

# Isolation, Sequence, and Differential Expression of a Human K7 Gene in Simple Epithelial Cells

Charles Glass and Elaine Fuchs

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

**Abstract.** Simple epithelial cells synthesize a different set of keratins than epidermal cells. In experiments reported in this manuscript, we show that the base level of keratin expression in simple epithelial cells is variable for different cell types, and that, in some simple epithelia, this level can be upregulated by increasing the exposure of cells to retinoids, but not glucocorticoids or estradiol. To elucidate the molecular mechanisms underlying simple epithelial keratin gene

regulation, we have isolated and characterized a human gene encoding the simple epithelial keratin K7. By examining the possible regulatory elements of this gene and by investigating the behavior of this gene introduced transiently into simple epithelial cells, we have uncovered a possible basis for the differential expression of epidermal and simple epithelial keratin genes.

ONLY epithelial cells express keratins, a large family of proteins ( $M_r$  40–67 kD) which assemble into a cytoskeletal network of 8-nm intermediate filaments (IFs).<sup>1</sup> There are two distinct classes of keratins: type I keratins are small ( $M_r$  40–56.5 kD) and relatively acidic (pKi, 4.5–5.5), whereas type II keratins are larger ( $M_r$  53–67 kD) and more basic (pKi, 5.5–7.5; for reviews, see Moll et al., 1982; Steinert et al., 1985; Fuchs et al., 1987). At least one member of each type of keratin is essential and sufficient for filament formation (Steinert et al., 1976; Lee and Baden, 1976; Hatzfeld and Franke, 1985; Eichner et al., 1986).

Simple epithelial cells differ from stratified squamous epithelia in their pattern of keratin expression. They typically express keratin K8 (53 kD) of the type II class and K18 (44 kD) of the type I class. Some, including certain strains of a marsupial kidney line PtK2, express additional keratins K7 (55 kD) of the type II class and K19 (40 kD) of the type I class (Wu et al., 1982; Moll et al., 1982). None of these keratins are major constituents of the IF proteins expressed in stratified squamous epithelial tissues.

Some simple epithelial cells, including PtK2, are interesting in that they express two classes of IFs: keratin filaments, which are heteropolymers composed of K7, K8, K18, and K19; and vimentin filaments, which are homopolymers composed of a single vimentin polypeptide. Recently, studies on the human mesothelial cell strain LP-9 revealed that cells with these two cytoskeletal networks may be able to select which IFs to make depending upon their growth conditions; in the presence of epidermal growth factor (Connell and Rheinwald, 1983) and in the absence of retinoids (Kim et al.,

1987), LP-9 cells synthesize predominantly vimentin; in contrast, keratin synthesis seems to be dependent upon the presence of retinoids in the culture medium (Kim et al., 1987).

To begin to examine the molecular mechanisms which underlie the regulation of the tissue specific and retinoid-mediated expression of simple epithelial keratins, we have isolated the gene encoding the human keratin K7. In this paper, we report the characterization and sequence of this gene, and preliminary studies aimed at exploring the nature of its regulation and expression.

## Materials and Methods

### Screening the Genomic Library

A human genomic library (Hae III/Alu I partial digest of human genomic DNA, cloned into the Eco RI site of lambda phage Charon 4A through the use of linker ligation) was obtained from Dr. Ed Fritsch (Genetics Institute, Boston, MA). Phage were screened essentially as described by Benton and Davis (1977). For probes, two subclones were made from fragments of previously sequenced K7 cDNAs (Glass et al., 1985). A subclone corresponding to the 5' end of the K7 cDNA was made by inserting a 144-bp Eco RI/Sac I fragment from KC-2 into the Eco RI/Sac I site of plasmid pGEM3 (Promega Biotech, Madison, WI). This subclone will be referred to as pK7-5', and encompasses 69 bp of 5' noncoding sequence and 75 bp of 5' coding sequence. To make cRNA probes, pK7-5' was linearized with Eco RI and transcribed using SP6 RNA polymerase and a 20-nucleotide SP6 promoter primer. A subclone corresponding to the 3' end of the K7 cDNA was made by inserting a 209-bp Eco RI/Pst I fragment from KC-1 into the Eco RI/Pst I site of plasmid pGEM3. This subclone will be referred to as pK7-3', and encompasses 46 bp of 3' noncoding sequence and 163 bp of 3' coding sequence. To make cRNA probes, pK7-3' was linearized with Eco RI and transcribed using SP6-RNA polymerase and a 20-nucleotide SP6 promoter primer.

1. *Abbreviation used in this paper:* IF, intermediate filament.

## Genomic Blots

To prepare the probe, we isolated an 850-bp Eco RI fragment of GK7-3' containing the 3' noncoding portion and polyadenylation signal of the K7 gene. This fragment was subcloned into the Eco RI site of pGEM3 in the 3' to 5' direction. For probe preparation, this clone (referred to as pGK7.4) was linearized with Sac I, and the plasmid was incubated in the presence of T7 polymerase, radiolabeled ribonucleotides, and a T7 promoter primer.

To prepare the blot, 20 µg of total human DNA from blood leukocytes was digested with the indicated restriction endonucleases. Digested DNAs were separated by electrophoresis through 0.8% agarose and transferred to nitrocellulose paper by blotting (Southern, 1975).  $1 \times 10^6$  cpm/ml of radiolabeled cRNA probe to pGK7.4 was hybridized with the blot for 48 h at 42°C as described previously (Marchuk et al., 1984). Following hybridization, blots were washed three times in 0.1% SDS, 0.00015 M sodium citrate, 0.015 M NaCl at 60°C, treated briefly with 0.1 µg/ml RNase, and exposed to x-ray film for 2 d.

## DNA Sequencing

Selected restriction endonuclease-excised DNA fragments from GK7-5' and GK7-3' were subcloned into plasmid pGEM3. DNA sequence analysis of each of these fragments was conducted using the dideoxy method of Sanger et al. (1977) and the double-stranded sequencing procedure described by Chen and Seeberg (1985). T7 and SP6 promoter primers (20-mers) were obtained from Promega Biotech, and were used to determine the sequences of the 5' and 3' ends of each fragment. 17-mer and 18-mer oligonucleotide primers were synthesized using an Applied Biosystems, Inc. (Foster City, CA) model 380B DNA synthesizer and beta-cyanoethyl chemistry, and were used to determine the internal K7 coding sequences of each fragment.

## S1 Nuclease Protection Analysis

S1 nuclease protection studies were conducted according to the method of Quarless and Heinrich (1986). A fragment of the K7 gene, extending from the Hinc II site ~500 bp 5' from the ATG translation-initiation site to the Sac I site 140 bp downstream from this ATG, was subcloned into the Sac I/Hinc II site of plasmid pGEM3. This hybrid plasmid will be referred to as pGK7.S1. Complementary RNA was prepared by linearizing plasmid pGK7.S1 with Hind III, and using T7 RNA polymerase in the presence of <sup>32</sup>P-radiolabeled ribonucleotides and a 20-mer primer to the T7 promoter.  $2 \times 10^6$  cpm of the cRNA was hybridized with 40 µg of total RNA from cultured LP-9 cells as described by Marchuk et al. (1985). The hybrid was then subjected to 2,500 U of S1 nuclease at 37°C for 60 min to digest any single-stranded RNA sequences. The size of the protected fragment was determined by electrophoresis on a 6% polyacrylamide, 5 M urea sequencing gel in the presence of a standard set of sequencing reactions.

## Construction of Plasmids Used for Transfections

Fragments from three separate plasmids were used to construct the hybrid plasmids used in our transfection studies. The lambda clone KC-2, containing a 1,319-bp insert for the human K7 cDNA (Glass et al., 1985), was first digested with Sac I and Eco RI to release a 1,250-bp K7 cDNA fragment extending from a 5' Sac I site (75 bp 3' from the ATG translation initiation-start codon) to a 3' Eco RI site (160 bp 5' from the TGA translation-stop codon). This fragment was cloned into the polylinker region of plasmid pGEM3 (Promega Biotech) and the subclone, KC-2.SE, was then linearized with Eco RI, and treated with mung bean exonuclease (New England Biolabs, Beverly, MA) at 30°C for 30 min to make the Eco RI ends blunt. The linearized plasmid was then purified by excision and extraction from a low temperature agarose gel.

The plasmid pGEM2K14-P-K14NC was constructed by inserting the following fragments in a 5' to 3' sequence into the Hind III/Eco RI sites of plasmid pGEM2: (a) the Hind III/Kpn I fragment of the genomic clone GK-1 (Marchuk et al., 1984) containing ~1.8 kb of 5' upstream sequences extending into exon I of the K14 gene; (b) the Kpn I/Ava II fragment of the cDNA clone KB-2 (Hanukoglu and Fuchs, 1982), extending from the unique Kpn I site in the coding sequence to the Ava II site located 18 bp 5' from the TGA stop codon; (c) the Hinc II/Sma I fragment of pAHP.2 containing the sequences coding for the antigenic portion (15 amino acid residues) of neuropeptide substance P (Albers and Fuchs, 1987); and (d) the Stu I/Eco RI fragment of the genomic clone GK-1, containing the K14 3' noncoding portion extending from 20 bp 3' from the TGA stop codon to ~800 bp downstream from the polyadenylation site. This plasmid was treated with Hinc

II and Sac I to liberate a 650-bp fragment extending from a 5' Hinc II site 3 bp into the 5' end of the P element to a 3' Sac I site, ~430 bp downstream from the K14 polyadenylation site. This fragment was purified by agarose gel electrophoresis as described above, and then ligated to the linearized, blunt-ended Eco RI plasmid KC-2.SE. The resulting DNAs were then treated with Sac I to yield a 1,900-bp fragment consisting of the Sac I/blunt Eco RI fragment from KC-2.SE ligated at its 3' end to the 5' end of the Hinc II/Sac I fragment from pJ2K14NC. This fragment, K7-P-3', was purified by agarose gel electrophoresis.

The genomic clone GK7-5' was treated with Sac I and Eco RI to liberate a 3-kb fragment extending from the 5' Eco RI site illustrated in Fig. 1 to the Sac I site 75 bp downstream from the ATG translation-initiation codon. The fragment was purified and cloned into the polylinker region of plasmid pGEM3. This subclone will be referred to as pGK7.1A.

Plasmid pGK7.1A was treated with Sac I, followed by calf thymus alkaline phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, IN). The 1,900-bp Sac I fragment, K7-P-3' (see above), was then subcloned into the Sac I site of pGK7.1A. Clones containing P-element sequences were digested with Eco RV and Hind III to determine the correct orientation of the Sac I ligation. The junction between the Hinc II site and the blunt-ended Eco RI site was sequenced using an oligonucleotide primer to confirm that the sequence was as expected and in the appropriate reading frame. This plasmid will be referred to as pGEM3-K7-P.

