obtained by proteolytic digestion in the stacking gel of an $NaDodSO_4$–polyacrylamide gel, as described by Cleveland et al. (1977). The proteins to be digested were cut from an $NaDodSO_4$–polyacrylamide gel that had been quickly (within 45 min) stained and destained. Each gel slice contained approximately 10–20 μg of protein. The electrophoresis was interrupted for 30 min just when the bromophenol blue marker was about to enter the resolving gel. Upon completion of electrophoresis, the gels were stained and destained as usual.

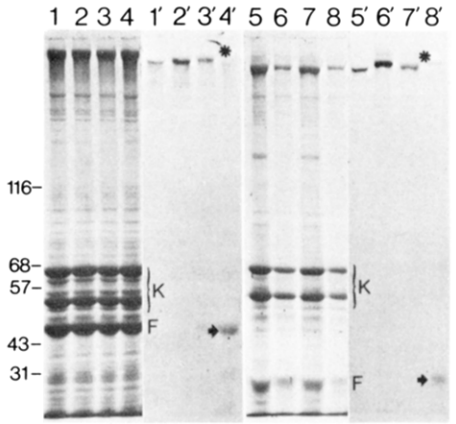

FIGURE 1: Time course of [^{3}H]histidine incorporation into epidermal proteins extractable in 8 M urea. Unlabeled histidine (100-fold molar excess) was injected 2 h after initial injection of labeled histidine. At various times after the initial injection into newborn rats and mice, the animals were killed and their epidermal proteins extracted and analyzed by electrophoresis on a 5–10% polyacrylamide gradient $NaDodSO_4$ gel and fluorography. (Lanes 1–4) Rat epidermal proteins at 0.5, 2, 6, 24 h after injection; (lanes 5–8) similar samples from mouse epidermis. The radiolabel is initially incorporated into proteins of high M_r (asterisk). By 24 h, the radiolabel is also found in filaggrin (F and arrow). Keratins (K) are poorly labeled. Positions of molecular weight markers ($\times 10^{-3}$) are indicated on the right.

Results

Radiolabel Incorporation into Epidermal Proteins. Pulse–chase studies using radiolabeled histidine were performed to identify precursors of filaggrin, a histidine-rich protein. All extractions in this study were performed rapidly and under nonreducing conditions, because the histidine-rich proteins of high molecular weight, especially those of rat, are very susceptible to breakdown upon freezing and thawing or in the presence of β-mercaptoethanol (J. D. Lonsdale-Eccles, unpublished observations) and upon dialysis (Scott & Harding, 1981). The fluorographs (lanes 1′–8′) and stained $NaDodSO_4$–polyacrylamide gels (lane 1–8) of pulse–chase experiments conducted in newborn rats (lanes 1–4) and mice (lanes 5–8) are shown in Figure 1. The predominant stained proteins in these gels are the keratins and filaggrins. Under the conditions used in these experiments, little or no radiolabel is observed in the keratins. The initial incorporation of radiolabeled histidine (within 30 min) is into proteins that have a very high molecular weight (Figure 1, lanes 1 and 5). At 2 and 6 h, radiolabel is still detectable only in high-M_r proteins; however, by 24 h, the radiolabel is found in filaggrin (rat M_r 49000, lane 4; mouse, M_r 26500, lane 8). This is consistent with a precursor–product relationship between the high-M_r proteins and filaggrin and indicates that processing of the precursor occurs between 6 and 24 h after its synthesis.

A separate experiment in which eight rats were injected with 50 μCi of [^{3}H]histidine is shown in Figure 2. The average amount of label incorporated after 2 h (96600 ± 24300 cpm/mg of wet wt tissue) was very similar to that incorporated after a 2-h pulse and a 22-h chase (104400 ± 19000 cpm/mg of wet wt tissue). Proteins were resolved by electrophoresis on 4–15% gradient gels and analyzed by scintillation counting of unstained gel slices, as well as staining and fluorography. The proportion of total incorporation in the high-M_r band was

