MOLECULAR AND CELLULAR BIOLOGY, Feb. 1986, p. 539–548
0270-7306/86/020539-10$02.00/0

Vol. 6, No. 2

Three Tightly Linked Genes Encoding Human Type I Keratins: Conservation of Sequence in the 5′-Untranslated Leader and 5′-Upstream Regions of Coexpressed Keratin Genes

AMLAN RAYCHAUDHURY, DOUGLAS MARCHUK, MARJORIE LINDHURST, AND ELAINE FUCHS*

Departments of Molecular Genetics and Cell Biology and Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637

Received 27 August 1985/Accepted 28 October 1985

We have isolated and subcloned three separate segments of human DNA which share strong sequence homology with a previously sequenced gene encoding a type I keratin, K14 (50 kilodaltons). Restriction endonuclease mapping has demonstrated that these three genes are tightly linked chromosomally, whereas the K14 gene appears to be separate. As judged by positive hybridization-translation and Northern blot analyses, the central linked gene encodes a keratin, K17, which is expressed in abundance with K14 and two other type I keratins in cultured human epidermal cells. None of these other epidermal keratin mRNAs appears to be generated from the K17 gene through differential splicing of its transcript. The sequence of the K17 gene reveals striking homologies not only with the coding portions and intron positions of the K14 gene, but also with its 5′-noncoding and 5′-upstream sequences. These similarities may provide an important clue in elucidating the molecular mechanisms underlying the coexpression of the two genes.

The keratins form a group of 20 related polypeptides (40 to 70 kilodaltons [kDa]) which comprise 8-nm cytoplasmic filaments in most if not all epithelial cells (for review, see reference 46). Although the proteins can be subdivided into two distinct sequence classes, type I and type II, all keratins have very similar secondary structures (8, 9, 19, 21–28, 60, 62, 63). The central 300-amino-acid portion is relatively constant in length and consists of several large α-helical domains, the sequences of which determine the classification of keratins according to type (25). These helical segments contain heptad repeats of hydrophobic residues, indicative of their ability to form coiled-coil subunits (22, 40, 41). The nonhelical termini are variable in length and in sequence and seem to play a role in lateral and end-to-end interactions necessary to pack the coiled-coil subunits into the filamentous structure (61, 62–64).

At least one member of each of the two keratin classes is expressed in all tissues, suggesting the importance of the two types of sequences in filament assembly (8, 14, 15, 30, 49, 59). Coexpressed keratins seem to have similar nonhelical terminal sequences. Cultured epidermal cells, for example, synthesize two to three pairs of type I and type II keratins, each having nonhelical termini rich in glycine and serine (16, 52, 65). When these cells undergo terminal differentiation, a shift to the synthesis of unusually large keratins takes place without disrupting the balance of type I and type II keratins (4, 13, 18, 29, 46, 56). These large epidermal keratins seem to differ from the smaller ones primarily in having extended glycine- and serine-rich termini (26, 62). In contrast, many other epithelial tissues express pairs of keratins with termini distinctly different from glycine- and serine-rich sequences of the epidermal keratins (8, 9, 23a). Although the role of the tissue-specific variability in the nonhelical ends of the keratins is not yet fully understood, these differences may enable the properties of the resulting filaments to be tailored to suit the particular protective needs of each epithelial cell.

The existence of constant central domains and variable terminal domains in the keratin polypeptide chains has led to the question of whether different keratin mRNAs might arise from differential processing of a smaller number of heterogeneous nuclear RNA transcripts (17, 18, 30, 35). Thus far, at least some keratins of each type have been shown to be encoded by separate genes (15, 26, 32, 36, 37, 67). Nonetheless, the features that determine the number and sequences of the transcripts encoded by these genes have not yet been explored in depth. To begin to elucidate the genetic complexity and organization of the keratin gene family and to investigate the possible regulatory mechanisms underlying keratin gene expression, we have examined a subfamily of closely related genes that seem to encode the type I keratins expressed in abundance in cultured human epidermal cells.

MATERIALS AND METHODS

Construction of subclones. (i) From KB-2 cDNA. KB-2, a 1,390-base-pair (bp) cDNA corresponding to the human 50-kDa (K14) type I keratin was isolated and sequenced as previously described (24). Restriction endonuclease fragments containing the 5′- and 3′-coding and the 3′-noncoding segments of KB-2 were isolated and subcloned into plasmids pUC8 (68) and later pSP65 (44). For the 5′-coding probe, we subcloned a 327-bp *Alu*I-*Pst*I fragment encompassing residues 25 through 352 at the 5′ end of KB-2. For the 3′-coding probe, we subcloned a 281-bp *Alu*I-*Stu*I fragment encompassing residues 947 through 1228 of KB-2. The TGA stop codon is positioned at residue 1,230. The 3′-noncoding probe was prepared by subcloning a 70-bp *Stu*I-*Alu*I fragment encompassing residues 1228 through 1298 of KB-2 into *Escherichia coli* vector pSP65, containing the *Salmonella* sp. SP6 polymerase promotor upstream from the polylinker region (44).

(ii) From genomic clones. Restriction endonuclease fragments (described below) containing portions of four type I keratin genes *A* through *D* were isolated from a λ Charon 4A library (Ed Fritsch, Genetics Institute, Boston) and subcloned into plasmid pSP64 or pSP65. A 4.25-kilobase (kb)

* Corresponding author.

*Bam*HI fragment containing gene *C* (encoding K17) was subcloned into plasmid pUC9 for DNA sequence analysis.

For positive hybridization analyses, a 931-bp *Xba*I-*Sac*I restriction endonuclease fragment containing part of intron VII, exon VIII, and the 3'-noncoding portion of gene *A* (encoding K14) was subcloned from a human genomic clone, GK-1. A 658-bp *Pst*I-*Eco*RI fragment containing the corresponding portion of gene *C* was subcloned from a second genomic clone, GK-3.

DNA sequence analyses. All DNA sequence analyses were conducted with the M13 dideoxy strategy (54) and the shotgun cloning method of Anderson (1). To obtain sequence information for the coding strand of gene *C*, the plasmid clone containing the 4.25-kb *Bam*HI fragment was first treated with *Kpn*I and then digested further by the action of exonuclease *Bal* 31 (1). Periodically, samples were removed, and the reaction was stopped. These fragments were subsequently treated with *Hin*dIII, and the resulting fragments were inserted into phage M13mp19 DNA (digested with *Hin*cII and *Hin*dIII) and transformed into *E. coli* JM101 (45). To obtain sequence information for the complementary strand, the clone was first treated with *Hin*dIII and then digested further with *Bal* 31 as described above. These fragments were then treated with *Kpn*I and inserted into phage M13mp18 DNA (digested with *Hin*cII and KpnI) and transformed in JM101. Positive clones were isolated and sequenced. Multiple overlapping and opposite strands were sequenced. Very few ambiguities arose in reading the sequences, and these were always resolved by additional sequencing.

RESULTS

Isolation and mapping of human genomic clones containing sequences related to keratin K14. When human genomic DNA was digested with restriction endonuclease *Eco*RI, eight fragments were generated which hybridized with various degrees of stringency to a K14 (50-kDa) epidermal keratin cDNA, KB-2 (Fig. 1, lane 1) (15, 24). These putative type I keratin genes were isolated by screening a human genomic library in λ Charon 4A. Recombinant phage were screened by plaque hybridization by the procedure of Benton and Davis (3). ^{32}P-labeled human K14 cDNA (KB-2) was used as a probe (15). This cDNA contains approximately 85% of the coding portion of the K14 keratin (1,380 bp), and its predicted amino acid sequence has been determined by DNA sequence analysis (24).