The 4.9-kb insert in pGEM3-K7-P was excised with the enzymes Eco RI and Hind III and was subcloned into the Eco RI/Hind III sites of the polylinker region of plasmid pJay2 (Land et al., 1986; Giudice and Fuchs, 1987), which contains an SV-40 enhancer 3' to the Hind III site. This plasmid will be referred to as pJ2-K7-P.

## Preparation of Human Cell Cultures

Mesothelial cell strain LP-9 was provided to us by Dr. James Rheinwald (Dana Farber Cancer Center, Boston, MA) and cultured by the methods developed in his laboratory (Wu et al., 1982). Culture medium for these cells consisted of a 3:1 mixture of DME and Ham's F12 medium supplemented with 15% FCS and 0.4 µg/ml hydrocortisone (Rheinwald, 1980). Mouse NIH 3T3 fibroblasts were maintained in this medium supplemented with 10% calf serum. MCF-7 human breast adenocarcinoma cells and potoroo kidney epithelial cells (PtK2) were cultured in the same medium supplemented with 10% FCS,  $1 \times 10^{-8}$  M cholera toxin,  $1 \times 10^{-11}$  M insulin,  $2 \times 10^{-10}$  M triiodothyronine, and 5 µg/ml human transferrin.

When needed, retinoic acid was added from a 4 mM stock (400×) in DMSO to a final concentration of 1 µM. The stock was stored in the dark at -20°C.

For experiments with estradiol and dexamethasone, steroid hormones were removed from the serum by adsorption by a dextran-coated charcoal suspension as described by Heynes et al. (1967), and phenol red was omitted from the medium. When needed, 17-beta estradiol (Sigma Chemical Co., St. Louis, MO) was added to a final concentration of  $1 \times 10^{-8}$  M from a  $4 \times 10^{-6}$  M stock in 100% ethanol. Dexamethasone (Sigma Chemical Co.) was added to a final concentration of  $2.8 \times 10^{-7}$  M from a  $2.8 \times 10^{-4}$  M stock in 40% ethanol.

For immunofluorescence, glass cover slides were placed in the plastic culture dishes before cell plating. The cover slides were prepared for tissue culture by soaking for 3 h in sulfur dichromate followed by extensive washing in glass-distilled water, and finally heat sterilization at 240°C. Immediately before use, the slides were flamed in 95% ethanol and then rinsed twice in complete culture medium.

## DNA Transfections

Plasmid DNAs were purified by CsCl density gradient centrifugation. All DNAs for transfection were subsequently dialyzed, extracted with phenol, and precipitated once with 2.5 M ammonium acetate, 2.5 vol of ethanol, and once with 0.2 M NaCl, 2.5 vol of ethanol. Following precipitation and washing, DNAs were stored at 1 mg/ml in 10 mM Tris-HCl (pH 7.9), 1 mM EDTA. Using a modification of the calcium phosphate precipitation method of Graham and van der Eb (1973) and a 15% glycerol shock step (Parker and Stark, 1979), cells were transfected with 20–60 µg of plasmid DNA per  $10^6$  cells. Cells were assayed at 65 h after transfection.

## Antibodies and Immunofluorescent Microscopy

Cells were grown on glass cover slides as described above. After transfection, cells were washed with PBS and then fixed in methanol (-20°C) for

15 min. After fixing, cells were washed again in PBS. To detect expression of the K7-P protein, we used the mouse mAb RCK105 (Ramaekers et al., 1987) for fibroblasts, and the rat anti-substance P mAb, NCI/34 (Cuello et al., 1979) for cells which expressed an endogenous K7. As a necessary prerequisite for NCI/34 antibody recognition, protein carboxyl groups were chemically converted to amides with 0.5 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride in 1 M ammonium chloride, as described by Munro and Pelham (1984). The mouse mAb LE-41 (anti-K8; Lane, 1982) was used to identify endogenous keratin networks.

To visualize primary antibody binding, we used the following fluorescently labeled secondary antibodies: FITC-conjugated goat anti-rat IgG (Cappel Laboratories, Inc., Cochranville, PA; and Tago Inc., Burlingame, CA) and Texas red-conjugated goat anti-mouse and anti-rabbit IgG (Tago Inc.).

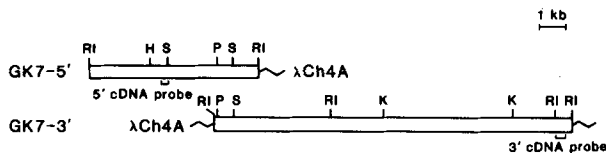
## Results

### Isolation of Genomic Clones Containing the K7 Gene

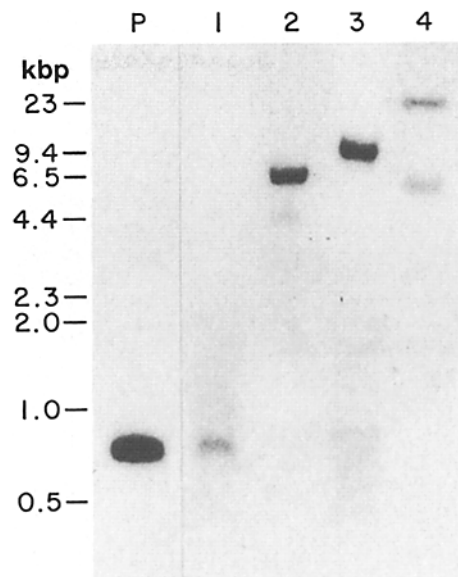
A human genomic library in lambda phage Charon 4A was screened with radiolabeled probes to the 5' and 3' sequences of a human K7 cDNA (see Materials and Methods). Two clones, GK7-5' and GK7-3', hybridized with the 5' and 3' cDNA probes, respectively. Further analyses using restriction endonuclease digestions and Southern blotting (Southern, 1975) indicated that these two clones contained ~1,650 nucleotides of common, seemingly overlapping sequences (Fig. 1). Partial sequence analysis of the Pst I/Sac I fragments from both clones indicated that the sequences in the overlapping region were identical. Additional analyses confirmed that the sequences contained in the genomic clones GK7-5' and GK7-3' corresponded precisely to the sequences in the isolated K7 cDNA clone (see below). The total distance between the 5' end of the genomic fragment hybridizing with the K7-5' cDNA probe and the 3' end of the genomic fragment hybridizing with the K7-3' cDNA probe was 14.5 kb, suggesting that the K7 gene is very large.

### There May Be More Than a Single Gene Encoding the Human K7

To verify that the K7 gene isolated from a human genomic library was also present in human genomic DNA, leukocyte DNA was digested with restriction endonuclease Eco RI,

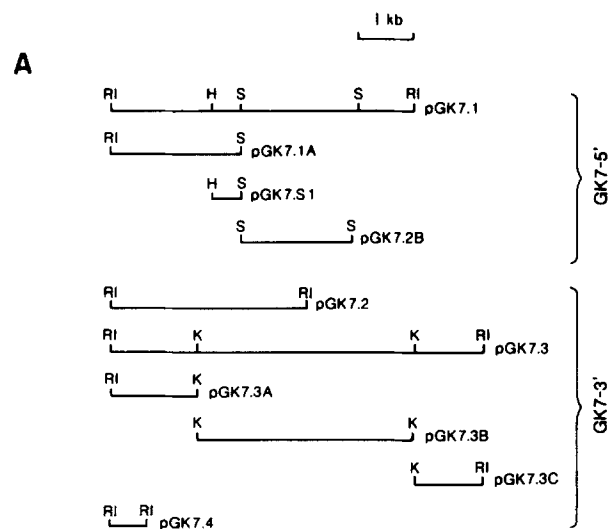


**Figure 1.** Genomic clones containing the K7 gene. Clones were isolated from a human genomic library as indicated in Materials and Methods. Clone GK7-5' hybridized strongly with K7-5', a subcloned segment of the 5' end of K7 cDNA. Clone GK7-3' hybridized strongly with K7-3', a subcloned portion of the 3' end of the K7 cDNA. These two genomic clones were partially mapped by restriction enzyme digestion and Southern blot hybridization (Southern, 1975). The apparent nucleotide overlap between the two clones is genuine and was verified as described in the text. The right phage arm and the 3' Eco RI fragment of human DNA insert is shown for the GK7-5' clone. The GK7-3' clone is shown in its entirety. The restriction endonuclease sites used for subcloning fragments for S1 mapping and in chromosomal DNA analysis, respectively, are marked above each stick diagram. RI, Eco RI; S, Sac I; P, Pst I; K, Kpn I; H, Hinc II.



**Figure 2.** Detection of the K7 genes in human chromosomal DNA. DNAs (20  $\mu$ g) from human leukocytes were digested with an excess of each restriction endonuclease: Eco RI (lane 1), Bam HI (lane 2), Kpn I (lane 3), and Hind III (lane 4). DNA fragments were resolved by electrophoresis through a 0.8% agarose gel and transferred to nitrocellulose paper (Southern, 1975). The blot was hybridized with  $^{32}$ P-labeled probe prepared from GK7.4, a plasmid containing an 850-bp genomic Eco RI fragment encompassing sequences homologous to the 3' noncoding portion of K7 mRNA. Lane P shows hybridization of radiolabeled probe with its 850-nucleotide complement from Eco RI-digested GK7-3' DNA. The intensity of the band in lane 1 corresponded to somewhat less than a single copy of DNA, presumably because very small fragments often do not adhere well to nitrocellulose. Lanes 2 and 3 are higher intensity and most likely represent fragments contributed by two different genes. The relative intensity of hybridization with the two putative genes can be seen in lane 4, where a Hind III digest shows two fragments even though there are no internal Hind III sites in the 850-nucleotide Eco RI probe. Sizes were estimated according to the migration of fragments from a Hind III digest of lambda DNA, which are shown at the left, in kilobase pairs.

and the fragments were subjected to Southern blot analysis (Southern, 1975). Hybridization with a  $^{32}$ P-labeled 3' noncoding probe (GK7.4) to the 850-bp Eco RI fragment of GK7-3' revealed the presence of a single hybridizing band of 850 bp (Fig. 2, lane 1). The mobility of this hybridizing fragment was identical to that observed with Eco RI-digested GK7-3' DNA (Fig. 2, lane P). This result suggested that the gene spanning GK7-5' and GK7-3' is indeed present in the human genome. Most of the hybridizing signal in Bam HI-digested DNA (Fig. 2, lane 2) and in Kpn I-digested DNA (Fig. 2, lane 3) was also found in a single band, suggesting at first glance that there may be only one K7 gene. However, for both of these samples, the intensity of hybridization was greater than that expected based on a single copy gene. Moreover, for Hind III-digested DNA, two distinct hybridizing bands were seen (Fig. 2, lane 4) and yet no Hind III sites exist within the 850-bp DNA fragment used as probe. Collectively, these data suggest that in fact, there may be more than a single gene encoding the human keratin K7.