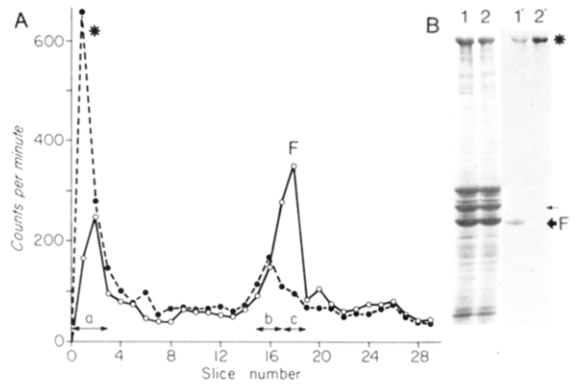

FIGURE 2: Pulse–chase study for quantitative evaluation of filaggrin labeling. Newborn rats (less than 12 h old) were injected with 50 μCi of [^{3}H]histidine (29 ci/mmol) in 50 μL of 0.01 N HCl with 0.9% NaCl. Four animals were used for extraction of epidermal proteins labeled for 2 h or 2 h with a 22-h chase . The latter group of animals was returned to the mother immediately and given a chase injection of 100-fold excess of unlabeled histidine after 2 h. Epidermal protein samples were prepared as described under Materials and Methods, except that the epidermis from each animal was separately extracted in 8 M urea–0.05 M Tris-HCl, with 0.1 M β-mercaptoethanol 20 μg/mL PMSF, 1 mM *o*-phenanthroline, and 10 mM EDTA. For the 2-h samples, an average of 4.6 ± 0.33 mg of wet wt tissue was extracted in 1 mL of buffer, yielding 2395 ± 587 cpm/5 μL. For the samples obtained 24 h after the initial injection, 5.2 ± 0.30 mg of wet wt of epidermis was extracted per 1 mL, yielding 2548 ± 735 cpm/5 μL. Proteins in identical volumes of extract were resolved by electrophoresis on 4–15% polyacrylamide gradient NaDodSO$_4$ gels and either stained or sliced for scintillation counting after being dissolved in protosol. (A) Results of counts per minute vs. slice number are shown for the combined samples obtained for 2 (●) or 24 h (○) after injection. Slice no. 1 was the origin of the gradient gel. The amount of label and the specific activity of the two combined samples agree within 10%. Percentage of total counts in the high-molecular-weight peak (slices 0–3) is 48% and 20% for 2- and 24-h samples, respectively, while that in the filaggrin peak (slices 17–19) is 6 and 27%, respectively. Approximately 10% of the label in each extract is in slices 15–17 in a band that migrates more slowly than filaggrin (see part B). (B) Stained gel and fluorograph showing that the high-molecular-weight band (*) labeled 2 h after injection corresponds to a heavily staining band in the gel and the band labeled in the 24-h sample corresponds to rat filaggrin (F, approximately 49 kDa). Both extracts also have a weakly labeled protein band (←) of approximately 55 kDa that corresponds to a keratin band on the stained gel. (Lane 1) Extract obtained 24 h after injection; (lane 2) extract obtained 2 h after injection, stained for protein; (lanes 1′ and 2′) fluorograph of some lanes.

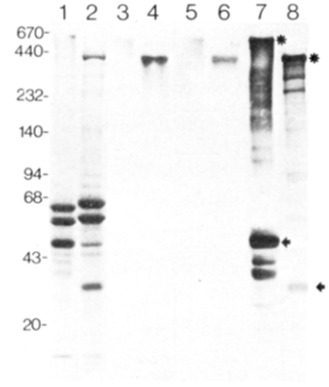

FIGURE 3: Immunologic identification of filaggrin-related proteins in newborn rat and mouse epidermal extracts labeled with [^{3}H]histidine or [^{32}P]phosphate. Proteins from 8 M urea epidermal extracts were separated by electrophoresis on 3–10% polyacrylamide gradient NaDodSO$_4$–urea gels and analyzed by fluorography or electrophoretically transferred to nitrocellulose, reacted with antibody to rat filaggrin, and then with peroxidase–antiperoxidase, as described under Materials and Methods. Lanes 1, 3, 5, and 7 are from rat, and lanes 2, 4, 6, and 8 are from mouse. Lanes 1 and 2 are stained for protein; lanes 3 and 4 show [^{3}H]histidine-labeled protein 2 h postinjection; lanes 5 and 6 show ^{32}P-labeled protein 4 H postinjection; lanes 7 and 8 show filaggrin-immunoreactive proteins. The numbers on the left are the molecular weight markers (×10^{-3}). The high-M_r cross-reactive proteins (*) correspond to the histidine- and phosphate- labeled bands. Filaggrin of both rat and mouse (arrows) reacts with the antibody, but these bands are not labeled with phosphate or with histidine in the sample shown. Immunoblot controls with preimmune serum from the same rabbit show little or no reaction under these conditions.

48% at 2 h and decreased to 20% at 24 h after the initial injection, while the percentage of incorporation in the filaggrin peak increased from less than 6 to 27%. Approximately 10% of the total label at each time was in a band migrating slightly slower than filaggrin, consistent with a keratin of 55 kDa.

Densitometry of the fluorograph of the mouse epidermal proteins in Figure 1 shows a similar trend. At 2 h after injection, 75% of the label incorporated is associated with the high-M_r band, and none is in the filaggrin band. At 24 h after the initial injection, 16% remains associated with the high-M_r band, but 37% is in the filaggrin band. Thus for both rat and mouse, the amount of label in filaggrin at 24 h is approximately 50% of that seen in the high-M_r precursor at 2 h, even though the high-M_r band is more than 10 times larger than filaggrin in each species.

Our earlier studies on a putative precursor of rat filaggrin (M_r 54 000) showed that the precursor was a highly phosphorylated protein (Lonsdale-Eccles et al., 1980). This is also true of the high-M_r proteins described in this paper. After subcutaneous injection of radiolabeled phosphate, the [^{32}P]-phosphate is incorporated into both the rat and mouse high-M_r proteins (Figure 3, lanes 5 and 6).