Eight genomic clones, labeled GK-1 to GK-8, were isolated and shown to hybridize strongly with the total K14 cDNA probe and with sublconed probes to either the 3'- or 5'-coding portions of K14 cDNA. Some of the clones were shown to contain more than one gene. A map of all of the KB-2 hybridizing clones is given in Fig. 2. Three of the four genes, labeled *B* through *D* in Fig. 2A, were found to be tightly linked chromosomally and in a head-to-tail arrangement. Only 5 to 6.5 kb of spacer DNA separates each of the genes. Thus far, we have not been able to establish any linkage with the fourth gene, labeled *A* in Fig. 2B, which was previously shown to contain a coding region identical to the K14 mRNA (36, 37). If indeed gene *A* is linked to the other three, it may be more distantly so, since for at least 10 kb of DNA beyond the 3' end of this gene, and for 5 kb beyond the 5' end of this gene, there are no other sequences that hybridize with the K14 cDNA.

To determine which of the eight human genomic *Eco*RI fragments correspond to which of these four keratin genes, we subcloned into plasmid SP64 a number of keratin gene fragments (marked by the overhead bars in Fig. 2). Each of these subcloned fragments was radiolabeled and used to probe human genomic DNA that was digested with restriction endonuclease *Eco*RI (Fig. 1, lanes 2 through 11). When the Southern hybridizations of Fig. 1 are coupled with the mapping data of Fig. 2, it is apparent that (i) gene *A* is most likely contained in a large fragment of at least 15 kb (Fig. 1, lanes 10 and 11; the transfer of this large fragment to nitrocellulose paper seemed to be incomplete, thereby reducing the relative level of hybridization); (ii) genes *B* and *C* are contained in a 12-kb fragment (lanes 2 through 4); and (iii) gene *D* is contained within two *Eco*RI fragments of 6.4 and 5.6 kb (lanes 5 through 9). Genes *A* and *D* appear to be highly homologous, since both the 5' and the 3' probes for gene *A* hybridized strongly with the corresponding gene *D* segments (lanes 10 and 11).

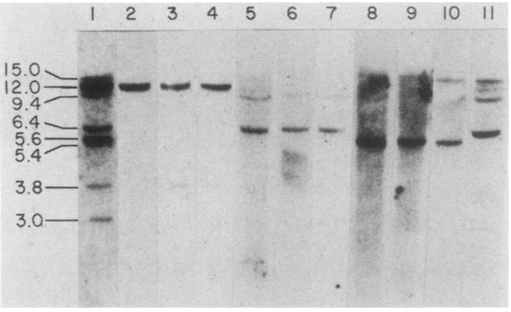

FIG. 1. Southern hybridization analysis of the type I keratin genes in the human genome. Human genomic DNA was isolated, digested with *Eco*RI, and subjected to agarose gel electrophoresis. After transfer to nitrocellulose paper (57), each lane of the blot was hybridized with a different radiolabeled probe to each of the gene segments that are indicated by the overhead bars in Fig. 2. Probes: 1, KB-2 cDNA; 2, gene *B*, subcloned from human genomic clone GK-3; 3, gene *C*, subcloned from GK-3; 4, gene *C*, subcloned from GK-6; 5, 5' portion of gene *D* isolated from GK-6; 6, 5' portion of gene *D* isolated from GK-7; 7, 5' portion of gene *D* from GK-8; 8, 3' portion of gene *D* from GK-7; 9, 3' portion of gene *D* from GK-8; 10, 3' portion of gene *A* subcloned from GK-1; 11, 5' portion of gene *A* subcloned from GK-1. Lane 1 was washed at 50°C with 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate to identify most if not all of the type I keratin gene family. Subsequent lanes were washed at 65°C to minimize cross-hybridization. Fragment sizes are given at left in kilobase pairs.

Although some of the *A* through *D* probes showed hybridization with the *Eco*RI fragments of 9.4, 3.8, and 3.0 kb, these hybridizations were generally weaker, and it seems most likely that these fragments represent additional type I genes which are more distantly related and as yet uncharacterized.

At least two of the four genes encode a subfamily of related but distinct epidermal keratins. Previous sequence analyses of gene *A* (36, 37) and K14 cDNA (24) demonstrated unequivocally that gene *A* encodes keratin K14, expressed in abundance in cultured human epidermal cells. To determine what keratins might be encoded by the closely related genes *B* through *D*, we conducted positive hybridization and translation analyses with linearized subcloned plasmid DNAs containing portions of these three genes (Fig. 3). Since the K14 cDNA has been shown to hybridize strongly with some of the other type I epidermal keratin mRNAs, but only weakly with nonepidermal type I mRNAs (30), we chose cultured human epidermal mRNA for hybridization. Under conditions of reduced stringency, clones containing genes *B* through *D* behaved similarly to gene *A* in hybridizing

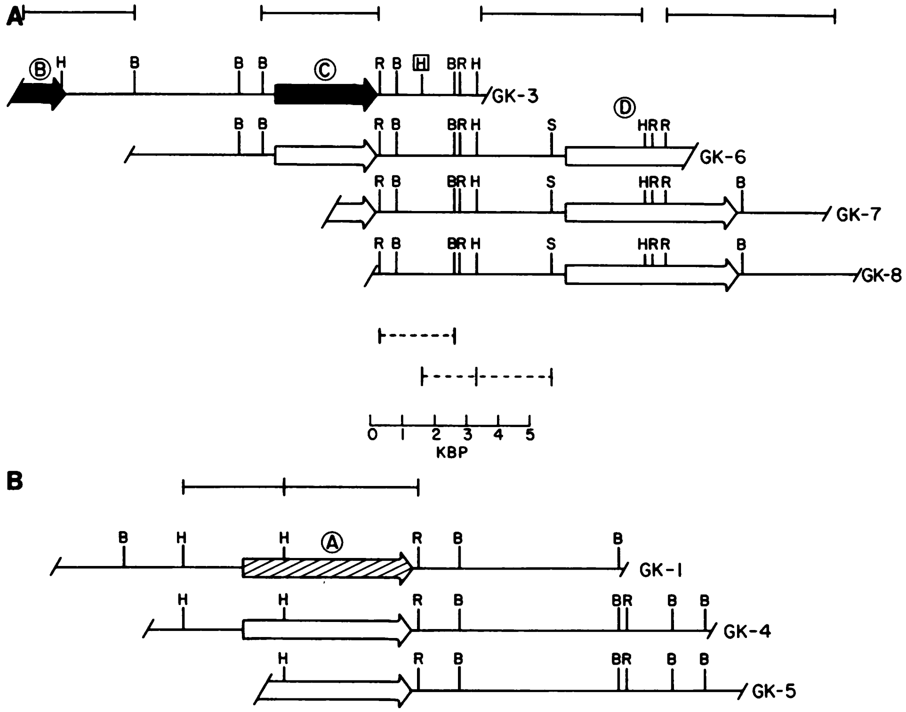

FIG. 2. Chromosomal organization of the human type I keratin genes containing sequences similar to keratin K14. Human genomic clones GK-1 to GK-8 were digested with different restriction endonucleases, and the fragments were fractionated by agarose gel electrophoresis. After transfer of the resolved fragments to nitrocellulose paper (57), the blots were hybridized with radiolabeled probe either to the total KB-2 cDNA or to subclones containing the 5′ or 3′ end of the KB-2 sequence. Two sets of overlapping clones are shown; one (A) contains three linked genes (labeled *B*, *C*, and *D*), and one (B) contains a single gene (labeled *A*). The thick arrows mark the approximate boundaries of each gene, and the direction of the arrow indicates the 5′-to-3′ direction of the coding strand. Note that gene *B* is only partially cloned in the cluster. It is assumed that the remainder of the gene resides in the sequences upstream from this region which have not yet been isolated. The solid arrows represent regions of the genome which have been sequenced in this report. The hatched arrow represents previous sequence analysis (36, 37). Restriction endonuclease sites: B, *Bam*HI; H, *Hin*dIII; R, *Eco*RI; S, *Sac*I. The diagonal lines mark the synthetic *Eco*RI sites of the vector. The *Hin*dIII site enclosed in a box in clone GK-3 was not found in clones GK-6, GK-7, and GK-8. The solid bars above the genes represent fragments subcloned for probes used in Southern hybridizations of human genomic DNA (Fig. 1) and in positive hybridization-translation analyses (Fig. 3). The dashed lines below the genes in A represent fragments subcloned for probes in Northern blot analyses (Fig. 7). The size scale is shown in kilobase pairs (KBP).