Primer Sequence	Region	Strand Sequenced	clone(s)
ACCCCTTCCCAGTCTC	ATG-5'	C	pGK7.1A, pGK7.S1
TCCCCGCTGTTCTCTG	ATG-5'	C	pGK7.S1
TGCCCGCCCGCCCGCC	ATG-5'	N	pGK7.1A, pGK7.S1
GGCGGGGCGGGCGGGCA	ATG-5'	C	pGK7.1A, pGK7.S1
CCTTTCTCTGCTGCCAG	ATG-5'	N	pGK7.1A, pGK7.S1
AGGAGGAGCGCCTCTC	ATG-5'	N	pGK7.1A, pGK7.S1
TCGCCCGCCGCTAGGTC	ATG-5'	C	pGK7.1A, pGK7.S1
GATGGACATGGTGGCTG	ATG-5'	N	pGK7.1A, pGK7.S1
GCCACGCGCGCCCGCGA	exon 1	C	pGK7.1, pGK7.1B
Sp6 promoter primer (pGEM3)	exon 1	N	pGK7.1A, pGK7.S1
Sp6 promoter primer (pGEM3)	exon 1	C	pGK7.1B
CGCGAGGTCAACATTAA	exon 1	C	pGK7.1B
GAGCAGCAGAACAAGCTG	exon 2	C	pGK7.1, pGK7.2
TGAAGTCCTCCACCACAT	exon 2	N	pGK7.1, pGK7.2
GAAGATAAATTAACCGC	exon 3	C	pGK7.2
TTCAGGACCACAAACTC	exon 3	N	pGK7.2
TGGATGCTGCCTACATG	exon 4	C	pGK7.2
TCATTGAGGGTCTGAG	exon 4	N	pGK7.2
Sp6 promoter primer (pGEM3)	exon 5	N	GK7.3A
Sp6 promoter primer (pGEM3)	exon 5	C	GK7.3B
ACCCTCCAGGCCAGGC	exon 6	C	GK7.3
GTTCTTGATGTTGTCGA	exon 6	N	GK7.3
CCAAGTTGGAGCCGCC	exon 7	C	GK7.3
TCTCCTCGCCCTCCAGC	exon 7	N	GK7.3
GTTGGCTGGAGATGGAG	exon 8	C	GK7.3A
AGATATTCACGCTCCCA	exon 8	N	GK7.3A
TCTAACAGCACAAGGG	exon 8	N	GK7.3
T7 promoter primer (pGEM3)	exon 9	N	pGK7.3
T7 promoter primer (pGEM3)	exon 9	C	pGK7.4
GCGGGCACTCCTGCGAC	TGA-3'	N	pGK7.4
CCCACCCCTCCTCCCAT	TGA-3'	C	pGK7.4
ATGTGCCTTCTAGTCTC	TGA-3'	N	pGK7.4

**Figure 3.** Sequencing strategy for the human K7 gene. (A) Genomic sequences were excised from lambda clones GK7-5' and GK7-3' using the appropriate restriction endonucleases (see Fig. 1). Fragments were subcloned into the bacterial plasmid vector pGEM3. The stick diagrams illustrate the subclones which were used for

### The Sequence of the Gene Spanning GK7-5' and GK7-3' Corresponds Perfectly to the Sequence of the Human Keratin K7 cDNA

To determine whether the isolated K7 gene was indeed the gene corresponding to the K7 cDNA cloned from human mesothelial mRNA, the entire coding portion, exon-intron junctions, and 5' and 3' flanking regions of the gene were sequenced. The strategy used for sequencing is shown in Fig. 3. Fig. 4 illustrates the structure of the K7 gene, aligned with the human K6b epidermal type II gene, whose sequence was determined previously (Tyner et al., 1985). The sequences of the coding and noncoding segments matched perfectly with the corresponding sequences in the cDNA clones (Glass et al., 1985). Only a single nucleotide difference was observed with the previously reported sequence (*asterisk*, Fig. 4), and upon rereading the cDNA sequencing gel, it was realized that this difference was due to a prior reading error, rather than a true sequence difference.

In our earlier study, the structural similarities among members of the type II keratin protein family were demonstrated by aligning the K7 cDNA with the coding sequences of the K6b gene (Glass et al., 1985). In this paper, the similarities in gene structure could be shown by aligning the coding portions of the K7 gene with the corresponding regions of the K6b gene (Fig. 4). All eight intron/exon junctions of the K7 and K6b genes were identically positioned, even though their encoded mRNA sequences shared only 69% nucleotide similarity. Hence, at least one simple epithelial keratin gene has been found to share a common structure which is highly homologous to that of the epidermal keratin genes. As noted previously for human epidermal type II keratin genes (Tyner et al., 1985; Johnson et al., 1985), the intron/exon junctions did not seem to demarcate the structural boundaries of the alpha-helical domains of the keratin protein, predicted by the computer methods of Chou and Fasman (1978, 1979).

### The Transcription-Initiation Site and Possible Regulatory Sequences for the Human K7 Gene

The 5' upstream sequence of the K7 gene is shown in Fig. 5. To determine the precise transcription-initiation site, mRNAs from the human mesothelial cell strain LP-9 were hybridized with a 600-nucleotide radiolabeled cRNA extending upstream from a Sac I site located 75 nucleotides 3' of the putative translation-initiation codon (sequences ho-

subsequent double-stranded dideoxy sequencing. RI, Eco RI; H, Hind II; S, Sac I; K, Kpn I. (B) Exon regions and intron/exon junctions were sequenced as described in Materials and Methods. Sequencing was conducted using 17-20-bp oligonucleotide primers obtained either commercially (for K7 sequences adjacent to the plasmid multiple-cloning region) or synthetically (for sequences complementary to the known K7 cDNA). 5' and 3' flanking sequences were obtained using synthetic primers complementary to the 5' and 3' ends of newly deduced sequences. For all sequences shown, both strands were analyzed. A complete list of all the primers used in sequencing is given in B, along with the clones to which each primer hybridized. ATG-5' corresponds to sequences that are located 5' from the ATG translation-initiation start site. TGA-3' corresponds to sequences that are located 3' from the TGA translation termination codon. N, noncoding strand; C, coding strand.

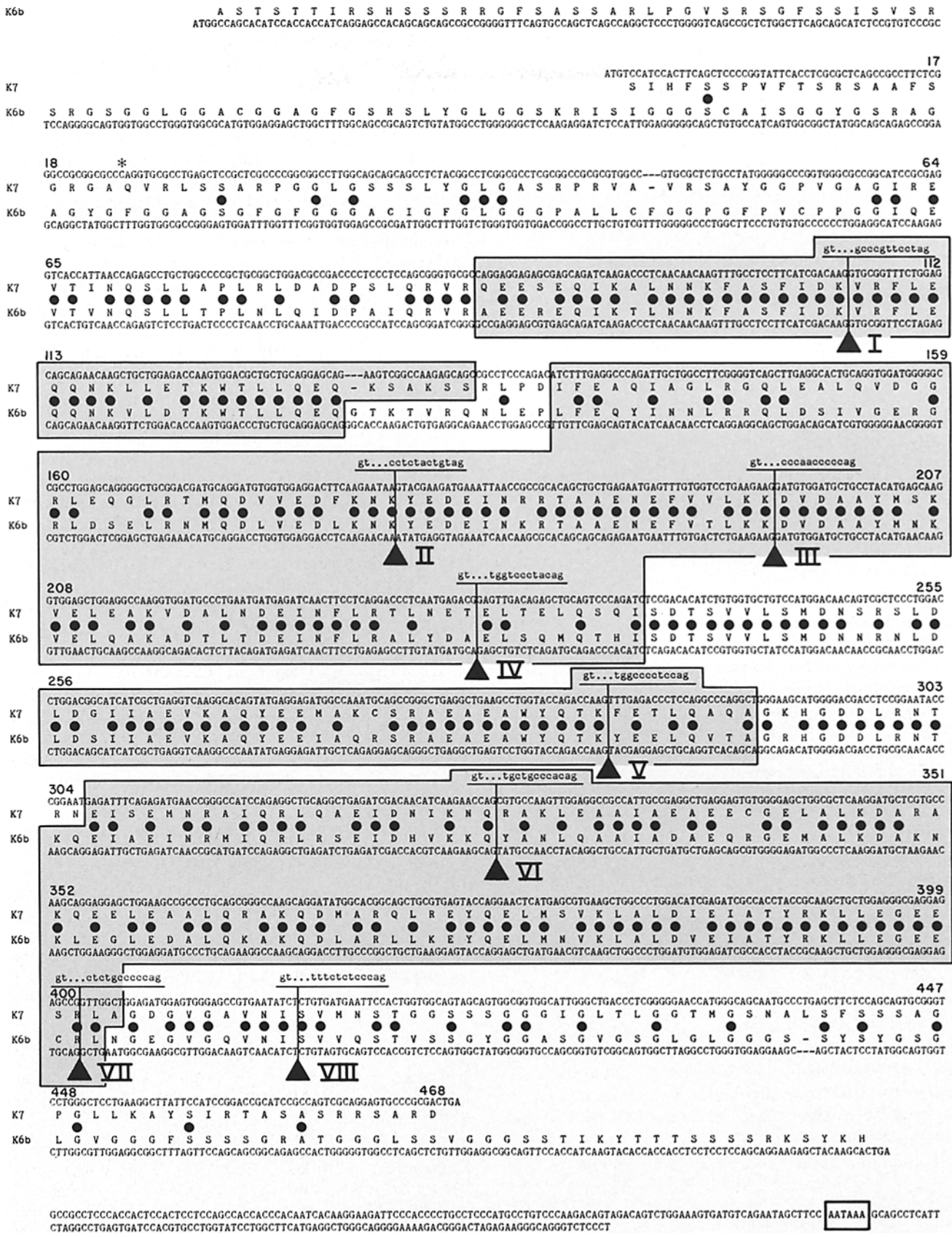
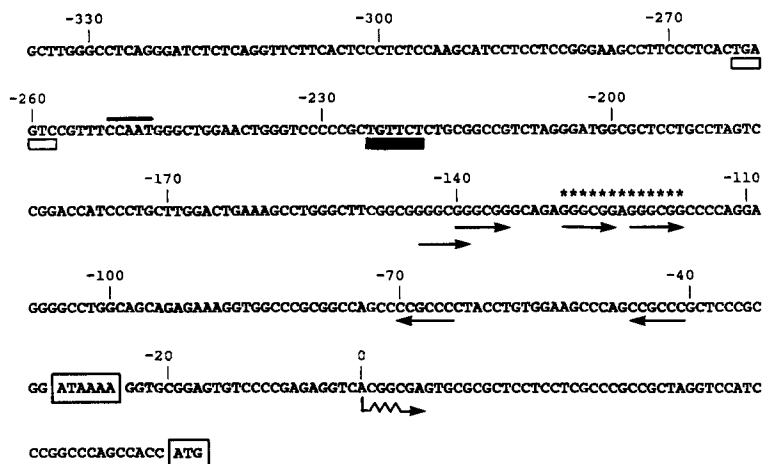