Immunologic Reactivity of the High-Molecular-Weight Proteins. Epidermal proteins separated on NaDodSO$_4$–urea–polyacrylamide gels were transferred to nitrocellulose and then analyzed for their reactivity with antibody to rat filaggrin. Filaggrins (rat, M_r 49 000; mouse, M_r 26 500) are readily identified in these epidermal extracts (Figure 3, lanes 7 and 8). In addition, proteins of very high M_r in each extract are strongly cross-reactive. The epidermal keratins show no reactivity with the antibody to filaggrin. The highest M_r cross-reactive protein seen in Figure 3, lanes 7 and 8, corresponds to the high-M_r proteins rapidly labeled with histidine and phosphate identified in Figures 1 (lanes 2–4 and lanes 5–7), 2, and 3 (lanes 3–6). The cross-reaction of the high-M_r bands in both rat and mouse with antiserum to rat filaggrin confirms their relationship to filaggrin. Rapid labeling of these proteins with histidine confirms their status as filaggrin precursors. Other protein bands also cross-react, but most of these are present in only trace amounts. They may be a result of limited proteolysis of the high-M_r protein rather than representing true processing of filaggrin precursor.

Molecular Weight. We initially estimated that the molecular weight of the mouse high-M_r filaggrin precursor is 300 000 and that of rat is 370 000 (from gradient polyacrylamide gels as in Figure 1); however, this estimate depended heavily upon extrapolation. When high-M_r markers were used in gradient polyacrylamide–NaDodSO$_4$–urea gels in an attempt to overcome this problem, the apparent M_r of the mouse filaggrin precursor was approximately 390 000, and that of rat was approximately 550 000–600 000.

Peptide Mapping. Peptide mapping was performed in order to verify the relationship between the high-M_r precursor and filaggrin. The limited proteolysis of the high-M_r filaggrin precursor can be divided into two stages. The first requires only small amounts of enzyme and is indicative of exposed, readily proteolyzed bonds. The second stage requires more enzyme for hydrolysis and suggests that the cleavage at these

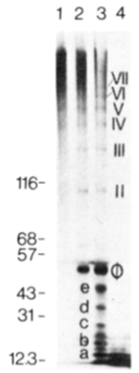

FIGURE 4: Chymotryptic digestion of the rat high-M_r filaggrin precursor generates an oligomeric series of peptides (II–VII), whose molecular weights are discrete multiples of fragment ϕ (M_r 54 000). The digestions were performed as described by Cleveland et al. (1977), with (1) 0, (2) 8 , (3) 80, and (4) 800 ng of enzyme. The gel is a 5–12% polyacrylamide gradient $NaDodSO_4$ gel stained with Coomassie brilliant blue. The positions of the molecular weight markers ($\times 10^{-3}$) are shown on the left. The peptides a–e stain with a distinct reddish color. They have M_r 15 200, 20 000, 28 800, 35 200, and 44 500, respectively.

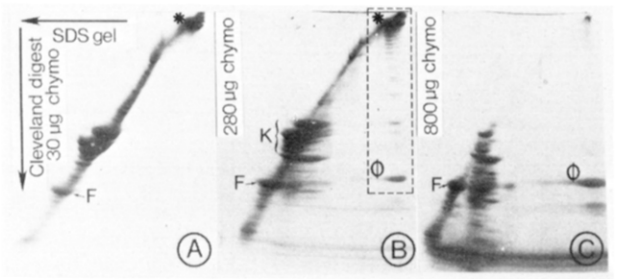

FIGURE 5: Two-dimensional chymotryptic peptide maps of mouse epidermal proteins. Proteins in 8 M urea extracts of mouse epidermis were separated by analytical $NaDodSO_4$–polyacrylamide gel electrophoresis in the first dimension (right to left) and then digested with chymotrypsin (as indicated in the figure) during electrophoresis in the second dimension in a 4–15% polyacrylamide gradient $NaDodSO_4$–urea gel. The asterisks indicate the position of the high-M_r filaggrin precursor. Filaggrin is indicated by (F). The digestion of the high-M_r filaggrin precursor seen in (A) proceeds through a ladderlike polymeric series of peptides [(B) rectangle; some of these can be identified as doublets] and results mainly in fragment ϕ [(C) M_r approximately 30 000]. In contrast, the keratins (K) are digested to form a complex pattern of peptides (B and C).

positions is more difficult. The effects of different amounts of enzyme on the filaggrin precursor can be seen in Figures 4–6.

When only a trace amount of enzyme is used to digest the rat filaggrin precursor (Figure 4), a relatively stable peptide, fragment ϕ, M_r approximately 54 000, is formed. A number of oligomers of ϕ can be seen. These oligomers are much less stable in the presence of proteases than is fragment ϕ and are presumably hydrolyzed rapidly to form ϕ. The formation of fragment ϕ and its oligomers occurs in the presence of either chymotrypsin (Figure 4) or trypsin (not shown). Fragment ϕ and its oligomers all contain phosphate.

