Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4848–4852, July 1989
Biochemistry

Characterization of a cDNA clone encoding human filaggrin and localization of the gene to chromosome region 1q21

(expression library/polyprotein repeat/polymorphism/*in situ* hybridization)

LYNN J. MCKINLEY-GRANT*, WILLIAM W. IDLER*, ISADORE A. BERNSTEIN*†, DAVID A. D. PARRY*‡, LINDA CANNIZZARO§, CARLO M. CROCE§, KAY HUEBNER¶, STUART R. LESSIN‖, AND PETER M. STEINERT*

*Dermatology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; †Departments of Biological Chemistry and Industrial and Environmental Health, University of Michigan, Ann Arbor, MI 48109; ‡Department of Physics and Biophysics, Massey University, Palmerston North, New Zealand; §Southwest Biomedical Research Institute, Scottsdale, AZ 85251; ¶Fels Institute for Cancer Research and Molecular Biology, Philadelphia, PA 19140; and ‖Dermatology Department, University of Pennsylvania, Philadelphia, PA 19104

Communicated by Igor B. Dawid, March 24, 1989

ABSTRACT **Filaggrins are an important class of intermediate filament-associated proteins that interact with keratin intermediate filaments of terminally differentiating mammalian epidermis. They show wide species variations and their aberrant expression has been implicated in a number of keratinizing disorders. We have isolated a cDNA clone encoding human filaggrin and used this to demonstrate that the human gene encodes a polyprotein precursor containing numerous tandem filaggrin repeats. This structure is similar to that of mouse; however, the human filaggrin repeat is much longer (972 base pairs; 324 amino acids) and shows little sequence homology to the mouse protein. Also, data presented here reveal that the human filaggrin repeats show considerable sequence variations; such polymorphism is not found in the mouse. Furthermore, chromosomal mapping data revealed that the human gene is located at 1q21, indicating that the polymorphism is confined to a single locus. By peptide mapping, we define a short linker sequence within the human filaggrin repeat that is excised by proteolysis to yield functional molecules. Finally, we show by *in situ* hybridization that human filaggrin precursor gene expression is tightly regulated at the transcriptional level in terminally differentiating epidermis and that this represents a useful system in which to study intermediate filament–intermediate filament-associated protein interactions as well as disorders of keratinization.**

Intermediate filaments (IFs) are complex constituents of the cytoskeletons of most types of eukaryotic cells and indeed some 30–40 proteins exist in each mammalian species (1). While their precise functions are unknown, they are thought to be involved in the physical coordination of cell shape and nuclear centration and may also be involved in maintaining the differentiation state of the cell (1–3). In addition, IFs are usually closely associated with numerous other proteins (IF-associated proteins or IFAPs) that are involved in the supramolecular organization of the IF in cells (1, 2).

One of the most quantitatively significant IF–IFAP systems is the interaction of filaggrin with epidermal keratin IF (1, 4–7). These proteins are expressed in terminally differentiating epidermis and constitute the "keratin" of mammalian epidermis as a filament–interfilamentous matrix. To understand this interaction in molecular terms, we recently cloned and sequenced mouse filaggrin (8) and found that its gene encodes a polyprotein precursor consisting of many filaggrin repeats arranged in tandem perhaps separated by a linker peptide (7–11). This precursor protein initially accumulates in the epidermis as keratohyalin granules (8, 12); it is highly phosphorylated and incapable of interaction with keratin IFs. At the time of terminal differentiation, the precursor is proteolytically processed by excision of the linker to individual filaggrin molecules, which are then able to interact with keratin IFs (7, 9–11). Since the aberrant expression of filaggrin has been implicated in a number of keratinizing disorders (13–16), it is important to answer fundamental questions concerning the structure and expression of its gene. Work with the human system, however, has been complicated by the realization that antibodies against mouse filaggrin (17, 18) or cDNA clones encoding mouse filaggrin (8, 19) do not recognize human filaggrin, implying some important differences in their properties.

In this paper, we describe the isolation and characterization of a clone encoding human filaggrin.** We find that human filaggrin is also expressed as a polyprotein precursor, but the size and sequence of the repeats are quite different from those in the mouse. In addition, we present evidence that the human repeats are polymorphic. Accordingly, we have localized the gene to demonstrate that such variations are confined to a single locus on chromosome 1.