Figure 4. Nucleotide and predicted amino acid sequence of the type II human simple epithelial keratin gene contained in the genomic clones GK7-5' and GK7-3'. The coding and 3' noncoding sequences of the human K7 gene are shown with 144 nucleotides per line. The sequence is shown aligned with the coding portion of a human epidermal type II keratin gene encoding K6b (Tyner et al., 1985). Solid circles indicate the amino acid residues identical for K7 and K6b. The two sequences show 56% homology at the amino acid level and 69% at the nucleic acid sequence level. 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 K7 gene with the K7 cDNAs sequenced previously (Glass et al., 1985). The gray boxes mark the four alpha-helical domains in the K7 keratin. Throughout these domains are the heptad repeats of hydrophobic residues which identify the portions of the polypeptide that are involved in the coiled-coil interactions with a second keratin (Steinert et al., 1985; Fuchs et al., 1987). An asterisk marks the insertion of a single nucleotide, which was inadvertently omitted in a previous report (Glass et al., 1985).



mologous to the 5' cDNA probe in Fig. 1). After subsequently treating the hybrid with endonuclease S1 to digest the unprotected (single-stranded) RNA, two major mRNA-protected cRNAs of 140 and 139 nucleotides were generated (Fig. 6). These two bands most likely represent protected fragments from capped and uncapped mRNAs, respectively. Based on these data, the predicted transcription-initiation start site is located at 56 nucleotides 5' upstream from the ATG start site as shown in Fig. 5 (squiggly arrow). The length of the K7 mRNA from the transcription-initiation site to the polyadenylation signal is therefore predicted to be 1,593 nucleotides, in close agreement with the 1.7-kb size of the polyadenylated mRNA estimated from Northern blot analysis (Glass et al., 1985).

At 25 nucleotides 5' upstream from the putative transcription-initiation site, the sequence ATAAA was found. This sequence has been shown to function as a TATA box sequence for a number of human genes (for examples, see Karlsson and Nienhuis, 1985). In the region extending to 145 nucleotides 5' upstream from the transcription-initiation site were six sites (arrows, Fig. 5) bearing the sequence CCGCCC or GGGCGG. This sequence is identical to the core element shown to have strong affinity for the DNA-binding protein SP1. SP1 has been shown to enhance by 10–100× the transcription of a number of different eukaryotic genes having SP1 core elements within 40–150 nucleotides 5' upstream from their transcription start sites (for review see Kadonaga et al., 1986). Interestingly, two of these putative SP1 binding sites are located within a 12-nucleotide DNA sequence which shares 100% identity with a 5' sequence in a type I simple epithelial keratin gene encoding K18 (Kulesh and Oshima, 1988).

Of the core DNA sequences presently known to bind specific eukaryotic transcription factors, three other sequences were found to be of potential interest. One is the sequence TGAGTCC (nucleotides –263 to –258), which shares 70–90% homology with other sequences known to bind the transcription factors AP1 (Lee et al., 1987a, b; Jones et al., 1988) and *c-fos* (Rauscher et al., 1988; Franza et al., 1988). Another is the sequence CCAAT (nucleotides –252 to –248), which is identical to the core element shown to have strong affinity for a number of CCAAT-binding fac-

Figure 5. 5' upstream region of the K7 gene. The 5' upstream sequence of the K7 gene is shown beginning at the ATG translation-initiation codon and extending 392 nucleotides 5' upstream from this site. The transcription-initiation site is marked by a squiggly arrow. Possible SP1 binding sites are indicated with arrows showing the direction of the SP1 consensus sequence (GGGCGG) relative to the K7 coding strand. The sequence TGTTCT corresponding to the core element of the glucocorticoid receptor DNA binding domain is underscored by a solid black bar. A sequence TGAGTCC sharing 70–90% homology with known AP1/*fos* binding sites is underscored by an open bar. A CCAAT box is indicated by an overhead thin bar. Asterisks mark a 12-bp sequence sharing 100% identity with a 5' upstream sequence of a human simple epithelial type I keratin gene encoding K18 (Kulesh and Oshima, 1988).

tors (Dorn et al., 1987). The other is the sequence TGTTCT (nucleotides –225 to –220), which is identical to the core sequence of the DNA element necessary for induction of transcription by glucocorticoids (Zaret and Yamamoto, 1984) and progesterone (Strahle et al., 1987). Whether any

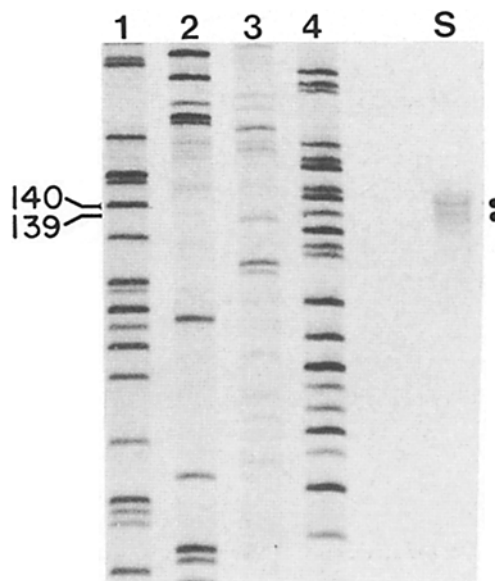


Figure 6. Mapping the transcription-initiation site of the K7 gene. A 600-bp fragment extending from the Sac I site at 75 nucleotides 3' from the putative ATG translation-initiation codon to the Hinc II site at 525 nucleotides 5' from the ATG was subcloned into pGEM3 (pGK7.S1, see Fig. 3). Plasmid pGK7.S1 was used as a template to make <sup>32</sup>P-labeled cRNA complementary to the coding strand of the K7 mRNA. Radiolabeled cRNA was hybridized with LP9 mRNA in solution as described in Materials and Methods. Hybrids were subsequently treated with S1 endonuclease to digest any unprotected, single-stranded RNA sequences. Hybrids were resolved on a sequencing gel. Four dideoxy reactions (A, C, G, and T) using a known primer and plasmid sequence were included as markers (lanes 1–4, respectively). Two S1 bands of 140 and 139 nucleotides most likely represent nuclease S1 bands from capped and uncapped mRNAs respectively. Based on these fragment sizes, the major transcription-initiation site is located 56 bp 5' upstream from the ATG start site (see Fig. 5).



of the putative regulatory sequences identified in the 5' upstream region of the K7 gene are important in regulating K7 expression in simple epithelial tissues remains to be determined.

### The Gene Contained in GK7-5' and GK7-3 Encodes a Functional K7 Protein

The previous identification of two K7 cDNAs (KC-2 and KC-1) was based on positive hybridization and translation analyses (Glass et al., 1985). To verify that this assignment was correct, we prepared a plasmid which would enable us to transfect and express the K7 gene in a variety of different cells, and also to distinguish the transfected gene product in cells which express endogenous K7. The constructs which we made for this purpose are illustrated in Fig. 7. The hybrid gene is composed of (a) 3 kb of K7 gene sequence 5' from the translation-initiation site, (b) the K7 coding sequence extending from the translation-initiation site to the Eco RI site (eliminating 53 amino acid residues of the nonhelical carboxy-terminal end domain of the protein), (c) an oligonucleotide encoding the antigenic portion (15 amino acid residues) of the neuropeptide substance P (Albers and Fuchs, 1987) placed immediately 5' from the translation-stop codon, and (d) the 3' noncoding sequence and polyadenylation signal of the K14 gene. This hybrid gene was cloned into plasmid pJay2, containing the 72-bp repeats of the SV-40 enhancer (pJ2-K7·P; Fig. 7, top), or plasmid pGEM3, containing no additional eukaryotic regulatory elements (pGEM3-K7·P; Fig. 7, bottom). Complete details of the construction of these hybrid plasmids is given in Materials and Methods.

To identify the transfected gene product in the absence of any endogenous keratins, pJ2-K7·P was first transfected into mouse NIH 3T3 fibroblasts, which only express vimentin IFs (Fig. 8). Immunofluorescence staining with an mAb

(NCI/34) specific for substance P (Fig. 8 A) revealed the presence of K7·P in ~2–5% of the transfected cells. The mAb (RCK105) monospecific for human K7 also detected K7·P in transfected 3T3 cells (Fig. 8 B), confirming that the protein encoded by the cDNAs KC-1 and KC-2 and by the gene contained in GK7-5' and GK7-3' is in fact bona fide K7.

To verify that the K7·P protein is capable of assembling into IFs, we transfected pJ2K7·P into PtK2 cells, which express endogenous keratin K7, in addition to K8, K18, some K19, and vimentin. Staining with the anti-P antibody NCI/34 showed seemingly normal filamentous staining in 85% of the transfected cells (Fig. 8 C), with ~15% of the cells showing some punctate staining (Fig. 8 D). Colocalization of the K7·P filaments with the endogenous keratin filament network was demonstrated by double immunofluorescence using both anti-P antibody (Fig. 8 E) and an anti-K8 antibody, LE41 (Fig. 8 F).

In the few cases where punctate staining was observed (Fig. 8 D), the aggregate structures showed little, if any, costaining with antibodies that recognize the endogenous type I keratins (data not shown). It seems most likely that these cells represent cases where there was not enough endogenous type I keratin to form filaments with all of the available K7 protein.

### K7 Expression Is Sensitive to Retinoic Acid in PtK2 Cells

Previous studies in our laboratory indicated that removal of vitamin A from the serum in human mesothelial (LP-9) cell cultures causes a near complete inhibition of keratin mRNA expression without affecting vimentin mRNA levels (Kim et al., 1987). Retinoic acid at  $1 \times 10^{-6}$  M restored the expression of the simple epithelial keratins K7, K8, K18, and K19 in these cells.