A similar series of oligomeric proteins is formed by the limited chymotryptic hydrolysis of mouse high-M_r precursor (Figure 5). In this example, 8 M urea extracts of mouse epidermis were subjected to electrophoresis in $NaDodSO_4$ gels; then, an entire lane was cut out and positioned on top of another $NaDodSO_4$ gel for peptide mapping in the second dimension. As the amount of chymotrypsin is increased (from panel A to C), the high-M_r filaggrin precursor breaks down into a ladderlike series of oligomers included in the rectangle in panel B and then into a stable peptide, fragment ϕ (C), that

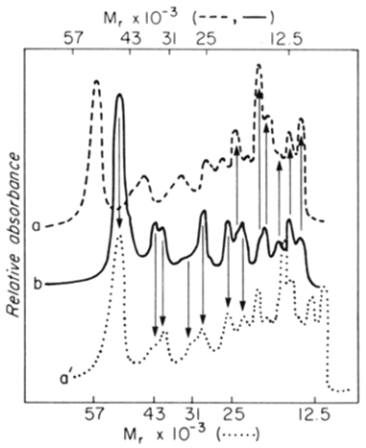

FIGURE 6: Comparison of the peptide maps of high-M_r filaggrin precursor and filaggrin. One-dimensional peptide mapping (Cleveland et al., 1977) was performed in $NaDodSO_4$ gels with 400 ng of chymotrypsin. The gels were stained and then scanned in a Helena Institute (Beaumont, TX) densitometer using a 595-nm filter. The scans (a and a′) are from a digest of high-M_r precursor similar to that shown in Figure 4, lane 3; however, only peptides up to the size of fragment ϕ (M_r 54 000) are shown for comparison with peptides from a similar filaggrin digest (b). The peptide map a′ has been displaced so that its M_r 54 000 peptide peak (fragment ϕ) lies under the peak of filaggrin (M_r 49 000). The displaced map and filaggrin map show considerable similarities (arrows between a′ and b) except in the lower M_r range where there are a few unpaired peptide peaks. However, these unpaired peaks in the filaggrin map are found to be paired when the precursor map is not displaced with respect to filaggrin (arrows between a and b). Thus, the digestion products from the high-M_r precursor fall into two populations when compared to those from filaggrin. The peptides in the lower M_r range migrate normally with respect to filaggrin, while the larger peptides from the precursor migrate consistently more slowly than those from filaggrin.

is slightly larger than mouse filaggrin (cf. rat fragment ϕ, which is slightly larger than rat filaggrin). The simple pattern of breakdown of the high-M_r band to yield fragment ϕ and a ladderlike series of peptides is in distinct contrast to the complex peptide map generated by the keratins (Figure 5C).

The oligomeric patterns generated in these digests are similar to our earlier preparations of rat filaggrin precursor (Dale & Ling, 1979; Lonsdale-Eccles et al., 1980, 1982) and suggest that the earlier preparations had undergone limited proteolysis. Although the extraction buffer used in the earlier experiments (1 M potassium phosphate, pH 7.6, 0.24 M histidine, 10 µg/mL PMSF) contained the proteolytic inhibitor PMSF, it also contained the strong nucleophile histidine, which may have caused the hydrolysis of this inhibitor. When histidine was omitted from the buffer, the amounts of M_r 54 000 peptide and of oligomers were reduced, and that of diffuse high-M_r protein was increased (B. A. Dale, unpublished observations). Thus, the M_r 54 000 monomeric precursor that we described previously (Dale & Ling, 1979; Lonsdale-Eccles et al., 1980, 1982) is most likely a proteolytic product of the high-M_r precursor described in the present paper and is similar to fragment ϕ.

Although fragment ϕ is relatively stable in the presence of small amounts of chymotrypsin, as shown in Figure 4, in the presence of larger amounts of the protease it is further digested to smaller fragments that may be identified by their phosphate content and metachromatic staining (Figure 4, lane 3). [The histidine-rich proteins exhibit a reddish tinge when the Coomassie brilliant blue stained gels are extensively destained with 10% acetic acid (Balmain, 1977).] The reddish peptides (a–e) have apparent molecular weights of 15 200, 20 000, 28 800, 35 200, and 44 500, while fragment ϕ has an M_r of 54 000. It appears that peptides averaging M_r of 8 700 are sequentially

removed from fragment ϕ. The significance of this value is discussed later.

