Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 4683–4687, July 1985
Biochemistry

The sequence of a type II keratin gene expressed in human skin: Conservation of structure among all intermediate filament genes

(gene structure/DNA sequence/positive hybrid translation/protein structure)

ANGELA L. TYNER, MARGARET J. EICHMAN, AND ELAINE FUCHS

Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, IL 60637

Communicated by A. A. Moscona, April 12, 1985

ABSTRACT **We report here the coding sequence of the gene for a 56-kDa type II keratin (designated K6b). Using a subclone specific for a unique 3′ noncoding region of the encoded mRNA, we have shown that this gene is one of at least two 56-kDa keratin genes expressed in abundance in human epidermis. Segmenting the coding portion of this gene are eight introns, six of which are identically positioned with those of a distantly related type III intermediate filament gene (vimentin), and five of which are identically positioned with those of a distantly related type I gene (50-kDa keratin). These results indicate a common ancestral origin for all three classes of intermediate filament genes. All of the highly conserved intron positions are located within, but do not demarcate, the four central α-helical domains common to all intermediate filament polypeptides, suggesting that these genes were probably not created piecemeal by recombination-mediated linkage of separate structural domains as they presently are known.**

The cytoskeleton of most mammalian cells includes a prominent network of 8- to 10-nm filaments called intermediate filaments (IFs) (1). There are three major classes of IF subunits: type I and type II keratins, which form the backbone of the intermediate filaments in epithelial tissues, and type III subunits (neurofilament protein, glial filament protein, vimentin, and desmin), which comprise the IFs in all other tissues (2). Each of the three classes of subunits form a subfamily encoded by at least 10 different genes, which are differentially expressed in different tissues (3, 4). Whereas the type III subunits are usually expressed by themselves and can likely form homopolymers, specific pairs of type I and type II keratins are coordinately expressed in epithelial cells, and both subunit types seem to be required for keratin filament assembly (3, 4–8).

The three types of IF subunits are only very distantly related, sharing 25–30% homology with each other (9–17). Nonetheless, they all have nearly identical secondary structures and are capable of assembling into filaments with similar structures (2, 11, 15). Each IF polypeptide has four large helical domains that are involved in forming the coiled-coil backbone (most likely a dimer) of the IF subunit (2, 11, 15, 18). The nonhelical amino and carboxyl termini of the IF proteins may play a role in end-to-end and in lateral interactions, both of which are important for packing these cylindrical subunits into the 8 to 10-nm filament (13, 19).

Recently, the structures of two IF genes were reported: the hamster vimentin (type III) IF gene (10) and a type I gene encoding the human 50-kDa keratin designated K14 (refs. 20 and 21; see ref. 5 for the nomenclature of keratins). These genes were shown to have remarkably similar structures despite their low homologies in sequence. We now have completed the sequence of the gene for a member of the type II keratin subfamily, K6b, which is coordinately expressed with the gene for K14 in epidermal cells. Our results demonstrate that the structures of all three classes of IF genes are highly conserved. This close structural relation provides a clue to the evolutionary constraints exerted against perturbing the length of the α-helical portion of the IF polypeptide. In addition, our data form a base with which we can now begin to analyze the regulatory regions responsible for the differential expression of IF genes in different tissues and at different stages of differentiation and development.

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.

MATERIALS AND METHODS

Clones and Subclones. Two fragments of a 1620-base-pair (bp) 56-kDa keratin (K6a) cDNA clone (3, 9) were subcloned into plasmid PUC-8: (*i*) a *Pst* I fragment of 470 bp located at the extreme 5′ end of the K6a cDNA and (*ii*) a *Pst* I fragment of 444 bp beginning 394 bp 5′ from the TAA stop codon and extending 50 bp 3′ from the TAA. Radiolabeled cDNA probe was transcribed using reverse transcriptase (3).

A 283-bp *Sac* I–*Rsa* I fragment of the GK-2 genomic clone extending from 169 bp 3′ from the TGA stop codon to 68 bp 5′ from the polyadenylylation signal was subcloned in the 3′-to-5′ direction into plasmid SP64 (22). Radiolabeled RNA complementary to the coding strand was transcribed by using *Salmonella* SP6 polymerase.