To determine whether the expression of K7 is dependent upon retinoids in other simple epithelial cell types, we analyzed IF proteins from PtK2 cells cultured in medium containing (a) delipidized (vitamin A-depleted) serum, (b) normal serum, or (c) normal serum supplemented with  $1 \times 10^{-6}$  M retinoic acid (Fig. 9). In contrast to the effect on LP-9 cells, removal of vitamin A from the serum seemed to have no significant effect upon keratin synthesis in PtK2 cells (Fig. 9 a, compare lanes 1 and 2). However, similar to LP-9 cells, addition of retinoic acid to PtK2 cultures elevated the synthesis of all keratins relative to that of vimentin (Fig. 9 a, compare lanes 2 and 3; also compare b and c). These results suggest that base-level (tissue-specific) expression of K7 in PtK2 cells may not require retinoic acid, although retinoic acid clearly has the capacity to upregulate its expression.

### Is the Expression of pJ2K7·P in Transfected PtK2 Cells Influenced by Retinoic Acid?

To begin to probe the molecular mechanism by which K7 expression is regulated, we tested the ability of retinoic acid to influence K7·P expression in transfected PtK2 cells. To accomplish this aim, it was first necessary to demonstrate that we could distinguish the expression of the K7·P protein from endogenous K7. Fig. 10 a shows the electrophoretic mobility of K7·P in pJ2K7·P-transfected NIH 3T3 cells (Fig. 10 a, lane 2) and in pJ2K7·P-transfected PtK2 cells grown in the

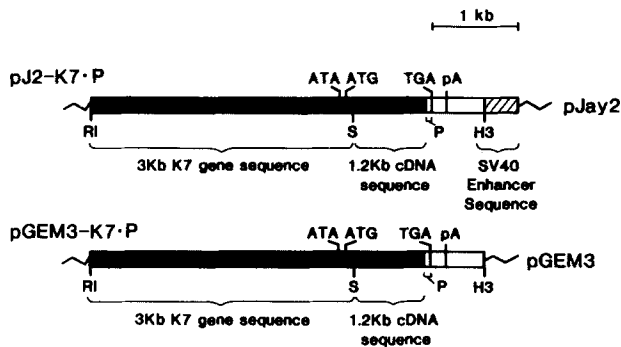
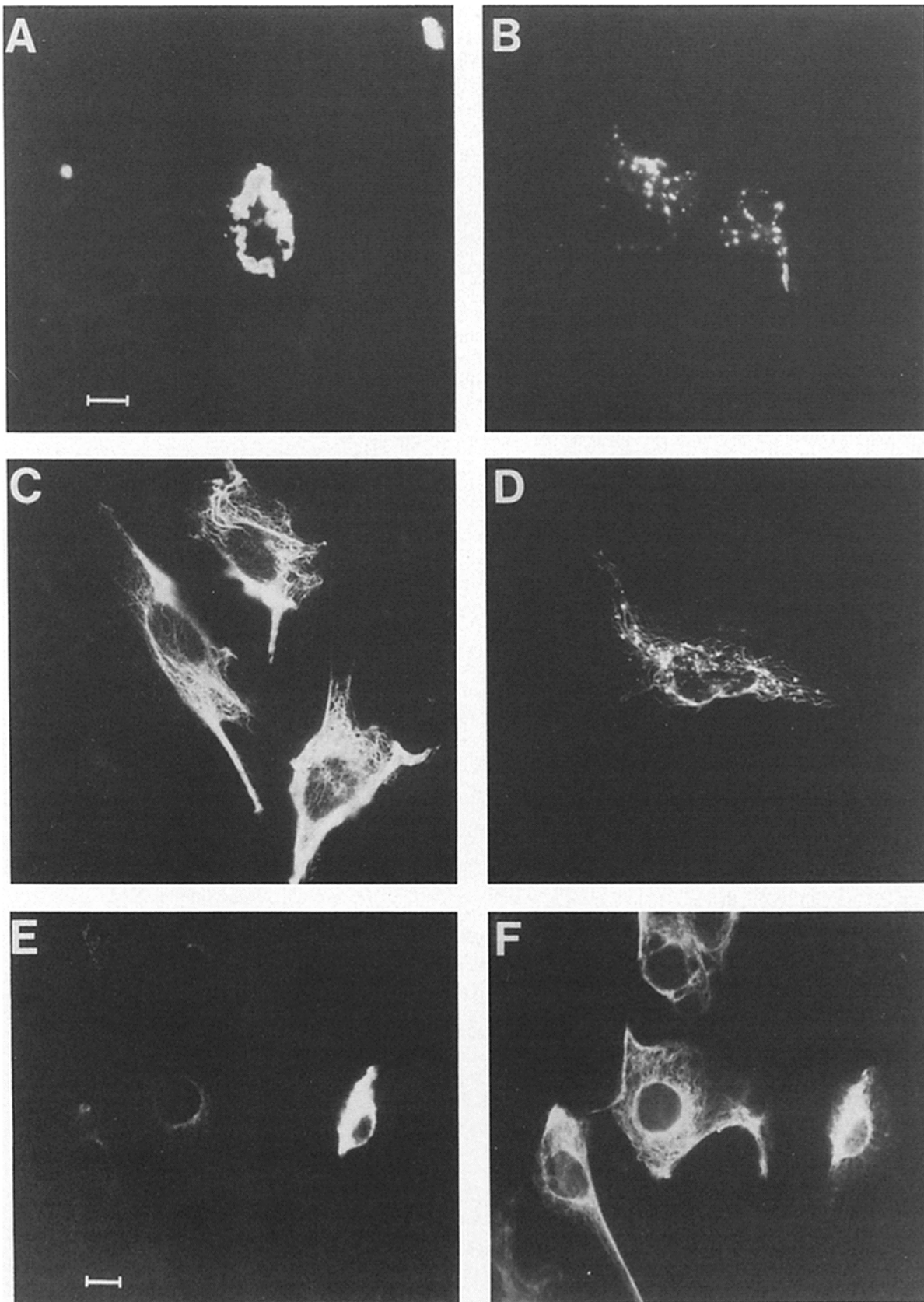
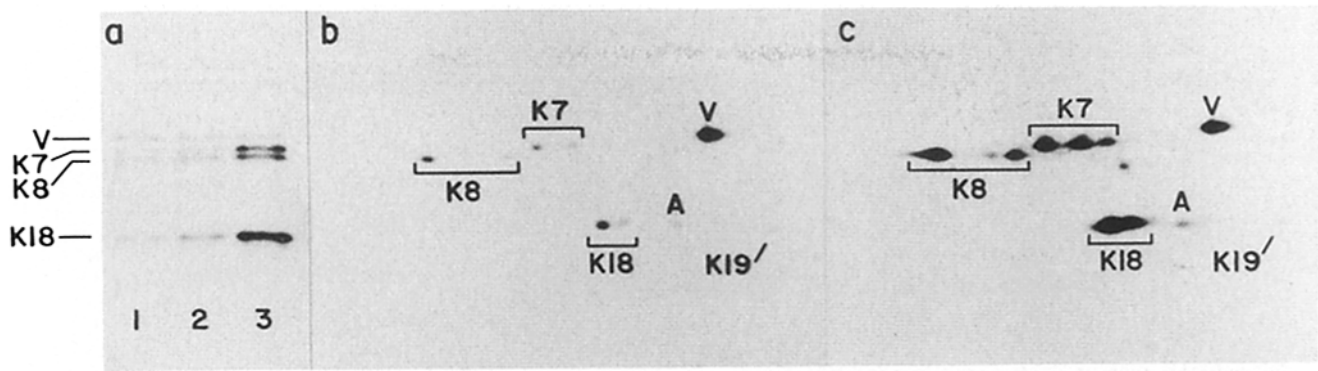


Figure 7. Construction of clones used for gene transfections. A detailed description of the steps involved in the preparation of these hybrid gene expression vectors is given in Materials and Methods. The jagged lines in the diagrams indicate vector sequences, either pJay2 (top; Land et al., 1986) or pGEM3 (bottom). The solid bars indicate K7 gene or cDNA sequences as indicated. The open bar includes the following sequences from 5' to 3': the neuropeptide substance P (Munro and Pelham, 1984; P), the K14 noncoding sequence and polyadenylation signal (pA), and additional 3' sequences from the human K14 gene in GK-1 (Marchuk et al., 1984). The hatched bar denotes SV-40 sequences, including the SV-40 72-bp repeats (marked as SV 40 Enhancer Sequence in the diagram). ATA, the recognition sequence for the binding of the transcription-initiation factor (TATA in most genes); ATG, the translation-initiation codon; TGA, the translation-termination codon.



**Figure 8.** Unequivocal identification of the human K7 gene. Plasmid pJ2-K7·P was transfected into either NIH 3T3 cells (*A* and *B*) or PtK2 cells (*C*–*F*). 65 h after transfection, cells were fixed in methanol (–20°C) and subjected to indirect immunofluorescence. To visualize the transfected gene product, cells were stained with either a rat mAb, NCI/34 (anti-P), specific for the carboxy-terminal sequence of substance P (*A* and *C*–*E*); a mouse mAb, RCK105, monospecific for K7 (*B*); or a mouse mAb specific for K8 (LE-41; anti-K8) (*F*). Antibody staining was followed by fluorescein- or Texas red-conjugated secondary antibodies as described in Materials and Methods. *E* and *F* represent a double immunofluorescence; the others are single. Bars, 10 μm.





**Figure 9.** The effect of retinoic acid on K7 expression in PtK2 cells. PtK2 cells were grown in medium supplemented with delipidized serum, normal serum, or normal serum plus  $1 \times 10^{-6}$  M retinoic acid. After 5 d, cells were radiolabeled with [ $^{35}$ S]methionine for 8 h, and IF proteins were extracted as described by Wu et al. (1982). Proteins were resolved by either one-dimensional SDS-PAGE (a) or by two-dimensional gel electrophoresis, with IEF in the first dimension and SDS-PAGE in the second dimension (b and c; O'Farrell et al., 1977). Following electrophoresis, gels were subjected to fluorography and autoradiography. Samples were from cells grown in medium containing delipidized serum (a, lane 1), normal serum (a, lane 2 and b), or normal serum plus  $1 \times 10^{-6}$  M retinoic acid (a, lane 3 and c). Identification of bands corresponding to IF proteins is indicated at the left. V, vimentin.

presence of  $1 \times 10^{-6}$  M retinoic acid (Fig. 10 a, lane 4). Relative to other IF proteins in the untransfected 3T3 and PtK2 cells (Fig. 10 a, lanes 1 and 3, respectively), K7·P showed an  $M_r$  of 51.5 kD. This value is reasonable considering the reduction in molecular mass expected from the deletion of 53 K7 carboxy-terminal amino acid residues and the addition of 15 substance P amino acid residues in the K7·P protein. The isoelectric mobility of the K7·P protein also differed considerably from the endogenous K7 protein (Fig. 10 b). This change is consistent with the net loss of three basic amino acid residues which occurred as a consequence of the carboxy-terminal alteration. Thus, the altered size and isoelectric mobility of K7·P enabled us to distinguish this protein from K7 in IF extracts from retinoic acid-treated, transiently transfected PtK2 cells.