MATERIALS AND METHODS

Using poly(A)-enriched mRNA obtained from human foreskin epidermal tissue (20), we constructed a cDNA library by use of a random oligonucleotide primer (21). cDNA species >0.5 kilobase pairs (kbp) were used to construct a library in λgt11 (0.9×10^6 plaque-forming units unamplified). This was screened by standard procedures (21, 22) with a human anti-filaggrin monoclonal antibody (17) kindly provided by Howard Baden and Joseph Kubilus (Massachusetts General Hospital). Following plaque purification of antibody-positive clones, the insert of the longest clone (λHF10; 1.248 kbp) was transferred to pGEM-3B (Promega Biotec) for synthesis of RNA probes for *in situ* hybridization (8) or M13mp18 or M13mp19 vectors for DNA sequencing with Sequenase 2 (United States Biochemical) (8). Northern and Southern blots were done as before (8, 20). Chromosomal mapping by somatic cell hybrid analysis and chromosomal *in situ* hybridization was done as described (23, 24). Fresh human foreskins were cultured in medium 199 and 10% fetal calf serum for 4 hr with L-[*ring*-2,3,4,5,6-^{3}H]phenylalanine (100 μCi/ml; 1 Ci = 37 GBq) (New England Nuclear/DuPont). Filaggrin and its precursor were isolated (5, 25) and subjected to amino acid analysis by established procedures (26). About 0.5 mg (2000–3000 cpm) of each protein was also digested with trypsin (Sigma, type IX) for 2 hr at 37°C at ≈1 mg/ml in 0.1 M *N*-ethylmorpholine acetate (pH 8.3). Peptides were re-

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

Abbreviation: IF, intermediate filament.

**The sequence reported in this paper has been deposited in the GenBank data base (accession no. M24355).

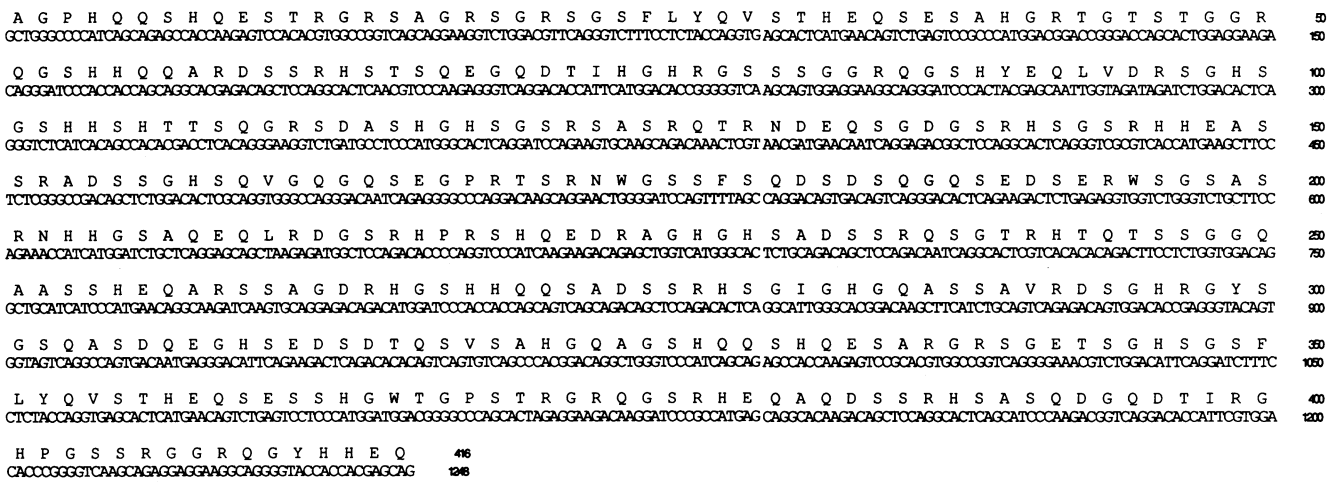

FIG. 1. Nucleotide sequence of the cDNA clone λHF10 and its deduced amino acid sequence (using single-letter code).

solved with a peptide analyzer and selected radioactive peaks were recovered for amino acid analysis.