RESULTS

Isolation, Nucleotide Sequence, and Predicted Structure of the Type II Keratin Gene Contained in Clone GK-2. When human genomic DNA digested with restriction endonuclease *Eco*RI is hybridized with ^{32}P-labeled probe made to a cloned K6a cDNA complementary to 63% of the coding region of a 56-kDa type II keratin mRNA of cultured human epidermal cells (12), about 10 fragments react with varying degrees of stringency (3). These putative type II genes were isolated from a human genomic library by plaque hybridization at reduced stringency (23) by using radiolabeled K6a cDNA as probe. One of the genomic clones, GK-2, was shown to hybridize with both the 5′ and the 3′ subcloned probes of K6a cDNA at increased stringency. An 8.8-kilobase (kb) *Eco*RI restriction fragment encompassing the gene contained in GK-2 was subcloned and then sequenced (24) by using the cloning method of Anderson (25).

The coding portion of the GK-2 gene was identified by aligning the gene sequence with K6a cDNA (12) and with a cDNA sequence of a mouse epidermal 60-kDa (K5) keratin (14). The gene contained in GK-2 is 6 kb in length and is predicted to encode a polypeptide chain of 562 amino acid residues (Fig. 1). The nucleotide sequence within the coding

Abbreviations: IF, intermediate filament; bp, base pair(s); kb, kilobase(s).

portion of GK-2 shares a high degree of homology with K6a (97%) and with the mouse K5 cDNA sequence (85%), but it is clearly distinct.

A 52-bp sequence that shares 77% homology with the partially sequenced 5′ noncoding segment of the mouse K5 cDNA is found 5′ upstream from the putative ATG translation initiation codon of GK-2. A promoter sequence, T-A-T-A-A (26), is located 95 nucleotides 5′ from this ATG codon.

The alignment of the GK-2 gene with either cDNA sequence predicts eight introns, noted by the triangles in Fig. 1. As indicated in lowercase letters, all of the introns begin with the sequence G-T and end with the sequence A-G.