with all of the type I epidermal keratin mRNAs. Most of the mRNA encoding keratin K14 was not eluted from any of the clones until the temperature was raised to 85°C (Fig. 3, second lane of each set), indicating that all of the clones shared very strong homology with the 50-kDa keratin mRNA. At lower wash temperatures (65°C; first lane of each set), the mRNA encoding keratin K13 (52 kDa) was frequently eluted, whereas the mRNAs encoding the other two epidermal keratins K16 (48 kDa) and K17 (46 kDa) showed different degrees of hybridization; clones containing portions of genes *A* and *D* (sets 3 through 5 and 9 through 11) showed weaker homology to the mRNAs for these keratins than clones containing portions of genes *B* and *C* (sets 6 through 8).

The substantial cross-hybridization of different epidermal keratin mRNAs with each of the genomic clones suggested that the four genes might encode a subfamily of closely related type I keratins expressed in epidermal cells. To unequivocally demonstrate which keratins are encoded by each of the four genes, it was necessary to obtain subcloned probes to unique coding portions of these genes. Previous sequence data on gene *A* enabled the construction of a subcloned probe containing the 3′-noncoding segment of its mRNA. This probe does not hybridize with any of the three linked genes, and it is represented only once in the human genome (36). When this probe was hybridized with epidermal mRNA, mRNA encoding keratin K14 was the only mRNA selected, even under conditions of reduced stringency (Fig. 3, set 5).

To determine the keratin encoded by gene *C*, we sequenced this gene (see below) and prepared a subcloned probe containing exon VIII and the 3′-noncoding portion of gene *C*. This subclone selected only the mRNA encoding a 46-kDa keratin (Fig. 4, lanes 3 and 4). Thus, gene *C* contained in clone GK-3 seemed to encode K17, which was previously shown to be expressed in cultured human epidermal cells (65).

We do not yet know which keratins, if any, are encoded by genes *B* and *D*. Although extensive cross-hybridizations currently preclude definite assignments, the degree of stringency with which mRNAs encoding 46- to 52-kDa keratins hybridize with these genes eliminates the possibility that the genes encode the more distantly related keratins of simple epithelial tissues (69). Thus, a subfamily of type I keratin genes has been isolated, which appear to encode the 46- to 52-kDa keratins expressed in the epidermis and in some other epithelial tissues (30).

Complete sequence of gene *C* and the predicted amino acid sequence of the encoded type I keratin K17: similarities between genes *A* and *C*. The sequence for the K14 gene (gene

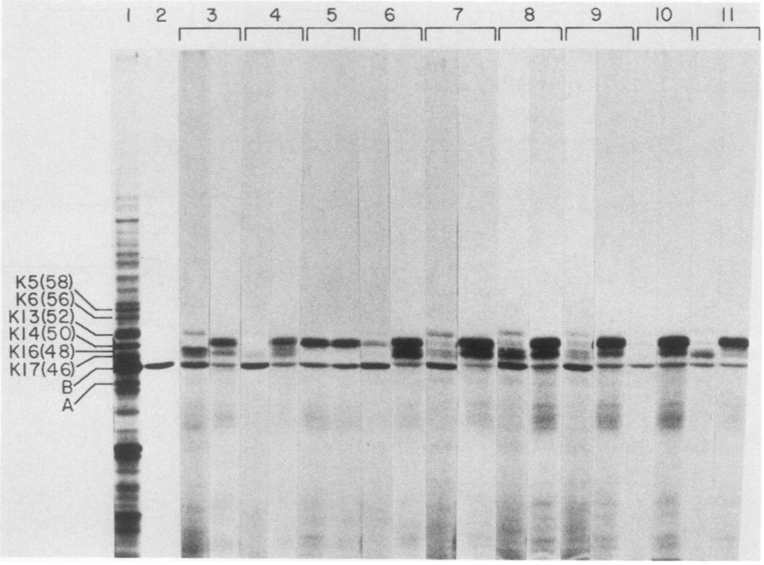

FIG. 3. Positive hybridization-translation analysis of the four type I epidermal keratin genes *A* through *D*. The subcloned gene fragments illustrated by the overhead bars in Fig. 2 were denatured and bound to nitrocellulose filters. The filters were hybridized with polyadenylated RNA from cultured human epidermal cells (17, 18), and the mRNAs were eluted sequentially at 65°C and the 85°C (30). After translation of the eluted mRNAs in a reticulocyte lysate system (51), the [^{35}S]methionine-labeled translation products were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis and fluorography. Lanes: 1, translation of total epidermal mRNA; 2, translation in the absence of added mRNA. Subsequent lanes are in pairs; the first lane shows the translation products of the mRNAs eluted at 65°C from a specified gene fragment, and the second shows the translation products of the mRNAs eluted at 85°C. mRNAs were translated after hybridization with the following: 3, 5′ portion of gene *A*; 4, 3′ portion of gene *A*; 5, 3′-noncoding segment of gene *A*; 6, 3′-coding portion of gene *B*; 7, entire gene *C* isolated from GK-3; 8, entire gene *C* isolated from GK-6; 9, 5′ portion of gene *D* isolated from GK-6; 10, 5′ portion of gene *D* from GK-7; 11, 3′ portion of gene *D*. The keratins expressed in cultured human epidermal cells are marked at the left according to the nomenclature of Moll et al. (46) and according to estimated molecular masses (shown in parentheses in kilodaltons). Band B is an mRNA-independent artifact, and band A is actin.

A) has been reported previously (36, 37). The sequence for the K17 gene (gene *C*) is shown in Fig. 5. The coding portion of this gene was identified by aligning the gene sequence with K14 cDNA (24). Gene *C* is about 3.0 kb in length and is predicted to encode a polypeptide chain of 469 amino acid residues (50.471 kDa). The putative molecular mass is substantially larger than that estimated by polyacrylamide gel electrophoresis (46 kDa [65]). This feature has been noted before for other sequenced human epidermal keratins (36, 37, 67). The predicted amino acid composition matches well with that predicted by chemical means (Table 1).

The nucleotide sequence within the coding portion of gene *C* shares a high degree of homology (85%) with the K14-coding sequence. The predicted amino acid sequence of K17 is also highly similar to that of K14 (83% sequence identity; Fig. 5). The highest degree of homology is found in the α-helical domains (encompassed by the gray boxes) and in the nonhelical spacer regions of the proteins. The most striking difference between the two keratin sequences resides in exon VIII, where only 28% homology exists at the nucleotide level, and almost no similarity is seen in amino acid sequence (Table 2). Despite considerable divergence throughout the exon VIIIs of the two genes, the region in the vicinity of the translation termination codon shows significant intersequence homology. In addition, although the 3′-untranslated segments of the two genes are divergent in sequence, their length has been quite highly conserved.

The positions of the introns (triangles in Fig. 5) were determined by comparing the sequence of the gene with that of the previously published coding sequence of the K14 gene (36, 37). All seven intron positions are identical to those of the K14 gene. They begin with the sequence gt and end with the sequence ag. Immediately 5′ from each of the ag sequences is a string of 7 to 12 pyrimidine residues. These features match well with the intron consensus characteristics that have been described previously (48).

Although intron positions seem to have been highly conserved within the subset of type I keratin genes, intron sequence and size have diverged considerably. The intron sequences for both the K14 and K17 genes have been determined in their entirety, and Table 2 shows a summary of their size relation. The intron sizes of gene *C* are extraordinarily small, with only one intron larger than 300 bp in length. In contrast to the high level of homology within most of the exons of genes *A* and *C*, none of the seven introns showed intersequence homologies of 12 nucleotides or greater.