RESULTS AND DISCUSSION

Isolation and Characterization of cDNA Clones Encoding the Human Filaggrin Precursor. Using a well-defined monospecific monoclonal antibody against human filaggrin (17), we identified 12 antibody-positive clones for the precursor from $\approx 2 \times 10^5$ plaque-forming units of a random-primed human epidermal cDNA library ligated into λgt11. The longest clone, λHF10, was sequenced, found to be 1248 bp long, and contained a single open reading frame encoding 416 amino acids (Fig. 1). Notably, the first 92 amino acid residues (276 bp) share 78% identity and 86% homology with the last 92 residues, thereby delineating a 324-residue (972-bp) repeat. The sequence data reveal a conserved aliphatic region (residues 23–38 and 347–362). This is flanked by sequences that, interestingly, are homologous to regions in mouse filaggrin, which are thought to demarcate a linker sequence that is excised from the polyprotein precursor to produce filaggrin (8, 19). By alignment of these mouse and human sequences to maximize homology (Fig. 2), we can demonstrate that the human filaggrin repeats are likely to be defined by a short linker of length within the range of 7–13 residues: the most striking homology predicts a linker of 7 residues corresponding to residues 350–356 (and residues 26–32). To explore whether this region is indeed part of the linker, fresh foreskins were labeled in organ culture with [^{3}H]phenylalanine, filaggrin and its precursor were purified, and their tryptic peptides were resolved. Fig. 3 shows that only acidic peptides contain radioactivity and are similar, but the precursor yielded three labeled peptides (peaks 4, 13, and 14), whereas in purified filaggrin, peaks 13 and 14 were completely absent and only peak 4 was labeled. These peptides have amino acid compositions as follows (single-letter code): peptide 4, $D_{3.9}S_{6.7}E_{4.1}G_{2.0}F_{1.0}H_{0.9}R_{1.0}$; peptide 13, $S_{4.8}T_{1.0}E_{3.8}G_{2.0}A_{1.0}$-$V_{0.9}L_{1.0}F_{1.0}Y_{1.0}H_{1.9}R_{1.0}$; peptide 14, $S_{8.5}T_{3.8}E_{4.8}P_{1.0}G_{4.7}V_{1.0}$-$L_{1.0}F_{1.0}Y_{1.0}H_{3.1}R_{1.0}$, which correspond very closely to sequences 174–193, 23–42, and 340–372 of Fig. 1, respectively. Therefore, sequences in the vicinity of phenylalanines 26 and 350 are removed during processing of the precursor and must demarcate the linker region. Accordingly, we suggest that human filaggrin is defined by residues 33–349 of Fig. 1. Such a protein of 317 residues has a molecular mass of ≈34.2 kDa, which is very close to the size of human filaggrin estimated on denaturing polyacrylamide gels to be 34 kDa (17) or 38 kDa (25). Moreover, its amino acid composition is quite similar to acid hydrolyzates of isolated and purified human filaggrin and its precursor (Table 1), and it has a pI of ≈9.5 as predicted from its biochemical properties.

The data of Figs. 1 and 3 have revealed sequence variations amongst the acidic peptides flanking the linker of the human filaggrin precursor; that is, the various repeats are polymorphic. This means that human filaggrin itself consists of a heterogeneous population of similar molecules whose variations cannot be attributed to charge heterogeneities resulting from postsynthetic modifications (9). In contrast, sequencing (8) and peptide mapping (11) data of mouse filaggrin and its precursor revealed no such polymorphism. Including the region shown in Fig. 2, the overall homology of mouse and human filaggrins is ≈20%, but beyond this region, the homology is only slight. There are other quasi-repeating peptide sequences within human filaggrin that resemble but are not homologous to sequences within mouse filaggrin.

The Human Filaggrin Precursor Gene Encodes a Polyprotein. The cDNA clone was used for Northern blot analysis and revealed a significantly degraded major mRNA species of up to 15 kb in human poly(A)-enriched epidermal RNA (Fig. 4*A*) but it did not hybridize to mouse or rat RNA (data not shown). This lack of cross-hybridization between the species has been described (19) and is not surprising in view of the low sequence homology. We tested several methods and sources of isolating human epidermal RNA. While the mRNAs for other epidermal differentiation products such as for keratins 1 and 10 were always intact, the filaggrin mRNA was always partially degraded, suggesting that this message may be partially processed in the tissue.