The small peptides formed from the high-M_r precursor, via fragment ϕ, are compared in Figure 6 with peptides generated from filaggrin. Two populations of peptides are produced. The larger, phosphorylated peptide group migrates consistently more slowly than those from filaggrin, while the smaller peptides migrate normally with respect to the smaller peptides from filaggrin. Every peptide in the digest of filaggrin can be aligned to a corresponding peptide from one or the other of the two groups of peptides in the precursor digest. The amounts of the respective peptides differ, suggesting that there is a different rate of cleavage at the common sites within each protein. This is probably a consequence of conformational differences between the two proteins, caused by the presence of phosphate on the precursor peptides. The similarity between the high-M_r precursor and filaggrin peptide maps is consistent with earlier studies using filaggrin and its putative precursor with M_r of 54000 (Lonsdale-Eccles et al., 1982).

Discussion

Newborn rat and mouse filaggrin are epidermal histidine-rich proteins that have molecular weights of 49000 and 26500, respectively, by $NaDodSO_4$–polyacrylamide gel analysis. These proteins incorporate radiolabeled histidine only after a lag of more than 6 h. In the work presented here, we have shown that a high-M_r protein in both rat and mouse can be rapidly labeled with histidine, that this band is a phosphoprotein in each species, that it cross-reacts with antibody to rat filaggrin, and that peptide mapping yields a series of oligomers and eventually a series of peptides similar to those from filaggrin itself. These results suggest that filaggrin in rat and mouse is derived from a very high M_r phosphorylated precursor protein. These precursors have molecular weights (390000, mouse; 600000, rat) approximately 10–15 times as large as that of filaggrin. Recent work by Ball et al. (1978), Goldsmith & Wilkening (1980), Scott & Harding (1981), Okazaki et al. (1983), and Ramsden et al. (1983) also supports the conclusion that filaggrin precursor is very large.

Proteins as large as this can be made directly by the protein synthesis machinery of the cell or alternatively by cross-linking of smaller units into a large product. The evidence in this case suggests initial synthesis as a large polypeptide. During pulse–chase studies with radiolabeled histidine, there is no evidence of the synthesis of labeled proteins in the molecular weight range of 30000–60000 prior to the incorporation of radiolabel into the high-M_r proteins. The high-M_r forms of these histidine-rich proteins are stable in 9% formic acid (Dale et al., 1980; Lonsdale-Eccles et al., 1981), in 8 M urea, in $NaDodSO_4$, and in β-mercaptoethanol (Ball et al., 1978; this study), in EDTA (Ball et al., 1978), in 6 M guanidinium chloride (Ball et al., 1978; K. A. Resing, unpublished observations), and in sodium thiocyanate (Resing et al., 1983). But the strongest evidence to date is that rat and mouse filaggrin mRNAs are large, 34 S and 30 S, respectively, and can be translated in vitro, yielding filaggrin-immunoprecipitable proteins of $M_r > 200000$ (Meek et al., 1983). This evidence points to the direct synthesis of the histidine-rich proteins as high-M_r polypeptides, although because of the problems in interpreting molecular weights of greater than 200000 from $NaDodSO_4$ gels, we cannot exclude the possibility of dimerization of the translation product via cross-linking. Conclusive evidence must await protein or DNA sequence analysis.

If the high-M_r proteins are indeed synthesized as single long polypeptide chains, then two models can be constructed for the relationship between filaggrin and its precursor. Filaggrin

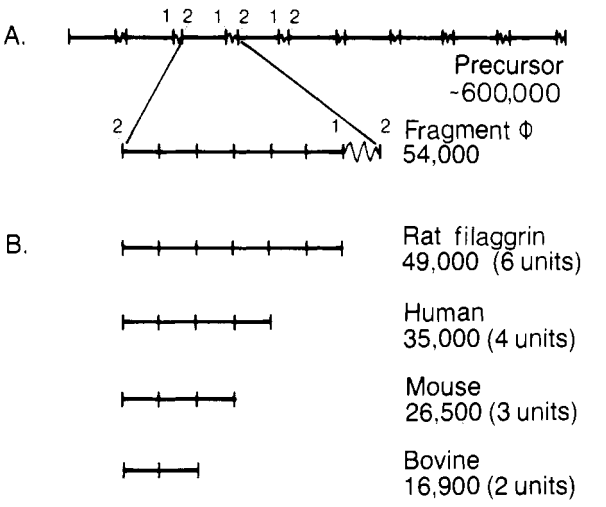

FIGURE 7: Proposed model for the repeated domain structure of filaggrin and its high-M_r precursor. (A) The high-M_r precursor is composed of tandem repeated domains of fragment ϕ, M_r 54000, in the rat. The fragment ϕ (rat) is composed of six smaller units (each approximately M_r 8200) plus a linker peptide. Multiple molecules of filaggrin are generated by proteolytic cleavages at all the sites indicated by 1 and 2. If the cleavages between adjoining units of fragment ϕ are not complete, then higher oligomers will be observed. (B) Filaggrins from several species are compared diagrammatically to illustrate the smaller repeated unit of approximately M_r 8200–8900 (see Table I).

may simply be only 5–10% of the high-M_r protein, with the remaining 90–95% removed during processing. Alternatively, the filaggrin could be repeated many times within the polypeptide chain of the precursor. Our data support the second possibility for several reasons.