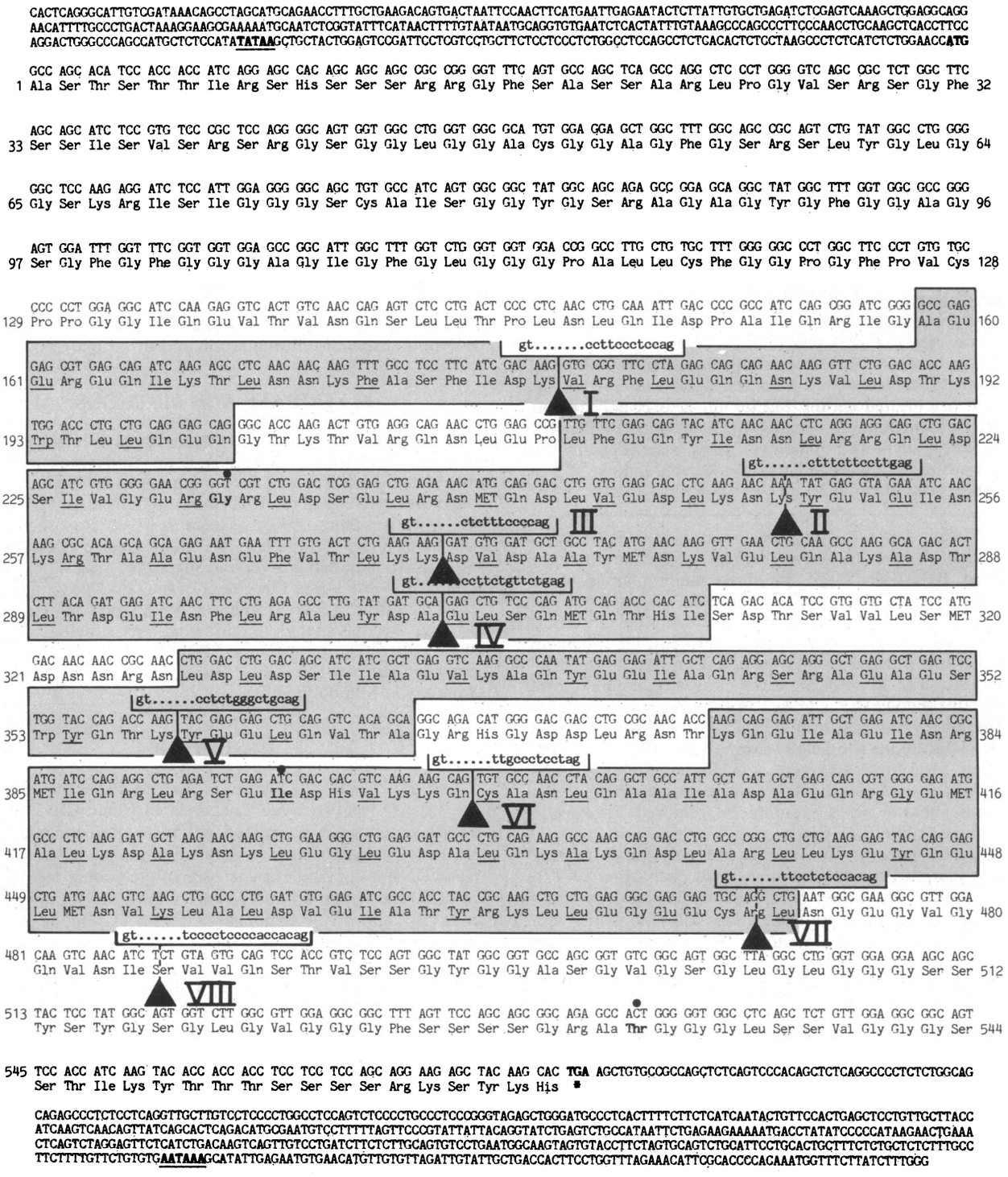

FIG. 1. Nucleotide sequence of the human type II keratin gene contained in genomic clone GK-2. The sequence of the gene and its 5′ and 3′ flanking regions are shown with 96 nucleotides per line. Intron positions are indicated by triangles. Intron–exon junctions and pyrimidine consensus sequences are shown for each intron in lowercase letters. The exons were identified by comparing the sequence of the 6-kb gene of GK-2 with K6a (12) and K5 (14) keratin cDNA sequences. The gray boxes mark the four α-helical domains in the GK-2 keratin. Throughout these domains are the heptad repeats of hydrophobic residues (underlined), which identify the portions of the polypeptide that are involved in coiled-coil interactions with a second keratin (2). The larger dots mark the 3-bp differences between GK-2 and the K6a cDNA that result in amino acid differences.

Immediately 5′ from each of these A-G sequences is a string of 7–12 pyrimidine residues. These features match well with the intron consensus characteristics that have been described previously (27).

At a position 1686 nucleotides 3′ downstream from the predicted translation initiation codon is a translation termination codon TGA. At 526-bp further downstream is found a single polyadenylylation signal, A-A-T-A-A-A (28). The 3′ noncoding segment of GK-2 is nearly identical in size to the 3′ noncoding portion of the K6a cDNA (Fig. 2). However, whereas the first 125 nucleotides of the 3′ noncoding segment of the GK-2 gene share a surprisingly high degree (92%) of homology with the corresponding segment of the previously sequenced human K6a cDNA, the remaining 400 bp are very divergent. In contrast, the 3′ noncoding segment of the mouse K5 cDNA is almost 200 bp shorter in size and shows only a low sequence homology with that of the GK-2 gene.

Positive Identification of the Gene in Clone GK-2 and Its Expression in Human Epidermal Cells. When *Eco*RI-digested human genomic DNA was hybridized with radiolabeled probe to the 3′ subclone of K6a cDNA encoding the carboxyl-terminal region of the 56-kDa keratin, four fragments (two at 8–9 kb are overlapping) were selected (Fig. 3, lane 1). The doublet band corresponded to the size of the *Eco*RI fragment containing the GK-2 gene (lane 2). Similarly, when human genomic DNA was digested with both *Eco*RI and *Kpn* I, four bands were selected (lane 3), one of which (5.8 kb) corresponded to the *Eco*RI/*Kpn* I fragment of GK-2 (lane 5). When subcloned DNA containing a highly specific 3′ noncoding segment of GK-2 (demarcated by arrows in Fig. 2) was used as probe, only a single human genomic *Eco*RI/*Kpn* I fragment (5.8 kb) was selected (Fig. 3, lane 4). Thus, there seems to be a single copy of the GK-2 gene in the genome, and the 3′ noncoding region subclone is selective for this gene and its transcript.