Southern blot analyses were performed on total human genomic DNA from a single placenta (Oncor, Gaithersburg, MD) (Fig. 4*B*). DNA cut by several restriction enzymes (for example, *Hgi*AI and *Sma* I and their isoschizomers) yielded a

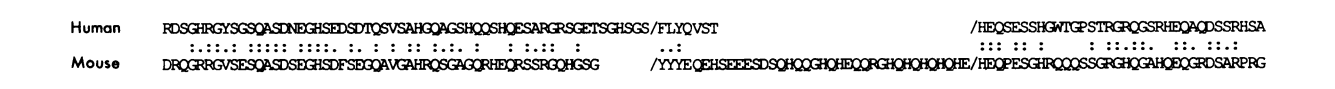

FIG. 2. Comparison of human and mouse precursor sequences to demonstrate region of high homology in the vicinity of the putative linker. Residues 292–390 of the human sequence (see Fig. 1) are aligned with the mouse sequence (8) to maximize homology. The single-letter code is used. :, exact match; ., conservative change; /, possible linker sequences that are excised during processing of the precursors to the filaggrin units. In the case of the human protein, the linker may be as short as 7 residues (as shown) or as long as 13 residues (including the residues inserted with respect to the mouse).

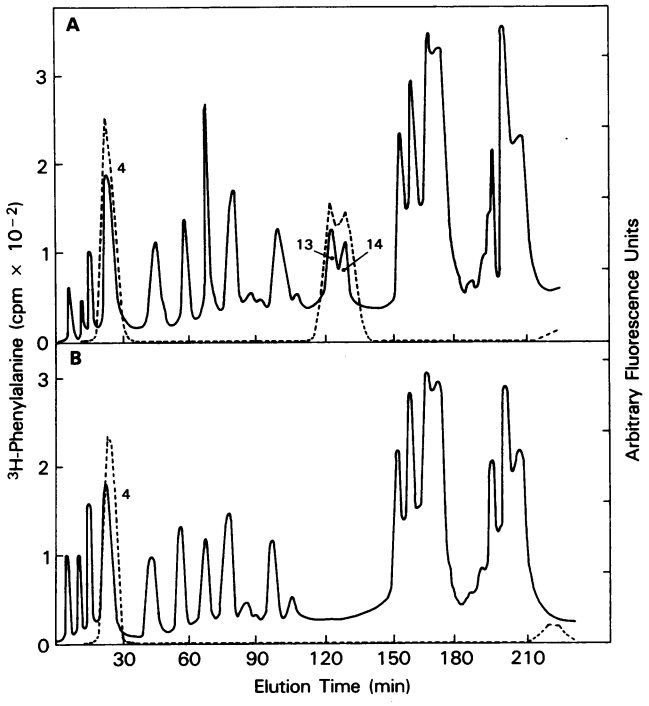

FIG. 3. Identification of [^{3}H]phenylalanine-labeled tryptic peptides of isolated human filaggrin precursor (*A*) and filaggrin (*B*). The peptides were separated with an initial buffer of 0.2 M sodium citrate (pH 3.25) for 140 min followed by 0.4 M sodium citrate (pH 6.5). Solid line, fluorescence trace; broken line, radioactivity. This efficiently separated the acidic peptides that contain label; optimal separation of unlabeled basic peptides was not attempted. Labeled peaks 13 (corresponding to residues 23–42; see Fig. 1) and 14 (residues 340–372) present in *A* are absent from *B*, while peak 4 is present in both (corresponding to residues 174–194). Accordingly, we predict that the phenylalanine residues 26 and 350 are involved in the linker region that is proteolytically removed during processing. Attempts to isolate the unique amino- and carboxyl-terminal peptides of filaggrin have not yet been successful because of sequence polymorphisms. Note that peak 7 is NH_3.