First, at least 50% of the histidine incorporated into the high-M_r band is retained in filaggrin for both the rat and mouse proteins. Second, peptides derived from filaggrin and its high-M_r precursor are quite similar (Figure 6), especially if the anomalous migration of phosphopeptides is taken into account [see also Lonsdale-Eccles et al. (1982)]. Thus, it is unlikely that the precursor contains large segments that are significantly different from filaggrin itself, even though their molecular weights differ at least by a factor of 10. Similar amino acid compositions and peptide maps have also been reported for the histidine-rich protein HRP2 (equivalent to filaggrin; Dale et al., 12981) and its high-M_r precursor HRP1 (Ball et al., 1978). Third, a repeating structure is suggested by cleavage of the high-M_r precursor into a series of oligomers, whose molecular weights are discrete multiples of the M_r 54000 phospho form (fragment ϕ) of rat filaggrin or the equivalent form (M_r approximately 30000) from mouse. Cleavages must occur at regular intervals along the polypeptide chain to generate this series. An oligomeric series was described for bovine and rat histidine-rich proteins (Ugel, 1971; Dale & Ling, 1979; Lonsdale-Eccles et al., 1980, 1982). We identified the M_r 54000 form as the putative precursor of rat filaggrin. We now believe that this was not the complete molecule but rather a basic repeating unit of the true high-M_r precursor.

The structure for the large filaggrin precursor that is most readily compatible with these data is one that encompasses multiple units of filaggrin, in the form of tandem repeated domains. A diagram of this model is shown in Figure 7A. A linker peptide is included to account for the difference in size between fragment ϕ and filaggrin. Although this type of repeating structure is not common, there is ample precedent for it. "Polyprotein" precursors to a number of regulatory

Table I: Interspecies Comparison of Filaggrin Molecular Weights

animal	M_r[a]	repeat no.[b]	M_r of repeated unit	ref
bovine	16 900	2	8450	Ugel, 1975
mouse	26 500	3	8830	Steinert et al., 1981
human	35 000	4	8750	Lynley & Dale, 1983
guinea pig	44 500	5	8900	Scott & Harding, 1981
rat	49 000	6	8200	Dale, 1977
rabbit	61 000	7	8710	Lonsdale-Eccles et al., 1981
			8600 (±300)[c]	

[a] Values calculated from $NaDodSO_4$-polyacrylamide gel electrophoresis. [b] Proposed number of tandem repeat sequences within each respective filaggrin molecule. [c] Average.

peptides have been identified and recently reviewed (Herbert & Uhler, 1982). Silk fibroin (Gage & Manning, 1980; Greene et al., 1975; Lucas & Rudall, 1968) and the maize storage protein zein (Pedersen et al., 1982; Marks & Larkins, 1982) are nonenzymatic proteins that have repetitive amino acid sequences indicative of tandem repeats.

The tandem repeating domain structure may have important consequences with respect to overall structure, function, and processing of the filaggrin precursor. Close packing of the molecule within the keratohyalin granule could expose phosphorylated residues and the sites for subsequent proteolytic cleavage to the action of enzymes released from lysosomes. This would facilitate rapid activation to form filaggrin that can immediately react with the adjacent keratin filaments during the conversion of granular cells to cornified cells.

Since the filaggrins from different species are all presumed to have the same role in epidermal differentiation, one might expect conservation of filaggrin structure. Indeed, the amino acid composition of filaggrins from different species is quite similar, and there is partial immunologic cross-reactivity. One would also expect that filaggrins would be similar in size. However, the filaggrins from various species studied to date all have different molecular weights (Table I). This variation in size can be reconciled with the conserved amino acid composition by assuming that a "primordial filaggrin" has undergone gene duplication. In this instance, however, the repeated unit is an unidentified polypeptide whose assumed molecular weight is 8 600 (±300). Each of the filaggrins identified thus far has a molecular weight that is a discrete multiple of 8 600. Mouse filaggrin may be considered to consist of three of these units, human filaggrin of four, etc. (Table I, Figure 7B). Although the argument for the repetitive M_r 8 600 unit is based primarily on numerology, chymotrypsin does cleave rat filaggrin precursor fragment ϕ (Figure 4) into several smaller peptides with sizes that are approximate multiples of 8 700. We speculate that the histidine-rich, M_r 8 500 "Fast Protein" from chick feather (Walker & Rogers, 1976; Powell & Rogers, 1979) may be related to the fundamental repeating unit hypothesized here.

The model for the structure of the high-M_r filaggrin precursor (Figure 7) has two types of repeats, the small unit of M_r 8 600 and the larger domain consisting of filaggrin and a possible linker. The model incorporates the known data on the structure of a filaggrin and its precursor . It assumes the simplest arrangement, but the actual structure may have a more complex topology, with the possibility of cross-links between domains. This is meant to be a general model, providing a basis for further study of biochemical aspects of the maturation of this protein. Efforts to test the model by sequence analysis of the proteins and of DNA complementary to the mRNA are under way.