When total mRNA from cultured human epidermal cells was hybridized with probe to the 3′ noncoding end of GK-2, a 2.1-kb epidermal mRNA was selected (Fig. 4*A*). Even when this hybrid was subjected to increasing temperature washes from 50°C to 80°C, some of the probe remained bound to the nitrocellulose filter. Under the equivalent conditions in solution, a perfect RNA–RNA hybrid of this sequence would have had a melting temperature (T_m) of ≈79°C (31).

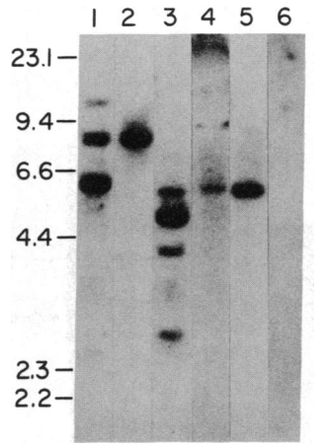

FIG. 3. Identification of a single gene encoding the human GK-2 keratin. Ten micrograms of total human DNA (lanes 1, 3, and 4), 500 ng of the subclone containing the 8.8-kb *Eco*RI fragment of clone GK-2 (lanes 2 and 5), and 500 ng of K6a cDNA (lane 6) were digested with restriction endonucleases *Eco*RI (lanes 1, 2, and 6) or *Eco*RI and *Kpn* I (lanes 3–5), and the fragments were resolved by agarose gel electrophoresis. Southern blots (29) were hybridized with ^{32}P-labeled probe made to either the subcloned *Pst* I fragment of K6a cDNA (see text) (lanes 1–3 and 5) or the subcloned 3′ noncoding portion of GK-2 (lanes 4 and 6). Size markers are indicated in kb at the right.

When either the entire 8.8-kb GK-2 clone or the 3′ noncoding subclone was denatured and hybridized at high stringency with total epidermal mRNA, an mRNA was selected that translated a 56-kDa keratin (K6b) *in vitro* (Fig. 4*B*, lanes 3–5). Since K6a cDNA also selects a 56-kDa keratin mRNA (6), there must be genes for at least two type II 56-kDa keratins, K6a and K6b, expressed in abundance in human epidermal cells. A detailed report of the relative amounts of the two K6 mRNAs synthesized in different epithelial cells will be published elsewhere.