band at ≈975 bp, which is clearly superstoichiometric with respect to weaker bands presumably representing flanking gene sequences. Interestingly, the partial *Sma* I digest shown in Fig. 3*B* (lane 3) revealed a series of bands spaced ≈975 bp apart corresponding to multiples of this repeat. *Pst* I revealed a repeat of ≈800 bp as expected (lane 2), since the cDNA sequence data contains two *Pst* I sites per 972 bp giving rise to fragments of 797 and 175 bp (Fig. 1). DNA cut by certain other enzymes such as *Dra* I and *Eco*RV (lanes 4 and 5) yielded a single large band of 15–16 kbp, thereby setting an upper limit on the size of the coding portion of the gene. Many other restriction enzymes that did not cut the cDNA clone nevertheless cut genomic DNA into a variety of intermediate-sized fragments (lanes 6 and 7 for *Eco*RI and *Sac* I). This indicates that there must be considerable nucleic acid sequence variation within the various repeats along the gene and/or between different individuals. We have found that such sequence variations do not occur in the mouse filaggrin gene, wherein the repeats share >99% sequence identity (J. A. Rothnagel and P.M.S., unpublished data). These Southern blotting data support the conclusions of the data of Figs. 1 and 3. Further work will be necessary to fully characterize the extent and complexity of the polymorphism of the human filaggrin precursor gene system. Nevertheless, like the mouse gene, the close similarity of the size of the repeat recognized in the cDNA clone and that seen in genomic DNA indicates that the human filaggrin precursor gene consists of a tandem array of repeating, if not polymorphic, filaggrin units devoid of introns in the coding regions. The data from the Southern and Northern blots suggest that there are at least 12 repeats in the human polyprotein precursor.

Table 1. Amino acid composition of human filaggrin (residues per 100 residues)

Amino acid	Deduced from cDNA clone (residues 33–349)	Determined by acid hydrolysis: Filaggrin Ref. 25	Filaggrin This work	Precursor This work
Asp + Asn	8.26	8.9	7.65	8.05
Thr	4.42	1.2	4.15	3.75
Ser	24.92	13.8	22.55	20.90
Glu + Gln	15.14	14.8	14.40	13.70
Pro	0.63		1.00	0.50
Citrulline			1.25	Trace
Gly	15.14	19.0	16.55	17.40
Ala	6.63	6.6	5.85	6.25
Val	1.26	4.3	1.35	2.00
Cys		Trace	Trace	Trace
Met		0.9	Trace	Trace
Ile	0.63	2.3	0.75	1.05
Leu	0.63	4.7	1.25	1.35
Phe	0.32	1.9	0.70	1.25
Tyr	0.63	2.1	1.05	1.60
Ornithine			0.45	
Lys		3.6	Trace	0.20
His	11.99	6.5	11.95	12.55
Trp	0.63		0.55	0.30
Arg	9.78	7.9	8.35	9.05
Ser(*P*)*			0.30	6.45

Note that human filaggrin contains some citrulline, which is thought to arise by desimidation of certain of the arginine residues (9). On acid hydrolysis, citrulline is partially degraded to ornithine. The total arginine + citrulline + ornithine content is similar to the total arginine of the deduced sequence.

*This shows that the human precursor, like the rat and mouse filaggrin precursors (7, 8), is highly phosphorylated.

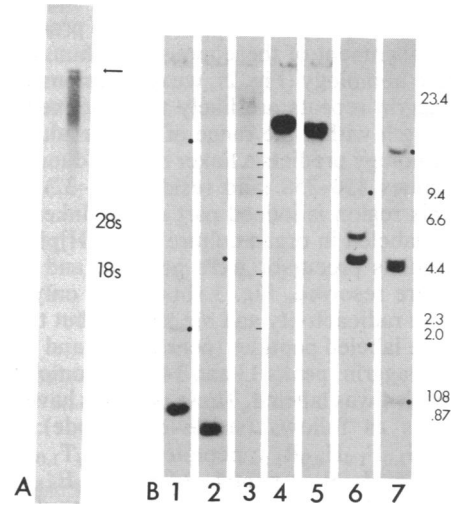

FIG. 4. (*A*) Northern blot of human epidermal poly(A)-enriched RNA probed with the λHF10 clone identifies a somewhat degraded but very large mRNA species with upper band (arrow) of ≈15 kb. The locations of the 28S and 18S ribosomal RNA species are shown. (*B*) Southern blots of human genomic DNA from a single placenta cut with *Hgi*AI (lane 1), *Pst* I (lane 2), *Sma* I (lane 3), *Dra* I (lane 4), *Eco*RV (lane 5), *Eco*RI (lane 6), *Sac* I (lane 7). The spots (·) in lanes 1, 2, 3, 7, and 8 mark minor bands that represent flanking sequences in relation to the superstoichiometric bands that delineate the existence of numerous filaggrin repeats. Lane 3 shows a partial digestion with *Sma* I demonstrating a ladder of at least 10 repeats.