Evolutionary basis for the high level of divergence in the nonhelical termini of the keratins. One of the most unusual features of the keratin family is the evolutionary divergence of the nonhelical termini of both the type I and the type II polypeptides. Inspection of the nonhelical amino termini of the two type I keratins K14 and K17 indicates that much of the sequence in this region was probably generated by repeated duplications of an oligonucleotide encoding the amino acid sequence Gly-Gly-Gly-Phe. The original nucleotide sequence was most likely GGCGGCGGCTTC. Given a limited number of deletions, insertions, and mutational variations, this sequence is repeated extensively throughout the amino-terminal segments of both of these type I genes as well as the corresponding region of several epidermal type II keratins. For some keratins, this sequence has even been found in the nonhelical carboxy-terminal segment (25–28, 32, 60, 62, 67).

Another interesting characteristic of the nonhelical termini of the type I and type II epidermal keratins is the abundance of serine residues. In the K14 sequence for example, there is a stretch of six serines in tandem. Five of these residues are also found in the corresponding region of K17. All of these serines are coded by AGC, and the stretch is flanked by phenylalanine. Since this codon is genetically less favorable, having only a twofold degree of conservancy at the third base (AGC/T) as opposed to a fourfold conservancy for the serine codon TCX, it is unlikely that such tremendous sequence preference is merely coincidental. A more plausible explanation for their origin might be that duplication of the oligonucleotide described above was followed by conversion of the glycine residues (GGC/T) to serines (AGC/T) by G-to-A transitions either before or after the duplication.

In contrast to the glycine- and serine-rich epidermal keratins, the hair and wool keratins have nonhelical termini rich in cysteine (8, 9). It is interesting that these seemingly unrelated cysteine-rich sequences could have easily arisen from one-base changes of either Gly (GGC/T) or Phe (TTC/T), forming a cysteine codon in the dodecanucleotide. Thus, it is likely that the variable nonhelical termini of the keratins had evolutionary origins which may have preceded the duplication of the primordial keratin gene. Extensive modification and duplication of the seminal oligonucleotide may have later led to a diversification of the keratins which ultimately found biological significance.

Comparison of the 5′-regulatory regions of the coexpressed K14 and K17 genes. At a single position 103 bp 5′ upstream from the putative translation initiation codon ATG of the K17 gene is found the sequence TATAAA. If the transcription initiation site is located 20 to 25 nucleotides downstream from this sequence, as is typical for many eucaryotic genes (for a review, see reference 5), then a 5′-untranslated sequence of about 75 bp would be transcribed. When this sequence is aligned with the corresponding sequence of the K14 gene, a remarkable degree of homology (86%) is revealed (Fig. 6). Even when this sequence is compared with the very distantly related, but coexpressed gene encoding the type II keratin K6b, the homology (67%) is still clearly evident. In striking contrast, no similarity was found between these sequences and that of the K1 (67-kDa) keratin gene expressed in differentiating epidermal cells (26). Thus, there seems to be a strong evolutionary pressure against changing the 5′-untranslated sequences of those keratin genes that are coexpressed in cultured human epidermal cells.

When the sequences 5′ upstream from the TATA boxes of the genes encoding K14, K17, and K6b were aligned for optimal homology, additional similarities were discovered (Fig. 6). These homologies (53 to 90%) appear in clusters extending as far as 230 bp 5′ upstream from the TATA sequence. The region of strongest homology for all three genes corresponds to nucleotides 30 to 65 bp 5′ upstream from the TATA box of the K17 gene, which shows 90% homology with the K14 gene and 69% homology with the K6b gene. Interestingly this region encompasses two se-

TABLE 1. Comparison of the predicted and actual amino acid compositions of the type I K17 keratin

Amino acid	% of indicated amino acid determined by:	
	Amino acid analysis[a]	cDNA analysis
Ala	9.9	7.8
Arg	7.6	6.8
Asn		3.0
Asp		4.3
Asn + Asp	8.7	7.3
Cys	ND	1.3
Gln		6.5
Glu		9.5
Gln + Glu	16.2	16.0
Gly	8.9	11.7
His	1.2	0.6
Ile	4.1	3.6
Leu	12.2	10.2
Lys	5.1	3.4
Met	2.0	2.1
Phe	2.7	2.8
Pro	1.8	1.3
Ser	7.7	13.2
Thr	4.3	4.3
Tyr	3.0	3.0
Trp	ND	0.4
Val	4.7	4.3

[a] Keratin K17 was isolated from cultured human epidermal cells and purified by polyacrylamide gel electrophoresis and electroelution (16). One preparation was carried out with glycine in the gels and buffers, and another was carried out by substitution of molar equivalents of alanine for glycine. After dialysis, 20 μg of purified K17 was hydrolyzed in 6 M HCl at 108°C for 36 h in evacuated sealed tubes. The sample was applied to a Dionex D-501 amino acid analyzer. ND, Not determined.

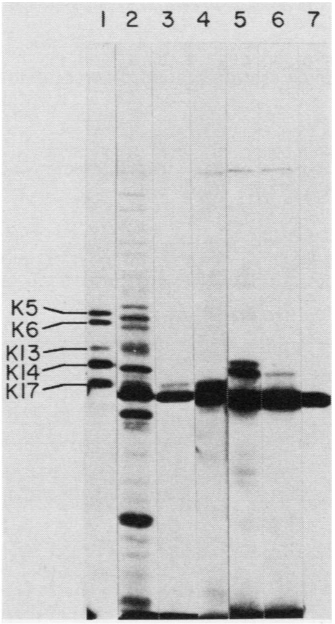

FIG. 4. Positive hybridization-translation analysis with specific subcloned portions of genes *A* and *C*.. Positive hybridizations were conducted as described in the legend to Fig. 3, but this time unique subcloned segments of genes *A* and *C* extending from exon VIII to the 3′-noncoding region of the genes were used to select keratin mRNAs. Lanes 1, [^{35}S]methionine-labeled keratins from cultured human epidermal cells; 2, radiolabeled products translated from human epidermal mRNA; 3 and 4, translation products of mRNA eluted at 65 and 85°C, respectively, from the probe to the exon VIII 3′-noncoding segment of gene *C*; 5 and 6, translation products of mRNA eluted from the probe to the exon VIII 3′-noncoding segment of gene *A*; 7, translation artifact of the reticulocyte lysate system. Note that mRNAs encoding two keratins of 50 kDa were selected at low stringency (lane 5). This doublet is less well resolved, but still apparent, in Fig. 3. These two keratins have nearly identical isoelectric mobilities on two-dimensional gels (unlike K15); thus there may be two highly similar K14 keratins in these cells.