Acknowledgments

We gratefully acknowledge the technical assistance of Julie Haugen Scofield and Janet Kimball. We thank Drs. Kenneth Walsh and David Teller for their critical reading of the manuscript and Alison Ross for editorial assistance.

References

Argos, P., Pederson, K., Marks, M. D., & Larkins, B. A. (1982) *J. Biol. Chem. 257*, 9984–9990.

Ball, R. D., Walker, G. K., & Bernstein, I. A. (1978) *J. Biol. Chem. 253*, 5861–5868.

Balmain, A., Lochren, D., Fischer, J., & Alonso, A. (1977) *Dev. Biol. 60*, 442–452.

Bernstein, I. A. (1970) *J. Soc. Cosmet. Chem. 21*, 583–584.

Bernstein, I. A., Brabec, R. K., Vadlamudi, B., Ohno, A. K., DeLap, L. W., & Gray, R. H. (1980) in *Biochemistry of Normal and Abnormal Epidermal Differentiation* (Bernstein, I. A., & Seiji, M., Eds.) pp 421–432, University of Tokyo Press, Tokyo.

Cleveland, D. W., Fischer, S. A., Kirschaer, M. W., & Laemmli, U. K. (1977) *J. Biol. Chem. 252*, 1102–1106.

Dale, B. A. (1977) *Biochim. Biophys. Acta 491*, 193–204.

Dale, B. A., & Ling, S.-Y. (1979) *Biochemistry 18*, 3539–3546.

Dale, B. A., Holbrook, K. A., & Steinert, P. M. (1978) *Nature (London) 226*, 729–731.

Dale, B. A., Vadlamudi, B., DeLap, L. W., & Bernstein, I. A. (1981) *Biochim. Biophys. Acta 668*, 98–106.

Gage, L. P., & Manning, R. F. (1980) *J. Biol. Chem. 255*, 9444–9450.

Goldsmith, L. A., & Wilkening, B. (1980) *Arch. Dermatol. Res. 267*, 253–263.

Greene, R. A., Morgan, M., Shatkin, A. K., & Gage, L. P. (1975) *J. Biol. Chem. 250*, 5114–5121.

Herbert, E., & Uhler, M. (1982) *Cell (Cambridge, Mass.) 30*, 1–2.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

Lonsdale-Eccles, J. D., Haugen, J. A., & Dale, B. A. (1980) *J. Biol. Chem. 255*, 2235–2238.

Lonsdale-Eccles, J. D., Lynley, A. M., & Dale, B. A. (1981) *Biochem. J. 197*, 591–597.

Lonsdale-Eccles, J. D., Teller, D. C., & Dale, B. A. (1982) *Biochemistry 21*, 5940–5948.

Lucas, F., & Rudall, K. M. (1968) *Compr. Biochem. 26B*, 475–558.

Lynley, A. M., & Dale, B. A. (1983) *Biochim. Biophys. Acta 744*, 28–35.

Marks, M. D., & Larkins, B. A. (1982) *J. Biol. Chem. 257*, 9976-9983.

Meek, R. L., Lonsdale-Eccles, J. D., & Dale, B. A. (1983) *Biochemistry 22*, 4867–4871.

Murozuka, T., Fukuyama, K., & Epstein, W. L. (1979) *Biochim. Biophys. Acta 579*, 334–345.

Okazaki, K., Horii, I., & Seiji, M. (1983) in *Biochemistry of Epidermal Differentiation* (Seiji, M., & Bernstein, I. A., Eds.) Tokyo University Press, Tokyo (in press).

Pederson, K., Devereux, J., Wilson, D. R., Sheldon, E., & Larkins, B. A. (1982) *Cell (Cambridge, Mass.) 29*, 1015–1026.

Powell, B., & Rogers, G. E. (1979) *Nucleic Acids Res. 7*, 2165–2176.

Ramsden, M., Loehren, D., & Balmain, A. (1983) *Differentiation (Berlin) 23*, 243–249.

Scott, I. R., & Harding, C. R. (1981) *Biochim. Biophys. Acta 669*, 65–78.
Sibrack, L. A., Gray, R. H., & Bernstein, I. A (1974) *J. Invest. Dermatol. 62* 394–405.
Steinert, P. M., Cantieri, J. S., Teller, D. C., Lonsdale-Eccles, J. D., & Dale, B. A. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4097–4101.
Ugel, A. R. (1971) *J. Cell Biol. 49*, 405–422.
Walker, I. D., & Rogers, G. E. (1976) *Eur. J. Biochem. 69*, 329–339.