The Complete Amino Acid Sequence and Predicted Second-

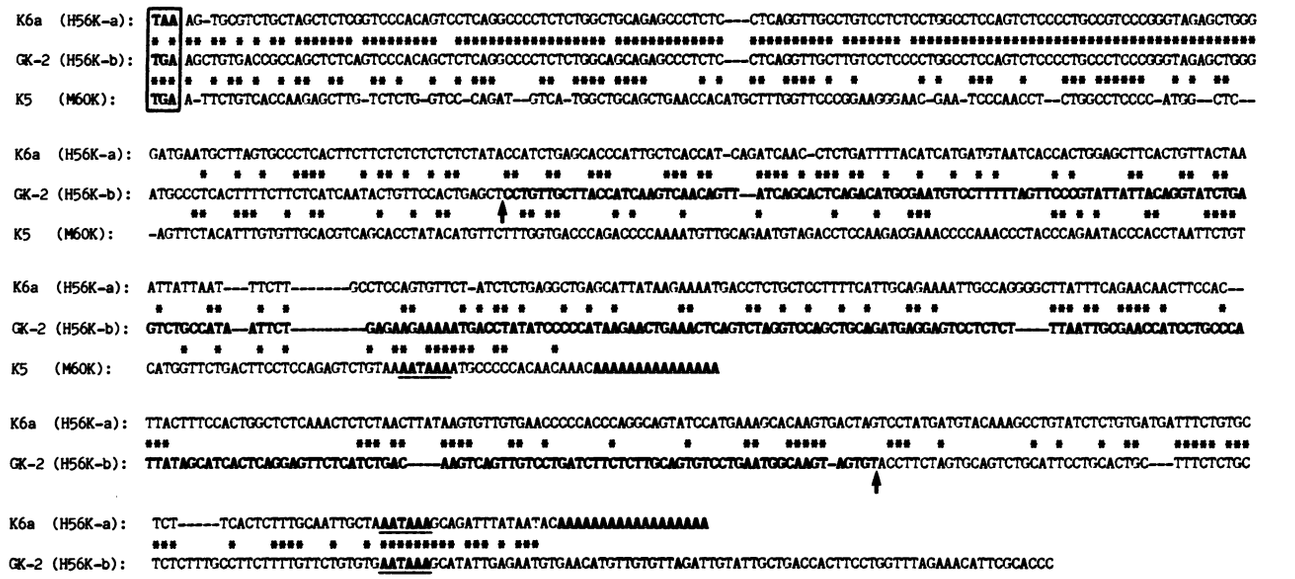

FIG. 2. Comparison of the 3′ noncoding portions of the keratin mRNAs encoded by the human GK-2 gene, the human K6a cDNA, and the mouse K5 cDNA. The noncoding segments 3′ to the translation termination codons of the three indicated sequences were aligned for optimal homologies. The asterisks indicate identical nucleotide residues between K6a and GK-2 and between GK-2 and the mouse K5 cDNA. Arrows demarcate the 283-bp *Sac* I/*Rsa* I restriction endonuclease fragment (in bold-face type) that was subcloned and used as a specific probe for the gene in GK-2.

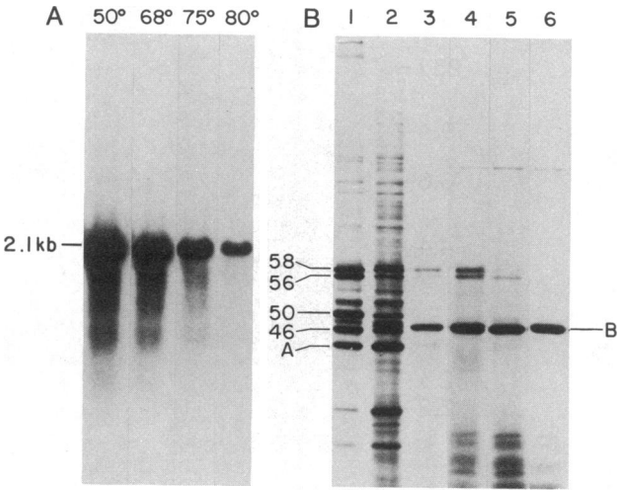

FIG. 4. Positive identification of clone GK-2 as a gene encoding one of at least two type II 56-kDa keratins expressed in abundance in human epidermal cells. (*A*) Poly(A)$^+$ RNA was isolated from cultured human epidermal cells and resolved by formaldehyde/RNA gel electrophoresis as described (3). After blot-transfer (30), the RNAs were hybridized with ^{32}P-labeled cRNA transcribed from the 3′ noncoding subclone of GK-2. This clone shows no crossreactivity with K6a cDNA (Fig. 3, lane 6). Hybridization was carried out in 50% formamide/750 mM NaCl, 41°C, followed by sequential washes in 0.1% $NaDodSO_4$, 15 mM NaCl, and 1.5 mM sodium citrate at 50°C, 68°C, 75°C, or 80°C. After each wash, the filter was exposed to x-ray film for 90 min. (*B*) DNAs from the entire GK-2 clone and the 3′ noncoding subclone of GK-2 were denatured and bound to a nitrocellulose filter (3). Filters were hybridized for 36 hr at 41°C with human epidermal mRNA. Specifically hybridizing RNAs were eluted at increasing temperatures and translated in a reticulocyte cell-free system. Translation products were resolved by $NaDodSO_4$/polyacrylamide gel electrophoresis. Lanes: 1, [^{35}S]methionine-labeled epidermal keratin extract; 2, radiolabeled translation products from total epidermal mRNA; 3 and 4, translation products from epidermal mRNAs that were hybrid-eluted from GK-2 DNA at 65°C and 85°C, respectively; 5, translation products from epidermal mRNAs that were hybrid-eluted from the 3′ noncoding subclone of GK-2 at 85°C; 6, translation with no added RNA. Molecular masses are indicated in kDa at the right.

ary Structure of the Human K6b Type II Keratin Encoded by the Gene in Clone GK-2. The complete sequence of the GK-2 gene yields a predicted molecular mass of 60,018 Da. Previous sequence analysis of a 50-kDa (K14) type I keratin gene had already indicated that the original estimates based on $NaDodSO_4$/polyacrylamide gel electrophoresis (32) are substantially smaller than the values calculated by actual sequence data (20, 21). The predicted amino acid composition matches very well with that predicted by chemical means (12).