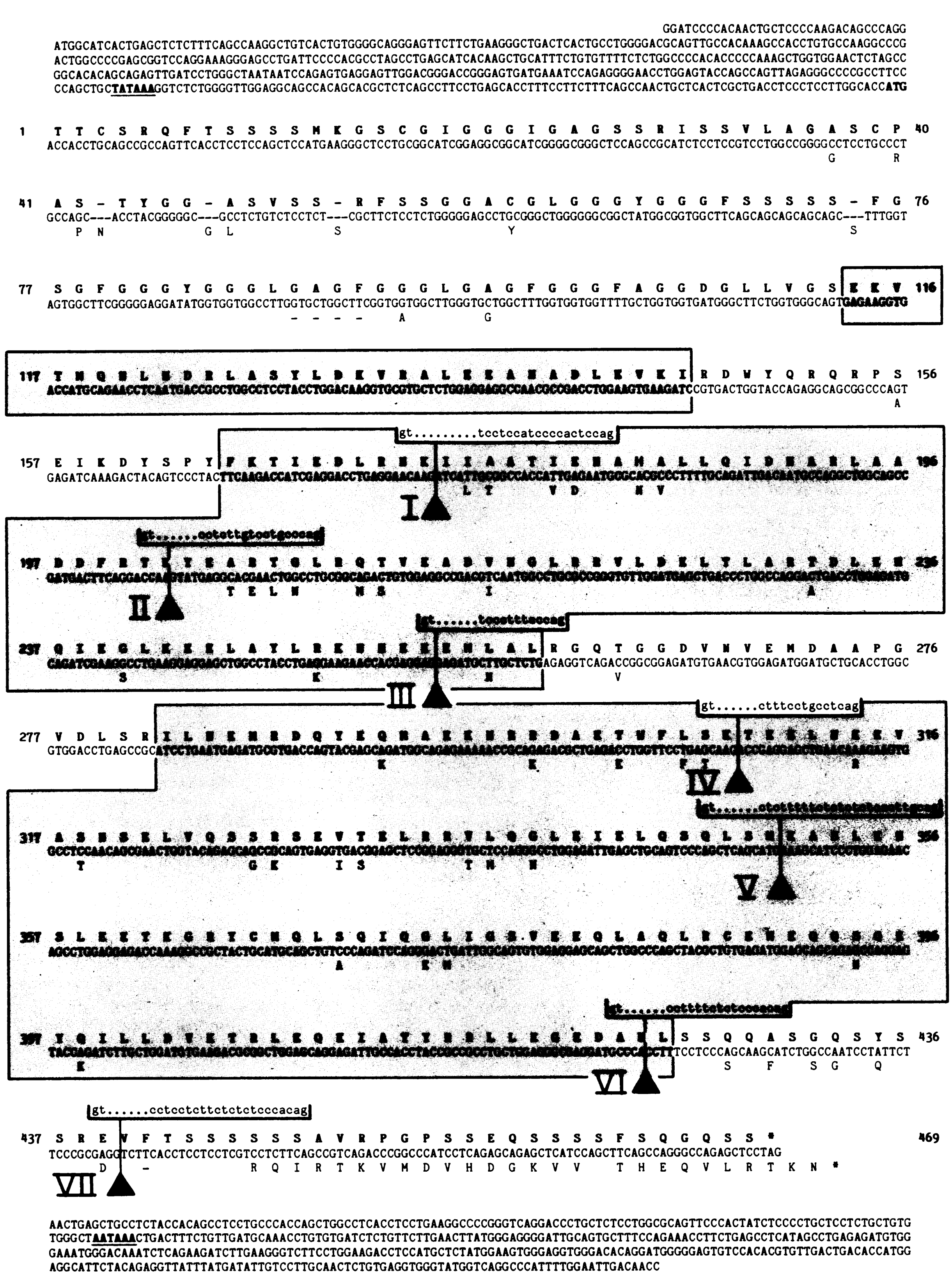
gt........tcctccatccccactccag
I
gt......ectcttgtcctgcccag
II
gt......tcectttecceag
III
gt......ctttcctgcctcag
IV
V
VI
gt......cctcctcttctctctcccacag
VII

FIG. 5. Complete nucleotide sequence of the human 46-kDa (K17) type I keratin gene contained in genomic clone GK-3. The sequence of gene *C* and its 5'- and 3'-flanking regions are shown with 120 nucleotides per line. The positions of the introns are indicated by triangles. Intron-exon junctions and pyrimidine consensus sequences are shown for each intron in lowercase letters. Complete intron sequences (not shown) were determined, and their exact sizes are listed in Table 1. The exons were identified by comparing the sequence of the gene *C* with K14 keratin cDNA sequences. The predicted amino acid sequence is shown in one-letter code above the nucleotide sequence. Amino acid residues encoded by the K14 gene that are different from those predicted by the K17 gene are shown below the nucleotide sequence. The gray boxes mark the α-helical domains predicted from the K17 amino acid sequence (6, 7, 20). Throughout these domains are the heptad repeats of hydrophobic residues, which identify the portions of the polypeptide that are involved in coiled-coil interactions with a second keratin. Note that an adenosine nucleotide was inadvertently omitted immediately preceding the TATA box.

TABLE 2. Comparison of exon and intron sizes of the type I keratin genes encoding K14 and K17

Keratin K17		Keratin K14[a]		% Nucleotide homology[b]	% Amino acid homology[b]	Keratin K17		Keratin K14[a]	
Exon	Size (bp)	Exon	Size (bp)			Intron	Size (bp)	Intron	Size (bp)
I	522	I	522	94	96	I	428	I	1,256
II	83	II	83	76	79	II	136	II	562
III	157	III	157	87	83	III	120	III	335
IV	162	IV	162	84	87	IV	88	IV	83
V	126	V	126	80	79	V	298	V	95
VI	221	VI	221	94	95	VI	94	VI	94
VII	47	VII	47	58	63	VII	159	VII	564
VIII	89	VIII	95	28	21				

[a] Values obtained from Marchuk et al. (37).

[b] Sequence homologies were calculated without penalty for introduction of gaps necessary for optimal alignment.

quences, first identified for the K14 gene, that share a substantial resemblance to the core transcription enhancer sequence of simian virus 40 (33, 47, 55; underlined sequences in Fig. 6). Whether these or other homologous sequences 5′ upstream from the TATA boxes of the three keratin genes play a role in their abundant or coordinate expression in human epidermal cells remains to be elucidated.

There is only one mRNA for the K17 gene. Because the K17 gene is tightly linked to gene *D*, it offers an ideal opportunity to examine the possibility that multiple polyadenylation signals or differential splice junctions might exist at the 3′ end of the gene. In 300 bp downstream from the first AATAAA, no additional polyadenylation signals were found (Fig. 5). To test for the possibility that other AATAAA signals might be located far downstream, and to look for any additional gene *C* exons, we prepared a series of subcloned probes containing sequences extending from the first TGA stop codon of gene *C* to a position within several hundred nucleotides of the start of gene *D* (probes indicated by the dashed lines at the bottom of Fig. 2A). These probes were tested for their ability to hybridize with polyadenylated epidermal RNA (Fig. 7). Whereas the probe containing the unique 3′-noncoding segment of gene *C* hybridized with a 1.6-kb band (Fig. 7, lane 2) previously shown to contain the mRNAs for the 46-kDa (K17) and 50-kDa (K14) keratins (lane 1) (15), no probes containing sequences 3′ downstream from the polyadenylation signal hybridized specifically with this or any other RNA bands. Thus, at least for the K17 gene, there seems to be only a single epidermal transcript giving rise to a single epidermal mRNA which encodes a single epidermal keratin polypeptide.

DISCUSSION

The mRNAs encoding the four type I keratins of cultured human epidermal cells (K13, K14, K16, and K17) were known to be highly homologous as judged by their ability to cross-hybridize under stringent conditions with a cloned K14 cDNA (15, 30). Their close relation made them an ideal subfamily to examine the molecular mechanisms underlying their coexpression.

In this paper, we have described the isolation and characterization of four human genes that share a high degree of homology with this group of type I epidermal keratin mRNAs. Three of the genes, referred to as genes *B* through *D*, are tightly linked chromosomally. The central gene (*C*) has been identified as the gene encoding K17 (46 kDa), and the unlinked gene (*A*) encodes the epidermal keratin K14. We are not yet certain whether the two other linked genes encode K13 and K16, respectively.