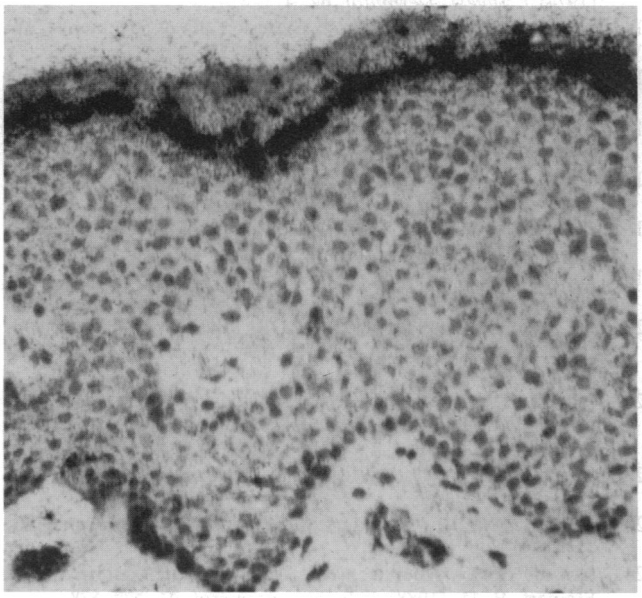

FIG. 5. *In situ* hybridization using antisense RNA probes from the λHF10 clone. The filaggrin probe decorates only a single cell layer high in the epidermis.

Transcriptional Control of Expression of the Human Filaggrin Precursor Gene. The cDNA clone was reconfigured into a pGEM-3B vector to generate sense and antisense RNA probes for *in situ* hybridization. Fig. 5 shows that the antisense orientation decorates a single cell layer high in full-thickness adult human epidermis, corresponding to the granular layer, indicating that filaggrin mRNA is actively transcribed only where the precursor accumulates as keratohyalin granules. In contrast, the keratin 1 gene is abundantly expressed throughout the suprabasal layers of the epidermis as expected (8, 27). These data indicate that filaggrin gene expression, in contrast to keratin 1 gene expression, is suppressed at the transcriptional level until very late in epidermal differentiation in normal epidermis, which therefore adequately explains why many hyperproliferative disorders of keratinization express diminished levels of filaggrin (16, 28). These RNA probes will now be useful in the further analysis of the expression of the filaggrin precursor during normal and abnormal epidermal differentiation.

Chromosomal Mapping and Localization of the Human Filaggrin Precursor Gene. DNA from a panel of mouse–human somatic cell hybrids retaining overlapping subsets of human chromosomal regions (see Fig. 6*A*) was digested with *Eco*RI and examined by Southern blots. The presence of the 5- and 6-kbp intermediate-sized fragments (Fig. 4*B*, lane 6), corresponding to the presence of unique human sequences was detected in five hybrids (Fig. 6*A*), all retaining chromosome 1. None of the hybrids that were negative for human filaggrin sequences retained chromosome 1, except for hybrid PB5, which only retains the short arm of chromosome 1. Thus, by somatic cell hybrid analysis, the human filaggrin precursor gene localizes to the long arm of chromosome 1. The precise assignment of the gene was established by chromosomal *in situ* hybridization (23, 24). The distribution of grains on 100 metaphases was assessed and is illustrated in Fig. 6*B*. Significant labeling occurs only over the proximal portion of the long arm of chromosome 1; 50.3% (200/398) of the total grains were found at the 1q21 region. This means that the polymorphism of the human filaggrin precursor gene is confined to a single locus at 1q21.