The amino acid sequence of the K6b keratin encoded by the gene in GK-2 is highly similar to the sequence of K6a. Only three amino acid residues are different, and none of these changes involves charged residues. Thus, it is not surprising that only one K6 spot is observed by two-dimensional gel electrophoresis of *in vitro* translation products from epidermal mRNA (6). In contrast, mouse K5, reportedly homologous to the human 58-kDa keratin (14), shows 76 variations in 562 residues when compared with the K6b keratin. Most of these changes are conservative ones (e.g., Asp → Glu or Thr → Ser).

As judged by the probabilistic method of Robson and co-workers (9, 12, 33), K6b contains four central α-helical domains characteristic of all IF proteins (marked by the gray boxes in Fig. 1). Throughout the helical domains of K6b is a heptad repeat, in which most *a* and *d* residues (underlined in Fig. 1) of every seven in the series *a* b c *d* e f g are hydrophobic. A similar repeat was initially observed for the coiled-coil helices of tropomyosin and has been identified since for the helical domains of all IF proteins (2, 9–18).

The amino- and carboxyl-terminal segments that flank the central helical domain are not compatible with α-helix or β-sheet conformations. They are particularly rich in glycine and serine residues and frequently contain inexact repeats of Gly-Gly-Gly-X, where X is either phenylalanine, tyrosine, leucine, or alanine. Similar sequences have been found previously in the amino- and/or carboxyl-terminal segments of all epidermal keratins (9, 12–14).

The Structures of All IF Genes Are Remarkably Similar. When the sequences of the hamster vimentin (10) and human K14 (20, 21) genes were aligned with that of the human K6b gene for optimal homology, the positions of the introns were found to be highly similar (Fig. 5). The conservation of intron position is quite striking, considering that all three types of IF proteins share only 23–25% amino acid sequence homology with one another (9, 10, 20). It provides a strong indication that all three types of intermediate filament genes had a common origin.

Seven of the introns (Fig. 1, introns II–VIII) show identical or nearly identical positioning for all three proteins. These introns are located in the central portion of the coding region but do not seem to demarcate any of the structural domains of the polypeptides. Despite the strict conservation of intron position, neither the size nor the sequence of the introns has been conserved.

No correlation was found between the regions of conserved sequence in the IF proteins and the location of the introns in the IF genes. In some cases, the conservation within an exon is fairly high, but in other cases, it is very low. Sometimes, regions of high and low homologies, respectively, are contained within a single exon. Not even the amino acid sequences immediately flanking the exon–intron junctions have been evolutionarily conserved. Thus, for the IF genes, exons do not seem to have evolved as separate units, even though the evolutionary pressures on maintaining exon size have been considerable.

Comparison of the 5′ Noncoding and 5′ Upstream Regions of the Coordinately Expressed K14 and K6b Genes. When the transcribed 5′ noncoding portions of the genes encoding the type II and type I epidermal keratins were compared, two 20-

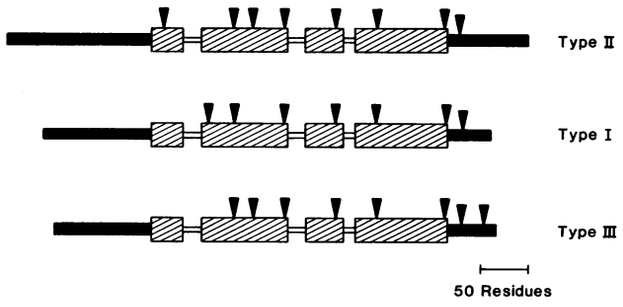

FIG. 5. Conservation of structure among all IF proteins and all IF genes. The secondary structures of the three IF proteins were determined from their amino acid sequences as described in the text. The complete amino acid sequences were obtained from the nucleic acid sequences of their genes: human K6b type II keratin (this paper); human K14 type I keratin (20, 21); and type III, hamster vimentin (10). The three sequences were aligned for optimal homology and drawn to scale, with the scale representing amino acid residue position. The boxed hatched regions indicate the four α-helical domains, and the thin connecting lines indicate the presence of helix-disrupting sequences. The thick black lines designate the nonhelical termini, which are variable in length. The positions of the introns are shown by arrowheads. Note that the intron positions do not demarcate the structural boundaries of the IF polypeptide chains.