The finding that the mRNAs encoding K13 and K16 hybridize with gene *C* under conditions of high stringency suggested that the gene might encode more than one epidermal mRNA. The possibility of differential usage of exon-intron splice junctions within the internal intron sequences of a keratin hnRNA transcript is highly unlikely, since a strong evolutionary pressure exists against shifting these intron positions (26, 34, 36, 37, 67). Two additional lines of evidence now indicate that K13 and K16 cannot arise from differential processing of the 3′ or 5′ external exon-intron junctions of the heterogeneous nuclear RNA transcript of gene *C*: (i) a unique subcloned probe for the known exon VIII of gene *C* selects only the mRNA for K17 and (ii) none of the mRNAs from cultured human epidermal cells hybridizes with sequences located 3′ to the known polyadenylation signal of gene *C* and 5′ to the approximate start of gene *D*. Although we have not yet ruled out the possibility that other keratin mRNAs expressed developmentally or in other tissues might arise from tissue-specific processing of the gene *C* transcript, this is unlikely due to the more distant relation of the mRNAs for nonepidermal type I keratins (30). Thus the processing of transcripts encoding the structural proteins of epithelial cells seems to differ from that of the structural proteins of muscle cells, where developmentally regulated differential splicing yields otherwise identical proteins with different carboxy-terminal segments (2, 43, 53).

We cannot yet account for the discrepancies between the keratin molecular weights predicted by sequence analyses and those predicted by electrophoretic mobilities. In particular, whereas the electrophoretic mobilities of K14 and K17 predicted a difference in protein size of almost 4 kDa (64), the amino acid sequences predict a difference of only 1 kDa. It seems unlikely that posttranslational processing of K17 or modification of K14 could account for this difference, since the products of their mRNAs translated in vitro show the same size differences as the proteins extracted from epidermal cells. We do not expect that the first ATG 3′ downstream from the TATA box of gene *C* is bypassed, since it is homologous to the conensus translation start sequence, A-X-X-A-U-G-A/G (31). Until further studies have been conducted, we can only suggest that despite the high degree of similarity between the two keratins, their electrophoretic mobilities may be anomalous. That small variations in sequence can give rise to substantial variation in electrophoretic mobility has been well documented (11, 38, 67).

The type II keratins K5 (58 kDa) and K6 (56 kDa) and the type I keratins K14 (50 kDa) and K17 (46 kDa) are coexpressed in a number of different epithelial tissues and cultured keratinocytes (13, 29, 46, 52). Although the regulatory factors that govern their tissue-specific expression have not yet

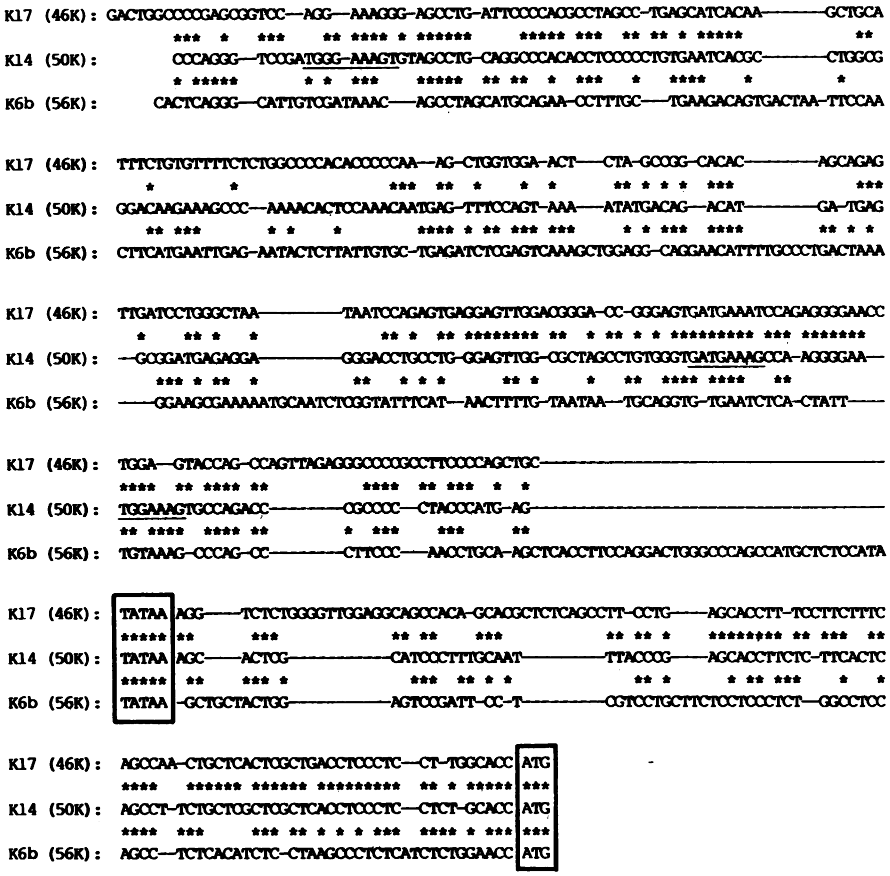
K17 (46K): GACTGGCCCCGAGCGGTCC—AGG—AAAGGG-AGCCTG-ATTCCCCACGCCTAGCC-TGAGCATCACAA————GCTGCA
K14 (50K): CCCAGGG—TCCGATGGG-AAAGTGTAGCCTG-CAGGCCCACACCTCCCCCTGTGAATCACGC————CTGGCG
K6b (56K): CACTCAGGG—CATTGTCGATAAAC——AGCCTAGCATGCAGAA-CCTTTGC——TGAAGACAGTGACTAA-TTCCAA
K17 (46K): TTTCTGTGTTTTCTCTGGCCCCACACCCCCAA—AG-CTGGTGGA-ACT——CTA-GCCGG-CACAC————AGCAGAG
K14 (50K): GGACAAGAAAGCCC—AAAACACTCCAAACAATGAG-TTTCCAGT-AAA——ATATGACAG—ACAT————GA-TGAG
K6b (56K): CTTCATGAATTGAG-AATACTCTTATTGTGC-TGAGATCTCGAGTCAAAGCTGGAGG-CAGGAACATTTTGCCCTGACTAAA
K17 (46K): TTGATCCTGGGCTAA————TAATCCAGAGTGAGGAGTTGGACGGGA-CC-GGGAGTGATGAAATCCAGAGGGGAACC
K14 (50K): —GCGGATGAGAGGA————GGGACCTGCCTG-GGAGTTGG-CGCTAGCCTGTGGGTGATGAAAGCCA-AGGGGAA—
K6b (56K): ——GGAAGCGAAAAATGCAATCTCGGTATTTCAT—AACTTTTG-TAATAA—TGCAGGTG-TGAATCTCA-CTATT——
K17 (46K): TGGA—GTACCAG-CCAGTTAGAGGGCCCCGCCTTCCCCAGCTGC————————
K14 (50K): TGGAAAGTGCCAGACC————CGCCCC—CTACCCATG-AG————————
K6b (56K): TGTAAAG-CCCAG-CC————CTTCCC——AACCTGCA-AGCTCACCTTCCAGGACTGGGCCCAGCCATGCTCTCCATA
K17 (46K): TATAA AGG——TCTCTGGGGTTGGAGGCAGCCACA-GCACGCTCTCAGCCTT-CCTG——AGCACCTT-TCCTTCTTTC
K14 (50K): TATAA AGC——ACTCG————CATCCCTTTGCAAT————TTACCCG——AGCACCTTCTC-TTCACTC
K6b (56K): TATAA -GCTGCTACTGG————AGTCCGATT-CC-T————CGTCCTGCTTCTCCTCCCTCT—GGCCTCC
K17 (46K): AGCCAA-CTGCTCACTCGCTGACCTCCCTC—CT-TGGCACC ATG
K14 (50K): AGCCT-TCTGCTCGCTCGCTCACCTCCCTC—CTCT-GCACC ATG
K6b (56K): AGCC—TCTCACATCTC-CTAAGCCCTCTCATCTCTGGAACC ATG

FIG. 6. Comparison of the nucleotide sequences 5′ upstream from the translation initiation codons of three genes that are coexpressed in cultured human epidermal cells. The 5′ sequences of two coexpressed type I keratin genes, K17 and K14, and one coexpressed type II keratin gene, K6b, were aligned for optimal homology. Putative translation initiation codons and TATA boxes are outlined by boxes. The stars indicate positions of sequence identity between the K17 and K14 sequences and the K14 and K6b sequences, respectively. The three sequences of the K14 gene previously shown to share homology with the simian virus 40 core enhancer sequence (37) are underlined.