To control for the possibility that the efficiency of transfection might be influenced by retinoic acid, we cotransfected PtK2 cells with plasmids pJK14 and pJ2K7·P. pJK14 is a plasmid containing the human epidermal keratin K14 gene driven by an SV-40 enhancer sequence (Giudice and Fuchs, 1987). In contrast to K7, K14 expression does not seem to be affected by retinoic acid (Fuchs and Green, 1981; Kopan et al., 1987). If retinoic acid is able to influence the expression of K7·P in the cotransfected PtK2 cells, then we should be able to detect an increase in K7·P expression relative to K14 expression as a consequence of retinoic acid treatment.

The results of the experiment are shown in Fig. 10 c. Although retinoic acid clearly stimulated the expression of the endogenous keratins in PtK2 cells (Fig. 10 c, compare lanes 2 and 4, with retinoic acid, with lanes 1 and 3, without retinoic acid), the level of K7·P relative to K14 was not significantly different in the cotransfected PtK2 cells with or without retinoic acid (Fig. 10, compare lanes 3 and 4).

One possible explanation for the lack of retinoic acid-mediated control of K7·P expression is that the regulation was overridden by the presence of the SV-40 enhancer on the pJ2 plasmid. To attempt to test this possibility, we repeated these experiments, this time using the plasmid pGEM3-

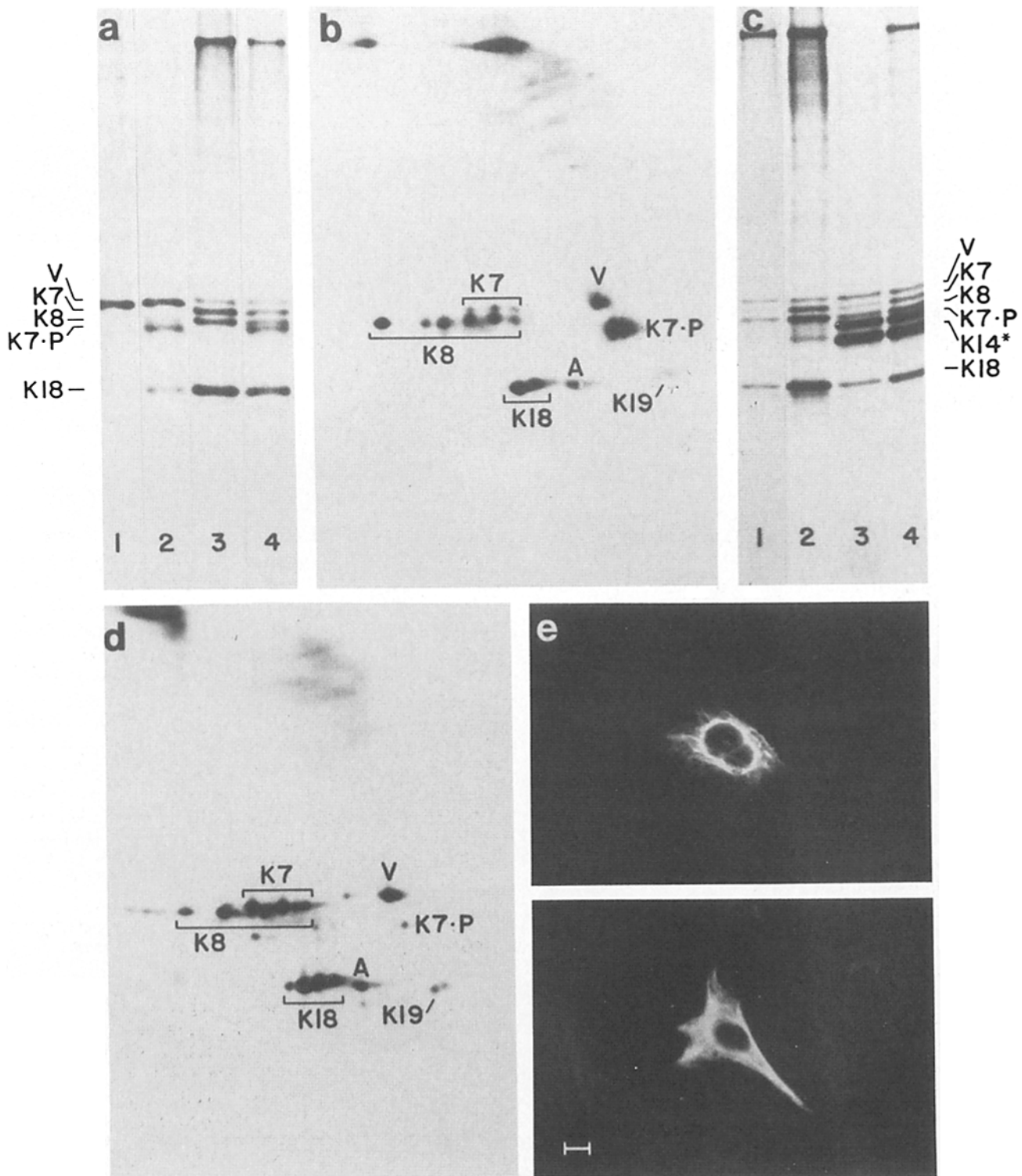
K7·P, which lacks the SV-40 sequences (see Fig. 7 for details).

pGEM3-K7·P-transfected PtK2 cells were readily detected by indirect immunofluorescence using anti-P (two examples in Fig. 10 e). These data demonstrate that the 3 kb of 5' sequences from the human K7 gene were sufficient to drive the cell-specific expression of K7·P. However, even in the presence of retinoic acid, the level of expression obtained with pGEM3-K7·P was much lower than that observed in pJ2-K7·P-transfected cells (compare Fig. 10, b and d). Although no retinoic acid response was detected with the pGEM3-K7·P construct (not shown), permanent cell lines will be necessary to provide sufficient sensitivity to unequivocally rule out this possibility.

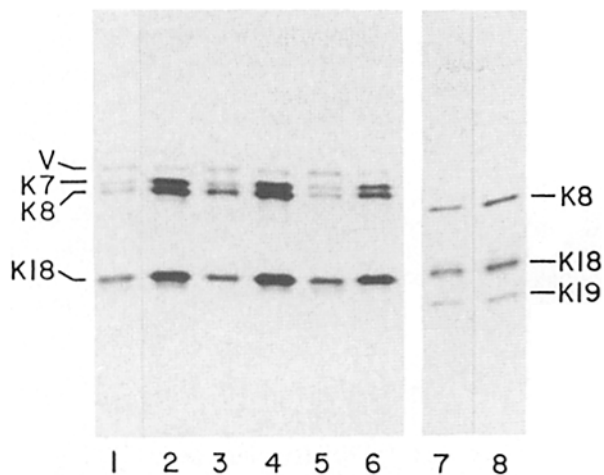
#### ***Glucocorticoids and Estrogens Do Not Seem to Influence Keratin Expression in PtK2 Cells or in MCF-7 Cells***

Recent evidence is accumulating to suggest that synthetic steroid hormone receptor-DNA binding domains are optimal when they are palindromic (Klock et al., 1987; Strahle et al., 1987). Nevertheless, most endogenous receptor DNA binding domains are not perfectly palindromic, and hence the presence of a nonpalindromic 5' K7 gene sequence sharing 100% identity with the glucocorticoid core regulatory element (Hollenberg et al., 1985; Miesfield et al., 1986; Greene et al., 1986) prompted us to investigate whether endocrine factors other than retinoids might influence keratin expression in simple epithelial cells.

Two cell types were used for our studies: PtK2 cells ( $K7^+$ ,  $K8^+$ ,  $K18^+$ ,  $K19^{+/-}$ , vimentin $^+$ ), which have not been analyzed for the presence of steroid hormone receptors, and the human breast adenocarcinoma line MCF-7 ( $K8^+$ ,  $K18^+$ , and  $K19^+$ ), which has been shown to contain significant levels of both glucocorticoid and estrogen receptors (Cato et al., 1986; Capony et al., 1987). All cells for this experiment were grown in medium containing charcoal-treated serum



**Figure 10.** Effects of retinoic acid on IF protein expression in transfected cells. NIH 3T3 cells and PtK2 cells were transfected with one or more of three different plasmids; i.e., pJ2-K7-P (Fig. 7), pGEM3-K7-P (Fig. 7), and pJK14 (containing the human K14 gene; Giudice and Fuchs, 1987). 65 h later, cells were radiolabeled with [<sup>35</sup>S]methionine. IF proteins were isolated and resolved by SDS-PAGE (a and c) or by two-dimensional gel electrophoresis (b and d) as described in the legend to Fig. 9. (a) IF samples are from (lane 1) untransfected NIH 3T3 cells, (lane 2) pJ2-K7-P-transfected NIH 3T3 cells, (lane 3) untransfected PtK2 cells cultured in the presence of normal serum supplemented with  $1 \times 10^{-6}$  M retinoic acid, or (lane 4) pJ2-K7-P-transfected PtK2 cells cultured as in lane 3. IF proteins are identified at left. V, vimentin. (b) Two-dimensional gel electrophoresis of the sample from a, lane 4. Note that K7-P migrates with an  $M_r$  of 51,500 and a p*K*i which is slightly more acidic than vimentin (V). (c) IF samples are from (lane 1) untransfected PtK2 cells cultured in normal medium, (lane 2) untransfected PtK2 cells cultured in normal medium supplemented with  $1 \times 10^{-6}$  M retinoic acid, (lane 3) pJ2-K7-P/pJK14 cotransfected PtK2 cells cultured in normal medium, or (lane 4) pJ2-K7-P/pJK14 cotransfected PtK2 cells cultured in normal medium supplemented with  $1 \times 10^{-6}$  M retinoic acid. IF proteins are identified at right. Asterisk, K14, the gene product encoded by pJK14 (Giudice and Fuchs, 1987). (d) Two-dimensional gel electrophoresis of an IF sample from PtK2 cells transfected with plasmid pGEM3-K7-P. Cells were cultured in the presence of normal medium supplemented with  $1 \times 10^{-6}$  M retinoic acid. (e) Indirect immunofluorescence of PtK2 cells cultured and transfected as in d. Cells were stained with NCI/34 antibody against substance P. Bar, 10  $\mu$ m.