to 30-bp stretches in comparable positions were found to share 75–80% homology. In contrast, much less homology was found between the transcribed 5′ noncoding regions of the K6b type II and vimentin type III genes. Similarly, when the sequences within a region 230 bp 5′ upstream from the "TATA" boxes of the type II and type I keratin genes were examined, three homologous (75–77%) stretches of 15–22 bp were found. No such homologies were found between these segments of the type III and type II genes. Whether any of these features play a role in the coordinate regulation of the two epidermal keratin genes or the differential regulation of the vimentin gene remains to be investigated.

DISCUSSION

The sequence of a type II keratin gene has made it possible to compare the structures of all three classes of IF genes. Our results clearly identify a common structure for all IF genes, a feature that was recently suggested from heteroduplex analyses (34). Two very distant duplications of a primordial IF gene most likely led to the early formation of three genes: a type I and type II keratin gene and a third (type III) IF gene. The multiplicity of related sequences within these three classes seems to have arisen from more recent gene duplications.

For the two keratin genes that we have now sequenced, there seem to be only single transcription-initiation and polyadenylylation sites. Furthermore, it is unlikely that differential usage of exon–intron splice junctions can occur within the internal intron sequences of a keratin heterogeneous nuclear RNA transcript, since our data indicate a strong evolutionary pressure against shifting the intron positions within the IF genes. Although these results do not exclude the possibility of differential splicing at the junctions of the external IF introns, we have not found any cases where multiple keratin mRNAs arise from a single gene. Because we now have demonstrated that there are at least two genes encoding K6 keratins, there may be more than the initial estimate of 20 keratin genes for the 20 keratin polypeptides of human epithelial cells (3).

The conservation of intron position in gene families has frequently provided clues to the piecemeal construction of a primordial gene by the recombination-mediated assembly of separate structural or functional domains (35–37). However, since the positions of introns in the IF genes do not seem to demarcate segments coding for distinct structural domains of the IF polypeptides, nor do they define any obvious functional domains of the protein, we have no evidence to suggest that the primordial IF gene was put together in this fashion. Nonetheless, the conservation of gene structure must indicate that there are important constraints exerted against sliding the introns within these genes.

Utilizing an alternative splice junction at a position $3n$ nucleotides from the original G-T or A-G consensus sequence would introduce or delete one or more amino acid residues at the intron–exon border. Since these junctions are located within the helical domains of the IF polypeptide, such changes might disrupt the heptad repeat which is necessary for the proper formation of the coiled coil. In turn, perturbation in this structure may alter the ability of the subunits to pack into filaments, thus making any change in the exon length deleterious to the subunit's function.

Although the evolutionary pressures exerted on maintaining exon length most certainly reflect at least in part the complex interactions of the coiled-coil subunits in forming the filamentous structure, it is also possible that as yet unexplored protein-, membrane-, or organelle-IF interactions might place restrictions on the overall design of the IF subunit. Unlike actins and tubulins, there are no known auxiliary proteins or factors necessary for filament assembly. However, there might be domains in the IF subunits important for interactions that determine overall cytoskeletal architecture. For the moment, the unequivocal role of subunit structure or function on maintaining IF gene structure must await further studies.

Note Added in Proof. Johnson *et al.* (38) have recently reported the sequence of a mouse type II keratin gene.

We thank Dr. Ed Fritsch (Genetics Institute) for providing us with a partial *Hae* III/*Alu* I human genomic library in phage λ Charon 4A. We thank Drs. Don W. Cleveland and Kevin Sullivan (Johns Hopkins University) and Dr. Kwan Hee Kim and Douglas Marchuk (our laboratory) for their valuable discussions and advice. This work was supported by a grant funded by the National Institutes of Health (R01-AM 27883). E.F. is a Presidential Young Investigator and the recipient of a National Institutes of Health Career Development Award. A.L.T. is a National Institutes of Health Predoctoral Trainee.