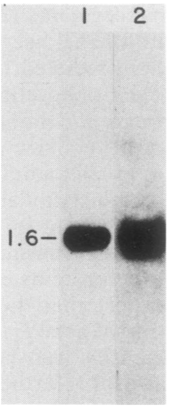

FIG. 7. A single transcript of 1.6 kb is encoded by the K17 gene. Polyadenylated RNA was isolated from cultured human epidermal cells and resolved by formaldehyde-RNA gel electrophoresis as described previously (15). After blot transfer (66), the RNAs were hybridized with ^{32}P-labeled cDNA transcribed from the following sources: 1, K14 cDNA; 2, an M13mp19 clone beginning in intron 7 of the K17 gene and extending through the polyadenylation signal (Fig. 5). Probes prepared from DNA fragments extending 3.5 kb 3′ downstream from this polyadenylation signal (Fig. 2A) gave no hybridization.

been determined, a comparison of the 5′-upstream sequences of three of the genes encoding these keratins has begun to provide some clues which might eventually assist us in this search. The most striking relation among the K6, K14, and K17 genes resides in their highly conserved 5′-untranslated sequence. In contrast, the type II gene encoding K1 (67 kDa) (26) does not share homology in this region, nor is it coexpressed in rapidly growing cultured keratinocytes. These findings are particularly intriguing in light of recent evidence that the preferential translation of heat shock-specific mRNAs may require sequences in the 5′-untranslated leader (39a). Whether the high homology in this region of the three keratin genes similarly reflects an important regulatory mechanism at the posttranscriptional level has not yet been examined.

In addition to the high degree of homology in the 5′-untranslated sequences of the K6, K14, and K17 genes, there appears to have been some evolutionary pressure exerted against changing segments of sequence in the 230 bp located 5′ upstream from their TATA boxes. This region has been shown to play an extremely important role in transcriptional regulation of many eucaryotic genes (10, 12, 39, 42, 50). The level of homology is as high as 90% for a 60-bp segment of the two type I genes and 69% for a 30-bp stretch of the type II and the type I genes. Once again, this degree of homology is not attained within the corresponding sequences of the K1 gene, whose expression is limited to terminally differentiating keratinocytes.

As more genomic sequences for coexpressed sets of keratins become available, it should be possible to more precisely define candidate sequences in the 5′-upstream regions that might play a role in regulating the tissue-specific expression of these genes. This will be an important prerequisite for identifying specific transcription factors involved in activating the synthesis of different keratins in different epithelial tissues and at different stages of differentiation and development.

ACKNOWLEDGMENTS

We thank Ed Fritsch (Genetics Institute) for providing us with a partial *Hae*III-*Alu*I human genomic library in phage λ Charon 4A. We thank Sean McCrohon for his expert technical assistance.

This work was supported by grants funded by the National Institutes of Health (Public Health Service grant R01-AM 27883), the National Science Foundation (PCM8351534), the Whirlpool Foundation, the Cancer Research Foundation, and Exxon. E.F. is a Presidential Young Investigator and the recipient of a Public Health Service Career Development Award from the National Institutes of Health. M.L. is a National Institutes of Health Predoctoral Trainee.

LITERATURE CITED

1. **Anderson, S.** 1981. Shotgun DNA sequencing using cloned DNAse I-generated fragments. Nucleic Acids Res. **9:**3015–3026.
2. **Basi, G., M. Boardman, and R. Storti.** 1984. Alternative splicing of a *Drosophila* tropomyosin gene generates muscle tropomyosin isoforms with different carboxy-terminal ends. Mol. Cell. Biol. **4:**2828–2836.
3. **Benton, W., and R. Davis.** 1977. Screening lambda-gt recombinant clones by hybridization to single plaques in situ. Science **196:**180–182.
4. **Bladon, P. T., P. E. Bowden, W. J. Cunliffe, and E. J. Wood.** 1982. Prekeratin biosynthesis in human scalp epidermis. Biochem. J. **208:**179–187.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:**349–383.
6. **Chou, P., and G. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47:**45–148.
7. **Chou, P., and G. Fasman.** 1979. Prediction of beta-turns. Biophys. J. **26:**367–383.
8. **Crewther, W. G., L. M. Dowling, and A. Inglis.** 1980. Amino acid sequence data from a microfibrillar protein of alpha-keratin. Proc. 6th Quinquennial Int. Wool Text. Conf. **2:**79–81.
9. **Crewther, W. G., L. M. Dowling, D. A. D. Parry, and P. M. Steinert.** 1983. The structure of intermediate filaments. Int. J. Biol. Macromol. **5:**267–282.
10. **Davidson, B. L., J.-M. Egly, E. R. Mulvihill, and P. Chambon.** 1983. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promotor sequences. Nature (London) **301:**680–686.
11. **DeJong, W. W., A. Zweers, and L. H. Cohen.** 1978. Influence of single amino acid substitutions on electrophoretic mobility of sodium dodecyl sulfate-protein complexes. Biochem. Biophys. Res. Commun. **82:**532–539.
12. **Dynan, W. S., and R. Tjian.** 1983. The promotor-specific transcription factor Spl binds to upstream sequences in the SV40 early promotor. Cell **35:**79–87.
13. **Eichner, R., P. Bonitz, and T.-T. Sun.** 1984. Classification of epidermal keratins according to their immunoreactivity, isoelectric point, and mode of expression. J. Cell Biol. **98:**1388–1396.
14. **Franke, W., D. Schiller, M. Hatzfeld, and S. Winter.** 1983. Protein complexes of intermediate-sized filaments: melting of cytokeratin complexes in urea reveals different polypeptide separation characteristics. Proc. Natl. Acad. Sci. USA **80:**7113–7117.
15. **Fuchs, E., S. Coppock, H. Green, and D. Cleveland.** 1981. Two distinct classes of keratin genes and their evolutionary significance. Cell **27:**75–84.
16. **Fuchs, E., and H. Green.** 1978. The expression of keratin genes in epidermis and cultured epidermal cells. Cell **15:**887–897.
17. **Fuchs, E., and H. Green.** 1979. Multiple keratins of cultured human epidermal cells are translated from different mRNA molecules. Cell **17:**573–582.
18. **Fuchs, E., and H. Green.** 1980. Changes in keratin gene expression during terminal differentiation of the keratinocyte. Cell **19:**1033–1042.
19. **Fuchs, E., and I. Hanukoglu.** 1983. Unravelling the structure of the intermediate filament. Cell **34:**332–334.
20. **Garnier, J., D. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120:**97–120.
21. **Geisler, N., E. Kaufmann, and K. Weber.** 1982. Protein chemical characterization of three structurally distinct domains along the protofilament unit of desmin 10 nm filaments. Cell **30:**277–286.
22. **Geisler, N., E. Kaufmann, and K. Weber.** 1985. Antiparallel orientation of the two double-stranded coiled-coils in the tetrameric protofilament unit of intermediate filaments. J. Mol. Biol. **182:**173–177.
23. **Geisler, N., and K. Weber.** 1982. The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins. EMBO J. **1:**1649–1656.

23a. **Glass, C., K. H. Kim, and E. Fuchs.** 1985. Sequence and expression of a human type II mesothelial keratin. J. Cell Biol. **101:**2366–2373.