**Figure 11.** The effects of dexamethasone and estradiol on IF protein synthesis in PtK2 and MCF-7 cells. PtK2 and MCF-7 cells were grown in steroid hormone-depleted medium supplemented with either (a)  $1 \times 10^{-8}$  M estradiol, (b)  $2.8 \times 10^{-7}$  M dexamethasone, (c) both a and b, (d)  $1 \times 10^{-6}$  M retinoic acid, or (e) no added hormones. After 5 d, cells were radiolabeled with [ $^{35}$ S]methionine for 8 h, and IF proteins were subsequently isolated and purified. IF extracts from equal numbers of cells were resolved by SDS-PAGE, and the gel was subjected to fluorography and autoradiography. Samples were from (lane 1) PtK2 cells with no added hormones; (lane 2) PtK2 cells with retinoic acid; (lane 3) PtK2 cells with estradiol; (lane 4) PtK2 cells with estradiol and retinoic acid; (lane 5) PtK2 cells with dexamethasone; (lane 6) PtK2 cells with dexamethasone and retinoic acid; (lane 7) MCF-7 cells with no added hormones; and (lane 8) MCF-7 cells with dexamethasone and estradiol. IF proteins are identified at the left (PtK2) and right (MCF-7). V, vimentin. Note that PtK2 cells express vimentin and MCF-7 cells do not. The differences in electrophoretic mobilities of PtK2 and MCF-7 keratins are due to species differences (potoroo vs. human).

and no phenol red to ensure that the medium was free of endogenous steroids. Dexamethasone ( $2.8 \times 10^{-7}$  M) and/or estradiol ( $1 \times 10^{-8}$  M) was added to cells after seeding, with or without retinoic acid ( $1 \times 10^{-6}$  M). After 5 d in the presence of hormone, cells were radiolabeled with [ $^{35}$ S]methionine, IF proteins were extracted, and proteins were resolved by SDS-PAGE (Fig. 11). In PtK2 cells, estradiol occasionally seemed to have some effect on K8 expression (Fig. 11, lane 3; compare with lane 1), but this was variable and relatively small. Dexamethasone also had little or no effect (Fig. 11, lane 5; compare with lane 1), whereas retinoic acid influenced keratin expression irrespective of the extracellular levels of the other hormones (Fig. 11, lanes 2, 4, and 6). MCF-7 were similar to PtK2 cells in showing no significant change in keratin expression in response to dexamethasone or estradiol (Fig. 11, lane 8; compare with the control in lane 7). In contrast to PtK2 cells, however, MCF-7 cells showed no change in keratin synthesis in response to retinoic acid (not shown). Collectively, our data do not support the notion that the TGTTCT sequence (residues -225 to -220 in Fig. 5) might be a functional steroid hormone-responsive element for glucocorticoids. Moreover, neither dexamethasone nor estradiol seem to have appreciable effects on the pattern of keratin expression in the simple epithelial cell lines which we have investigated.

## Discussion

### The Human K7 Gene and Its Encoded Protein

In this paper, we reported the isolation and characterization of the gene encoding a K7 mRNA which is expressed in human mesothelial cells. The isolation of the gene enabled us to determine the complete coding sequence of the human K7. Moreover, we were able to verify the identity of this sequence by using an SV-40 enhancer to force the expression of the K7 gene in fibroblasts, and detecting the transfected gene product with a monospecific mAb for K7.

The foreign expression of a substance P-tagged K7 protein in fibroblasts resulted in the accumulation of the keratin into aggregates. A priori, since it is known that no single keratin by itself is competent for filament formation (Steinert et al., 1976; Lee and Baden, 1976; Hatzfeld and Franke, 1985), these results might have been anticipated. However, comparable studies using another type II keratin, K6b, resulted in the assembly of the expressed foreign keratin into filamentous structures (Giudice and Fuchs, 1987). Concomitant with the expression of K6b was the induction of an endogenous type I keratin.

K7-P is fully competent to assemble into the existing keratin network of PtK2 cells, and hence, there is nothing seemingly aberrant about the hybrid gene construct or the P-tag alteration. However, there are several other possibilities as to why the expression of these two closely related human type II keratin genes in NIH 3T3 cells might result in very different phenotypes. It could be that an endogenous type I keratin is only induced when K6b, but not K7, is expressed in the transfected fibroblasts. Alternatively, it might be that the alteration of the carboxy-terminal end of K7 to add the P-tag resulted in the removal of sequences involved in the induction process. It is also possible that induction of a type I keratin did take place in K7-P-expressing fibroblasts, but that either due to the abundance of K7 expression or to a difference in filament-forming ability, aggregates still accumulated. Further studies should help to elucidate whether the induction of type I keratins as a consequence of type II expression is a universal phenomenon or unique to K6b gene expression.

### The Human K7 Gene and the Possible Role of SP1 in K7 Expression

While the simple epithelial keratin gene encoding human K7 is similar in structure to previously sequenced epidermal keratin genes, it differs in several major respects. Most notably, the gene is a large one, spanning  $\sim 15$  kb in the human genome. Most interestingly, the sequences in the putative 5' regulatory region are markedly different from those of the closely related human epidermal gene encoding K6b. In particular, six CCGCCC sequences are located within 110 bp 5' from the ATAAA box. A recent report of the 5' upstream sequence of a type I simple epithelial keratin gene, encoding K18, reveals four possible SP1 binding sites within 140 nucleotides 5' from the transcription-initiation site (Kulesh and Oshima, 1988). In fact, two of these sites are contained in a 12-nucleotide stretch which shares 100% identity with a corresponding region of the K7 gene (*asterisks*, Fig. 6). Although binding of the regulatory protein SP1 to the human K7 and K18 genes has not yet been examined, the presence

of multiple putative SP1 binding sites is highly suggestive that SP1 plays a major role in directing K7 and K18 gene expression.

Multiple, sometimes overlapping, CCGCCC sequences 50–150 nucleotides upstream of the start site of transcription have been observed for a number of eukaryotic promoters, including the SV-40 early promoter, the herpes simplex virus, thymidine kinase promoter, the human metallothionein I<sub>A</sub> and II<sub>A</sub> promoters, and the Ha-*ras* promoter (for review, see Kadonaga et al., 1986). SP1 binding sites seem to be especially prominent in the regulatory regions of genes (e.g., Ha-*ras*) that do not have a typical TATA box. Although the function of SP1 has not yet been fully elucidated, it appears to play a role not only in assisting the proper positioning of RNA polymerase II at the transcriptional initiation site, but also in increasing the transcriptional activity of promoters that harbor SP1 binding sites.

Although the 5' upstream sequences for the K8 and K19 genes have not yet been reported, the presence of multiple SP1 sites for both a type II and a type I simple epithelial keratin gene suggests the possibility that all simple epithelial keratin genes may have multiple SP1 sequences in their 5' upstream regulatory regions. If so, this difference may be a widespread one between simple epithelial and epidermal keratin genes, and it could be that the differential expression of these two sets of closely related genes might stem at least in part from tissue-specific variations in intracellular SP1 levels. Indeed, while SP1 is thought to be a transcription factor which is widely expressed in many cell types, it was initially purified from a simple epithelial cell line (HeLa), and its level may be particularly high in these cells.

### *The Regulation of Simple Epithelial Keratins by Retinoids*

It has been known for some time that concomitantly with the retinoic acid-mediated induction of differentiation, teratocarcinoma F9 cells begin to synthesize simple epithelial keratins (Kurkinen et al., 1983; Oshima et al., 1983). Recently, our laboratory has shown that retinoids might also play an important role in regulating the expression of these keratins (K7, K8, K18, and K19) in adult (differentiated) simple epithelial cells, and that this response extends to the mRNA level (Kim et al., 1987). Coupled with the experimental findings reported here, it seems that this response may be unique to cells which are capable of expressing both vimentin and keratin. Moreover, our data indicate that the seemingly nonretinoid-mediated, base level of keratin expression may vary widely among simple epithelial cells. Although the function of IFs remains to be elucidated, it may be that each epithelial cell requires a certain level of IF protein (either vimentin, keratin, or both) to fulfill the structural requirements of its cytoskeleton and that this base level cannot afford to be regulatable by extracellular factors.

The molecular mechanisms underlying the positive regulatory effects of retinoids on simple epithelial differentiation have not yet been elucidated. A number of years ago, it was suggested that retinoids might act in a manner analogous to steroid hormones (Chytil and Ong, 1979; Liau et al., 1981), a notion which received support when several cytosolic-binding proteins were identified: two for retinol and one for retinoic acid (Chytil and Ong, 1983). Although

the sequences of the two cellular retinol-binding proteins (and probably the retinoic acid-binding protein as well) bear no resemblance to steroid hormone receptors (Sundelin et al., 1985; Li et al. 1986), recent reports have demonstrated the existence of two additional receptors which bind retinoic acid with high affinity, and which have DNA binding domains that share strong homology with the glucocorticoid receptor (Giguere et al., 1987; Petkovich et al., 1987; Brand et al., 1988). If the differentiation of simple epithelial cells is mediated via these retinoic acid receptors, then it is likely that the expression of K7 is either indirectly or directly controlled via an association between a retinoic acid receptor complex and specific epithelial genes.

Although the method by which the retinoic acid receptor cDNAs were isolated certainly predicts that the retinoic acid DNA regulatory elements will be similar to that of the glucocorticoid receptor, the retinoic acid receptor DNA binding domains have yet to be identified. Is the TGTTCT sequence (residues -225 to -220 in Fig. 5) a retinoic acid receptor domain for the K7 gene? Several lines of evidence suggest that it is not. Not only have we been unable to demonstrate retinoic acid regulation of K7-P constructs containing this sequence, but in addition, we have previously shown that the induction of simple epithelial keratin synthesis by retinoids is a relatively slow process, occurring in cultured human mesothelial cells only after ~48–72 h after transfer from vitamin A-depleted to vitamin A-containing medium (Kim et al., 1987). Hence, it seems most likely that keratin gene expression is indirectly regulated by retinoids, although more extensive studies should answer unequivocally the question of whether this or any other sequence within the K7 gene is a bona fide retinoic acid binding domain.