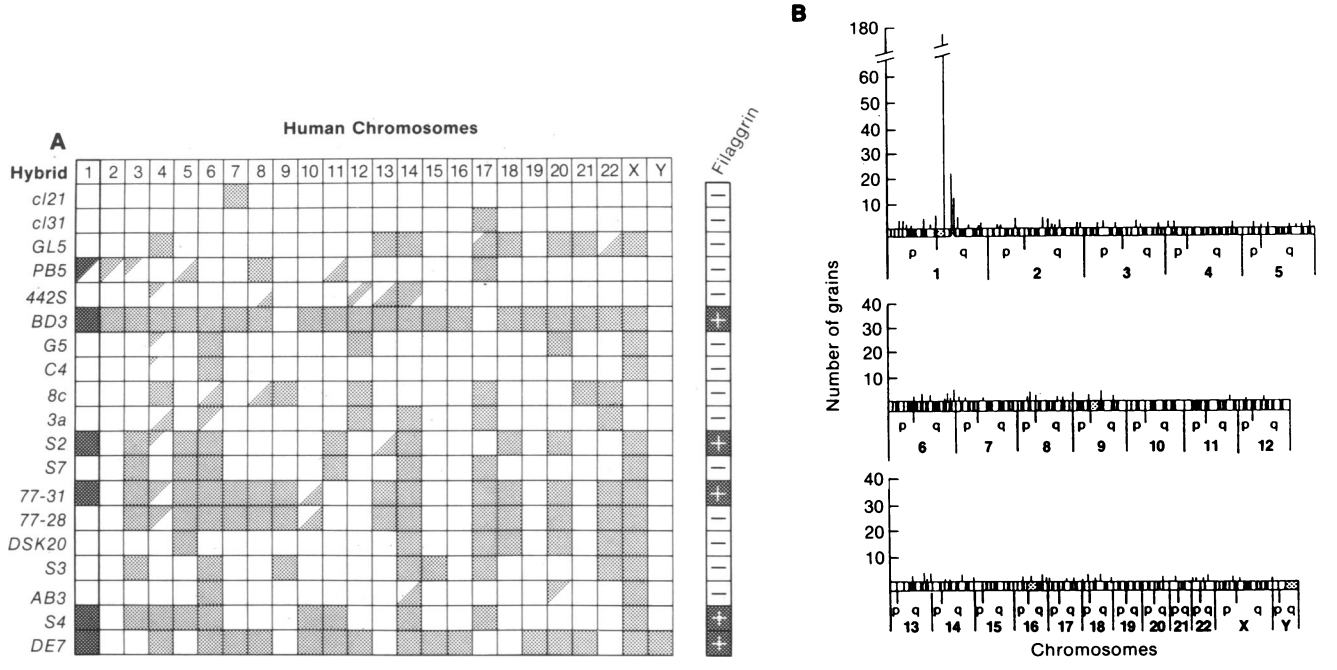

FIG. 6. (*A*) Presence of human filaggrin gene in a panel of 19 rodent–human hybrids. ■, The hybrid named in the left column contains the chromosome indicated in the upper row; ◪, presence of the long arm (or in some cases a part of the long arm, indicated by a smaller fraction of stippling) of the chromosome shown above the column; ◩, presence of the short arm of the chromosome listed above the column; □, absence of the chromosome listed above the column. The column for chromosome 1 is boldly outlined and stippled to emphasize correlation of this chromosome with the presence of filaggrin sequences. The pattern of retention of the filaggrin sequences is shown in the column to the right of the figure, where the presence of the sequences in the hybrid is indicated by a stippled box with a plus sign and the absence of the gene is indicated by an open box enclosing a minus sign. (*B*) Localization of the filaggrin gene to human chromosome region 1q21 by chromosomal *in situ* hybridization. Diagram shows the distribution of 398 grains over chromosomes of 100 metaphases. The abscissa represents the chromosomal banding pattern of each human chromosome in relative size proportion; the ordinate shows the number of silver grains. The highest grain count was found over the 1q21 region.

We are very grateful to Drs. Howard Baden and Joseph Kubilus (Massachusetts General Hospital) for their generous gift of the anti-filaggrin monoclonal antibody. We thank Dr. Joe Rothnagel for his assistance with the DNA sequencing and for numerous useful discussions. S.R.L. was supported by a Society for Investigative Dermatology Research Fellowship from the Dermatology Foundation and Grant AR01645 from the National Institutes of Health. C.M.C. was supported by Grant CA25875 from the National Institutes of Health.