24. **Hanukoglu, I., and E. Fuchs.** 1982. The cDNA sequence of a human epidermal keratin: divergence of sequence but conservation of structure among intermediate filament proteins. Cell **31:**243–252.
25. **Hanukoglu, I., and E. Fuchs.** 1983. The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structural domains among keratins. Cell **33:**915–924.
26. **Johnson, L., W. Idler, X.-M. Zhou, D. Roop, and P. Steinert.** 1985. Structure of a gene for the human epidermal 67-kda keratin. Proc. Natl. Acad. Sci. USA **82:**1896–1900.
27. **Jorcano, J. L., J. K. Franz, and W. W. Franke.** 1984a. Amino acid sequence diversity between bovine epidermal cytokeratin polypeptides of the basic (type II) subfamily as determined from cDNA clones. Differentiation **28:**155–163.
28. **Jorcano, J. L., M. Rieger, J. K. Franz, D. L. Schiller, R. Moll, and W. W. Franke.** 1984b. Identification of two types of keratin polypeptides within the acidic cytokeratin subfamily I. J. Mol. Biol. **179:**257–281.
29. **Kim, K. H., D. Marchuk, and E. Fuchs.** 1984. Expression of unusually large keratins during terminal differentiation: balance of type I and type II keratins is not disrupted. J. Cell Biol. **99:**1872–1877.
30. **Kim, K. H., J. Rheinwald, and E. Fuchs.** 1983. Tissue specificity of epithelial keratins: differential expression of mRNAs from two multigene families. Mol. Cell. Biol. **3:**495–502.
31. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12:**857–872.
32. **Krieg, T. M., M. P. Schafer, C. K. Cheng, D. Filpula, P. Flaherty, P. M. Steinert, and D. R. Roop.** 1985. Organization of a type I keratin gene: evidence for evolution of intermediate filaments from a common ancestral gene. J. Biol. Chem. **260:**5867–5870.
33. **Laimins, L., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA **79:**6453–6457.
34. **Lehnert, M. E., J. L. Jorcano, H. Zentgraf, M. Blessing, J. K. Franz, and W. W. Franke.** 1984. Characterization of bovine keratin genes: similarities of exon patterns in genes coding for different keratins. EMBO J. **3:**3279–3287.
35. **Magin, T. M., J. L. Jorcano, and W. W. Franke.** 1983. Trans-

lational products of mRNAs coding for non-epidermal cytokeratins. EMBO J. 2:1387–1392.

36. **Marchuk, D., S. McCrohon, and E. Fuchs.** 1984. Remarkable conservation of structure among intermediate filament genes. Cell **39**:491–498.

37. **Marchuk, D., S. McCrohon, and E. Fuchs.** 1985. Complete sequence of a gene encoding a human type I keratin: sequences homologous to enhancer elements in the regulatory region of the gene. Proc. Natl. Acad. Sci. USA **82**:1609–1613.

38. **Markham, B. E., J. W. Little, and D. W. Mount.** 1981. Nucleotide sequence of a lex A gene of *E. coli* K-12. Nucleic Acids Res. **9**:4149–4161.

39. **Matsui, T., J. Segall, P. A. Weil, and R. G. Roeder.** 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. **255**: 11992–11996.

39a. **McGarry, T. J., and S. Lindquist.** 1985. The preferential translation of Drosophila hsp70 mRNA requires sequences in the untranslated leader. Cell **42**:903–911.

40. **McLachlan, A.** 1978. Coiled-coil formation and sequence regularities in the helical regions of alpha-keratin. J. Mol. Biol. **98**:293–304.

41. **McLachlan, A., and M. Stewart.** 1975. Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. J. Mol. Biol. **98**:293–304.

42. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of an eukaryotic protein-coding gene. Science **217**:316–324.

43. **Medford, R., H. Nguyen, A. Destree, E. Summers, and B. Nadal-Ginard.** 1984. A novel mechanism of alternative RNA splicing for the developmentally regulated generation of troponin T isoforms from a single gene. Cell **38**:409–421.

44. **Melton, D., P. Krieg, M. Rebagliati, T. Maniatis, K. Zinn, and M. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.

45. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.

46. **Moll, R., W. Franke, D. Schiller, B. Geiger, and R. Krepler.** 1982. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. Cell **31**:11–24.

47. **Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. Gaub, and P. Chambon.** 1981. The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res. **9**:6047–6068.

48. **Mount, S.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.

49. **Nelson, W., and T.-T. Sun.** 1983. The 50- and 58-kdalton keratin classes as molecular markers for stratified squamous epithelial: cell culture studies. J. Cell Biol. **97**:244–251.

50. **Parker, C. S., and J. Topol.** 1984. A Drosophila RNA polymerase II transcription factor specific for the heat-shock gene binds to the regulatory site of an hsp70 gene. Cell **37**:273–283.

51. **Pelham, H., and R. Jackson.** 1976. An efficient mRNA-dependent translation system from reticulocyte lysates. Eur. J. Biochem. **67**:247–256.

52. **Roop, D., P. Hawley-Nelson, C. Cheng, and S. Yuspa.** 1983. Keratin gene expression in mouse epidermis and cultured epidermal cells. Proc. Natl. Acad. Sci. USA **80**:716–720.

53. **Rozek, C., and N. Davidson.** 1983. *Drosophila* has one myosin heavy-chain gene with three developmentally regulated transcripts. Cell **32**:23–34.

54. **Sanger, F., S. Nicklen, and A. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

55. **Scholer, H., and P. Gruss.** 1984. Specific interaction between enhancer-containing molecules and cellular components. Cell **36**:403–411.

56. **Skerrow, D., and C. J. Skerrow.** 1983. Tonofilament differentiation in human epidermis: isolation and polypeptide composition of keratinocyte subpopulations. Exp. Cell Res. **143**:27–35.

57. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

58. **Steinert, P., and W. Idler.** 1975. The polypeptide composition of bovine epidermal alpha-keratin. Biochem. J. **151**:603–614.

59. **Steinert, P., W. Idler, and S. Zimmerman.** 1976. Self-assembly of bovine epidermal keratin filaments in vitro. J. Mol. Biol. **108**:547–567.

60. **Steinert, P., D. Parry, E. Racoosin, W. Idler, A. Steven, B. Trus, and D. Roop.** 1984. The complete cDNA and deduced amino acid sequence of a type II mouse epidermal keratin of 60,000 Da: analysis of sequence differences between type I and type II keratins. Proc. Natl. Acad. Sci. USA **81**:5709–5713.

61. **Steinert, P. M.** 1978. Structure of the three chain unit of the bovine epidermal keratin filament. J. Mol. Biol. **123**:49–70.

62. **Steinert, P., R. Rice, D. Roop, B. Trus, and A. Steven.** 1983. Complete amino acid sequence of a mouse epidermal keratin subunit and implications for the structure of intermediate filaments. Nature (London) **302**:794–800.

63. **Steinert, P. M., A. C. Steven, and D. R. Roop.** 1985. The molecular biology of intermediate filaments. Cell **42**:411–419.

64. **Steven, A., J. Hainfeld, B. Trus, J. Wall, and P. Steinert.** 1983. Epidermal keratin filaments assembled in vitro have masses-per-unit length that scale according to average subunit masses: structural basis for homologous packing of subunits in intermediate filaments. J. Cell Biol. **97**:1939–1944.

65. **Sun, T.-T., and H. Green.** 1978. Keratin filaments of cultured human epidermal cells: formation of intermolecular disulfide bonds during terminal differentiation. J. Biol. Chem. **253**:2053–2060.

66. **Thomas, P.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.

67. **Tyner, A., M. Eichman, and E. Fuchs.** 1985. The sequence of a type II keratin gene expressed in human skin: conservation of structure among all intermediate filament genes. Proc. Natl. Acad. Sci. USA **82**:4683–4687.

68. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.

69. **Wu, Y.-J., L. Parker, N. Binder, M. Beckett, J. Sinard, C. Griffiths, and J. Rheinwald.** 1982. The mesothelial keratins: a new family of cytoskeletal proteins identified in cultured mesothelial cells and nonkeratinizing epithelia. Cell **31**:693–703.