Amino Acid Sequence of Crayfish (*Astacus fluviatilis*) Carboxypeptidase B†

Koiti Titani, Lowell H. Ericsson, Santosh Kumar, Frank Jakob, Hans Neurath, and Robert Zwilling*

ABSTRACT: The complete amino acid sequence of carboxypeptidase B from the crayfish *Astacus fluviatilis* has been determined. The S-carboxymethylated protein was cleaved with cyanogen bromide and with trypsin, before and after citraconylation of lysine residues. Peptides were purified by gel filtration followed by reverse-phase high-performance liquid chromatography and analyzed by a liquid-phase sequencer. The enzyme contains 303 amino acid residues and has a molecular weight of 33 899 (without zinc). The crayfish enzyme shows 44.7% sequence identity with bovine carboxypeptidase B as compared to 44.0% identity with bovine carboxypeptidase A. The constituents of the active site of the bovine enzymes are present in corresponding positions in the crayfish enzyme. Isoleucine-255, which, in carboxypeptidase A, is located at the pit of the substrate binding pocket, is occupied by aspartic acid in the crayfish enzyme, as in bovine carboxypeptidase B. When compared to bovine and rat carboxypeptidases, the amino acid sequence of the crayfish enzyme displays one deletion, two insertions, and 35 amino acid replacements but is fully compatible with the X-ray structure of bovine carboxypeptidase B. The present analysis affords the first detailed view into the evolution of carboxypeptidases at the invertebrate level.

Bovine pancreatic carboxypeptidase A (EC 3.4.12.2) is the most thoroughly characterized metallopeptidase and has served as a prototype in the study of the mechanism of action of metalloenzymes in general [for the most recent review, see Vallee et al. (1983)]. The amino acid sequences (Bradshaw et al., 1969; Titani et al., 1975) and the X-ray structures (Lipscomb et al., 1968; Schmid & Herriott, 1976) of both bovine carboxypeptidases A and B (EC 3.4.12.3) are known. By comparison, little is known of carboxypeptidases of other mammalian species (Folk, 1970; Reeck & Neurath, 1972; Everitt & Neurath, 1980; Quinto et al., 1982) and even less of those of invertebrate origin (Gates & Travis, 1973; Zwilling et al., 1979, 1980). To fill this gap of knowledge, and to establish a data base for a study of the evolution of carboxypeptidases, we have undertaken an investigation of the amino acid sequence of crayfish (*Astacus fluviatilis*) carboxypeptidase. This species was chosen for several reasons: (1) crayfish lies on the evolutionary pathway from which decapode crustacea and mammals diverged some 700 million years ago; (2) preliminary analysis of the amino-terminal sequence of crayfish carboxypeptidase indicated significant structural homology to bovine carboxypeptidase A (Zwilling et al., 1979); (3) in view of our recent findings that bovine and crayfish trypsins are homologous and evidently have diverged from a common ancestor (Titani et al., 1983), it appeared of interest to compare the rates of evolution of two enzymes that originated from analogous structural milieus, the pancreas in the case of the mammalian enzymes and the hepatopancreas in the case of the crayfish enzymes.

In this paper, we report the complete amino acid sequence of crayfish (*Astacus fluviatilis*) carboxypeptidase B and evaluate its homology to bovine and rat carboxypeptidases. On the basis of a comparison of bovine and crayfish carboxypeptidases and trypsins, respectively, we present arguments about the rates of evolution of these pancreatic proteases.

Materials and Methods

Crayfish carboxypeptidase originates in the hepatopancreas of the animal and is secreted into the stomachlike cardia where it is present in the active form at approximately 1 mg/mL digestive fluid. No evidence for the presence of a zymogen form has been obtained thus far.

Digestive fluid was collected from the cardia of a large number of animals by the method previously described for crayfish trypsin (Zwilling & Neurath, 1981) and stored in the frozen state until used. Carboxypeptidase was isolated by affinity chromatography on immobilized potato inhibitor as described by Ager & Hass (1977). Details of the isolation procedure have been previously published (Zwilling et al., 1979). Material prepared in this fashion was homogeneous as judged by sodium dodecyl sulfate gel electrophoresis, two-dimensional immunoelectrophoresis, and amino-terminal sequence analysis (Zwilling et al., 1979).

The protein was reduced with dithiothreitol and was S-carboxymethylated with either cold or 1-^{14}C-labeled iodoacetic acid (New England Nuclear). CM-protein[1] was cleaved with cyanogen bromide or with trypsin after labeling of methionyl residues with [^{14}C]methyl iodide (New England Nuclear) (Sasagawa et al., 1983). [^{14}C]CM-protein was digested with

† From the Howard Hughes Medical Institute and the Department of Biochemistry, University of Washington, Seattle, Washington 98195 (K.T., L.H.E., S.K., and H.N.), and the Department of Physiology, Institute of Zoology, University of Heidelberg, D-6900 Heidelberg, FRG (F.J. and R.Z.). *Received October 11, 1983.* This work was also supported by the National Institutes of Health (GM-15731) and by research grants from the Deutsche Forschungsgemeinschaft, Bonn-Bad Godesberg, FRG (ZW 17/7 and ZW 17/11-1).

[1] Abbreviations: CM, carboxymethyl; HPLC, high-performance liquid chromatography; Pth, phenylthiohydantoin.

0006-2960/84/0423-1245$01.50/0