1. Steinert, P. M. & Roop, D. R. (1988) *Annu. Rev. Biochem.* **57**, 593–625.
2. Goldman, R. D., Goldman, A. E., Green, K., Jones, J. C. R., Lieska, N. & Yang, S.-Y. (1985) *Ann. N.Y. Acad. Sci.* **455**, 1–17.
3. Franke, W. W. (1987) *Cell* **48**, 3–4.
4. Dale, B. A., Holbrook, K. A. & Steinert, P. M. (1978) *Nature (London)* **276**, 729–731.
5. Steinert, P. M., Cantieri, J. S., Teller, D. C., Lonsdale-Eccles, J. D. & Dale, B. A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4097–4101.
6. Steinert, P. M. (1983) in *Stratum Corneum*, eds. Marks, R. & Plewig, G. (Springer, Berlin), pp. 25–38.
7. Dale, B. A., Resing, K. A. & Lonsdale-Eccles, J. D. (1985) *Ann. N.Y. Acad. Sci.* **455**, 330–342.
8. Rothnagel, J. A., Mehrel, T., Idler, W. W., Roop, D. R. & Steinert, P. M. (1987) *J. Biol. Chem.* **262**, 15643–15648.
9. Harding, C. R. & Scott, I. R. (1983) *J. Mol. Biol.* **170**, 651–673.
10. Resing, K. A., Walsh, K. A. & Dale, B. A. (1984) *J. Cell Biol.* **99**, 1372–1378.
11. Resing, K. A., Dale, B. A. & Walsh, K. A. (1985) *Biochemistry* **24**, 4167–4176.
12. Fisher, C., Haydock, P. V. & Dale, B. A. (1987) *J. Invest. Dermatol.* **88**, 661–664.
13. Baden, H. P., Roth, S. I., Goldsmith, L. A. & Lee, L. D. (1974) *J. Invest. Dermatol.* **62**, 411–414.
14. Holbrook, K. A., Dale, B. A., Sybert, V. P. & Sagebiel, R. W. (1982) *J. Invest. Dermatol.* **80**, 222–227.
15. Holbrook, K. A., Dale, B. A. & Brown, K. S. (1982) *J. Cell Biol.* **92**, 387–397.
16. Sybert, V. P., Dale, B. A. & Holbrook, K. A. (1985) *J. Invest. Dermatol.* **84**, 191–194.
17. Kubilus, J., Scott, I., Harding, C. R., Yendle, J., Kvedar, J. & Baden, H. P. (1985) *J. Invest. Dermatol.* **85**, 513–517.
18. Dale, B. A., Gown, A. M., Fleckman, P., Kimball, J. R. & Resing, K. A. (1987) *J. Invest. Dermatol.* **88**, 306–313.
19. Haydock, P. V. & Dale, B. A. (1986) *J. Biol. Chem.* **261**, 12520–12525.
20. Steinert, P. M., Parry, D. A. D., Idler, W. W., Johnson, L. D., Steven, A. C. & Roop, D. R. (1985) *J. Biol. Chem.* **260**, 7142–7149.
21. Huynh, T. U., Young, R. A. & Davis, R. W. (1985) *DNA Cloning Techniques: A Practical Guide*, ed. Glover, D. (IRL, Oxford).
22. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
23. Harper, M. E. & Saunders, G. F. (1981) *Chromosoma* **83**, 431–439.
24. Lessin, S. R., Huebner, K., Isobe, M., Croce, C. M. & Steinert, P. M. (1988) *J. Invest. Dermatol.* **91**, 572–578.
25. Lynley, A. M. & Dale, B. A. (1983) *Biochim. Biophys. Acta* **744**, 28–35.
26. Steinert, P. M., Idler, W. W. & Goldman, R. D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4534–4538.
27. Roop, D. R., Lowy, D. R., Tambourin, P. E., Strickland, J., Harper, J. R., Balaschak, M., Spangler, E. F. & Yuspa, S. H. (1986) *Nature (London)* **323**, 822–824.
28. Dale, B. A., Holbrook, K. A., Kimball, J. R., Hoff, M. & Sun, T.-T. (1985) *J. Cell Biol.* **101**, 1257–1269.