1. Lazarides, E. (1982) *Annu. Rev. Biochem.* **51,** 219–250.
2. Fuchs, E. & Hanukoglu, I. (1983) *Cell* **34,** 332–334.
3. Fuchs, E., Coppock, S., Green, H. & Cleveland, D. (1981) *Cell* **27,** 75–84.
4. Steinert, P. M., Idler, W. W. & Zimmerman, S. B. (1976) *J. Mol. Biol.* **108,** 547–567.
5. Moll, R. Franke, W. W., Schiller, D., Geiger, B. & Krepler, R. (1982) *Cell* **31,** 11–24.
6. Kim, K. H., Rheinwald, J. G. & Fuchs, E. (1983) *Mol. Cell. Biol.* **3,** 495–502.
7. Franke, W. W., Schiller, D. L., Hatzfeld, R. & Winter, S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7113–7117.
8. Nelson, W. G. & Sun, T.-T. (1983) *J. Cell Biol.* **97,** 244–251.
9. Hanukoglu, I. & Fuchs, E. (1982) *Cell* **31,** 243–252.
10. Quax, W., Egberts, W. V., Hendriks, W., Quax-Jeuken, Y. & Bloemendal, H. (1983) *Cell* **35,** 215–233.
11. Geisler, N. & Weber, K. (1982) *EMBO J.* **1,** 1649–1656.
12. Hanukoglu, I. & Fuchs, E. (1983) *Cell* **33,** 915–924.
13. Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L. & Steven, A. C. (1983) *Nature (London)* **302,** 794–800.
14. Steinert, P. M., Parry, D. A. D., Racoosin, E. L., Idler, W. W., Steven, A. C., Trus, B. L. & Roop, D. R. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5709–5713.
15. Crewther, W. G., Dowling, L. M., Steinert, P. M. & Parry, D. A. D. (1983) *Int. J. Biol. Macromol.* **5,** 267–274.
16. Geisler, N., Fischer, S., Vandekerckhove, J., Plessmann, U. & Weber, K. (1984) *EMBO J.* **3,** 2701–2706.
17. Lewis, S. A., Balcarek, J. M., Krek, V., Shelanski, M. & Cowan, N. J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2743–2746.
18. McLachlan, A. D. (1978) *J. Mol. Biol.* **98,** 293–304.
19. Skerrow, D., Matoltsy, G. & Matoltsy, M. (1973) *J. Biol. Chem.* **248,** 4820–4826.
20. Marchuk, D., McCrohon, S. & Fuchs, E. (1984) *Cell* **39,** 491–498.
21. Marchuk, D., McCrohon, S. & Fuchs, E. (1985) *Proc, Natl. Acad. Sci. USA* **82,** 1609–1613.
22. Green, M. R., Maniatis, T. & Melton, D. A. (1983) *Cell* **32,** 681–694.
23. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
24. Sanger, F., Coulson, A. R., Barrel, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143,** 161–178.
25. Anderson, S. (1981) *Nucleic Acids Res.* **9,** 3015–3026.
26. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
27. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
28. Fitzgerald, M. & Shenk, T. (1981) *Cell* **24,** 251–260.
29. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
30. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5201–5205.
31. Casey, J. & Davidson, N. (1977) *Nucleic Acids Res.* **4,** 1539–1552.
32. Sun, T.-T. & Green, H. (1978) *J. Biol. Chem.* **253,** 2053–2060.
33. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.
34. Lehnert, M. E., Jorcano, J. L., Zentgraf, H., Blessing, M., Franz, J. K. & Franke, W. W. (1984) *EMBO J.* **3,** 3270–3287.
35. Gilbert, W. (1978) *Nature (London)* **277,** 501.
36. Yamada, Y., Avvedimento, V. E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. & de Crombrugghe, C. (1980) *Cell* **22,** 887–892.
37. Lonberg, N. & Gilbert, W. (1985) *Cell* **40,** 81–90.
38. Johnson, L. D., Idler, W. W., Zhou, X.-M., Roop, D. R. & Steinert, P. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1896–1900.