Irrespective of whether simple epithelial keratin expression is indirectly or directly controlled by retinoids, what is the level at which retinoids act? Previous Northern blot analyses have demonstrated that the effects of retinoids on simple epithelial keratin expression are paralleled by their effects on the corresponding keratin mRNAs (Kim et al., 1987). Although we have not yet been measured the influence of retinoids on the transcriptional rates of the simple epithelial keratin genes, the results of our transfection experiments make it less likely that retinoic acid acts at the level of keratin mRNA stability. Otherwise, our transient transfection studies using the SV-40-driven K7-P construct would have been expected to show a retinoic acid-mediated difference in K7-P expression (providing of course that the modifications caused by the introduction of the substance P tag and K14 3' noncoding sequence did not alter the K7 mRNA significantly in this regard). It seems more likely that failure of pJ2K7-P-transfected cells to mediate a retinoid response for K7-P expression resulted from a problem at the transcriptional level. For instance, the naked plasmid DNA might have assembled into an optimally active chromatin configuration, perhaps bypassing the retinoic acid response. Alternatively, while sequences necessary for base-level expression of K7 were clearly present in the construct, sequences necessary for retinoic acid regulation (either indirect or direct) might have been missing. Finally, we have not yet unequivocally ruled out the possibility that in our experiments either (a) the presence of SV-40 enhancer sequences caused dominant, constitutive expression which overrode the regulatory effects of retinoids, or (b) the SV-40 enhancer sequences caused such

abundant expression that retinoid-mediated regulatory factors were limiting in the transfected cells. As additional studies are conducted, we hope to further unravel the complex mechanisms underlying the effects of retinoic acid on epithelial differentiation in general and on K7 expression in particular.

We thank Dr. Frans Raemakers (University of Nijmegen, The Netherlands) for the anti-K7 mAb; Dr. E. Birgitte Lane (Imperial Cancer Research Fund, London, England) for the anti-K8 mAb LE41; Dr. Ed. Fritsch (Genetics Institute, Boston, MA) for the human genomic library; and Jay Morgenstern, Harmut Land, and Robert Weinberg (Massachusetts Institute of Technology, Cambridge, MA) for the pJay vector. A special thank you goes to Paul Gardner for his expert synthesis of oligonucleotide primers, Grazina Trazka for her excellent technical assistance in tissue culture, and Philip Galiga for his artistic assembly of the figures. Finally, we would like to thank Dr. Geoffrey Greene (Ben May Cancer Center, University of Chicago), Dr. Paul Gilna, and Ms. Carol Deshich for their advice concerning the estradiol/dexamethasone experiments; and Raphael Kopan, Marjorie Rosenberg, Dr. George Giudice, Dr. Angela Tyner, and Dr. Kathryn Albers (of our laboratory) for their valuable assistance in technical procedures and discussions.

This work was supported by grants from the National Institutes of Health, the Cancer Research Foundation, and the Whirlpool Foundation. E. Fuchs is a National Science Foundation Presidential Young Investigator.

Received for publication 1 April 1988, and in revised form 3 June 1988.

## References

- Albers, K., and E. Fuchs. 1987. The expression of mutant epidermal keratin cDNAs transfected in simple epithelial and squamous cell carcinoma lines. *J. Cell Biol.* 105:791-806.
- Benton, W. D., and R. W. Davis. 1977. Screening lambda gt recombinant clones by hybridization to single plaques in situ. *Science (Wash. DC)*. 196: 180-182.
- Brand, N., M. Petkovich, A. Krust, P. Chambon, H. de The, A. Marchio, P. Tiollais, and A. Dejean. 1988. Identification of a second human retinoic acid receptor. *Nature (Lond.)*. 332:850-853.
- Capony, F., M. Morisset, A. J. Barrett, J. P. Capony, P. Broquet, F. Vignon, M. Chambon, P. Louisot, and H. Rochefort. 1987. Phosphorylation, glycosylation, and proteolytic activity of the 52-kD estrogen-induced protein secreted by MCF-7 Cells. *J. Cell Biol.* 104:253-262.
- Cato, A. C. B., R. Miksicek, G. Schutz, J. Arnemann, and M. Beato. 1986. The hormone regulatory element of mouse mammary tumour virus mediates progesterone induction. *EMBO (Eur. Mol. Biol. Organ) J.* 9:2237-2240.
- Chen, E. Y., and P. H. Seeberg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA (N. Y.)* 4:165-170.
- Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* 47:45-148.
- Chou, P. Y., and G. D. Fasman. 1979. Prediction of beta-turns. *Biophys. J.* 26: 367-383.
- Chytil, F., and D. E. Ong. 1979. Cellular retinol- and retinoic acid-binding proteins in vitamin A action. *Fed. Proc.* 38:241-247.
- Chytil, F., and D. E. Ong. 1983. Cellular retinol- and retinoic acid-binding proteins. *Adv. Nutr. Res.* 5:13-29.
- Connell, N. D., and J. G. Rheinwald. 1983. Regulation of the cytoskeleton in mesothelial cells: reversible loss of keratin and increase of vimentin during rapid growth in culture. *Cell.* 34:245-253.
- Cuello, A. C., G. Galfre, and C. Milstein. 1979. Detection of substance P in the central nervous system by a monoclonal antibody. *Proc. Natl. Acad. Sci. USA.* 76:3532-3536.
- Dorn, A., J. Bollekens, A. Staub, C. Benoist, and D. Mathis. 1987. A multiplicity of CCAAT box-binding proteins. *Cell.* 50:863-872.
- Eichner, R., T.-T. Sun, and U. Aebi. 1986. The role of keratin subfamilies and keratin pairs in the formation of human epidermal intermediate filaments. *J. Cell Biol.* 102:1767-1777.
- Franza, B. R., F. J. Rauscher, S. F. Josephs, and T. Curran. 1988. The fos complex and fos-related antigens recognize sequence elements that contain AP-1 binding sites. *Science (Wash. DC)*. 239:1150-1153.
- Fuchs, E., and H. Green. 1981. Regulation of terminal differentiation of cultured human keratinocytes by vitamin A. *Cell.* 25:617-625.
- Fuchs, E., A. L. Tyner, G. J. Giudice, D. Marchuk, A. RayChaudhury, and M. Rosenberg. 1987. The human keratin genes and their differential expression. *Curr. Top. Dev. Biol.* 22:5-34.
- Giguere, V., E. S. Ong, P. Segui, and R. M. Evans. 1987. Identification of a receptor for the morphogen retinoic acid. *Nature (Lond.)*. 330:624-629.
- Giudice, G. J., and E. Fuchs. 1987. The transfection of human epidermal keratin genes into fibroblasts and simple epithelial cells: evidence for inducing a type I keratin by a type II gene. *Cell.* 48:453-463.
- 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.
- Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology.* 52:456-467.
- Greene, G. L., P. Gilna, M. Waterfield, A. Baker, Y. Hort, and J. Shine. 1986. Sequence and expression of human estrogen receptor complementary DNA. *Science (Wash. DC)*. 231:1150-1154.
- 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.
- Hatzfeld, M., and W. W. Franke. 1985. Pair formation and promiscuity of cytokeratins: formation in vitro of heterotypic complexes and intermediate-sized filaments by homologous and heterologous recombinations of purified polypeptides. *J. Cell Biol.* 101:1826-1841.
- Heynes, W., H. Van Baelen, and P. De Moor. 1967. Study of steroid-protein binding by means of competitive adsorption: application to cortisol binding in plasma. *Clin. Chim. Acta.* 18:361-370.
- Hollenberg, S. M., C. Weinberger, E. S. Ong, G. Cerelli, A. Oro, R. Lebo, E. B. Thompson, M. G. Rosenfeld, and R. M. Evans. 1985. Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature (Lond.)*. 318:635-641.
- 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.
- Jones, N. C., P. W. J. Rigby, and E. B. Ziff. 1988. Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. *Genes Development.* 2:267-281.
- Kadonaga, J. T., K. A. Jones, and R. Tjian. 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends. Biol. Sci.* 11:20-23.
- Karlsson, S., and A. W. Nienhuis. 1985. Developmental regulation of human globin genes. *Annu. Rev. Biochem.* 54:1071-1108.
- Kim, K. H., V. Stellmach, J. Javors, and E. Fuchs. 1987. Regulation of human mesothelial cell differentiation: opposing roles of retinoids and epidermal growth factor in the expression of intermediate filament proteins. *J. Cell Biol.* 105:3039-3052.
- Klock, G., U. Strahle, and G. Schutz. 1987. Oestrogen and glucocorticoid responsive elements are closely related but distinct. *Nature (Lond.)*. 329: 734-736.
- Kopan, R., G. Traska, and E. Fuchs. 1987. Retinoids as important regulators of terminal differentiation: examining keratin expression in individual epidermal cells at various stages of keratinization. *J. Cell Biol.* 105:427-440.
- Kulesh, D. A., and R. G. Oshima. 1988. Cloning of the human keratin 18 gene and its expression in nonepithelial mouse cells. *Mol. Cell. Biol.* 8:1540-1550.
- Kurkinen, M., A. R. Cooper, D. P. Barlow, J. R. Jenkins, and B. L. M. Hogan. 1983. Gene expression during parietal endoderm differentiation in mouse embryos and teratocarcinoma cells. In *Teratocarcinoma Stem Cells*. L. M. Silver, G. R. Martin, and S. Strickland, editors. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 389-402.
- Land, H., A. C. Chen, J. P. Morgenstern, L. F. Parada, and R. A. Weinberg. 1986. Behavior of myc and ras oncogenes in transformation of rat embryo fibroblasts. *Mol. Cell. Biol.* 6:1917-1925.
- Lane, E. B. 1982. Monoclonal antibodies provide specific intramolecular markers for the study of epithelial tonofilament organization. *J. Cell Biol.* 92: 665-673.
- Lee, L. D., and H. P. Baden. 1976. Organization of the polypeptide chains in mammalian keratin. *Nature (Lond.)*. 264:377-379.
- Lee, W., A. Haslinger, M. Karin, and R. Tjian. 1987a. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. *Nature (Lond.)*. 325:368-372.
- Lee, W., P. Mitchell, and R. Tjian. 1987b. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell.* 49:741-752.
- Li, E., L. A. Demmer, D. A. Sweetser, D. E. Ong, and J. Gordon. 1986. Rat cellular retinol-binding protein II; use of a cloned cDNA to define its primary structure, tissue-specific expression, and developmental regulation. *Proc. Natl. Acad. Sci. USA.* 83:5779-5783.
- Liau, G., D. E. Ong, and F. Chytil. 1981. Interaction of the retinol/cellular retinol-binding protein complex with isolated nuclei and nuclear components. *J. Cell Biol.* 91:63-68.
- Marchuk, D., S. McCrohon, and E. Fuchs. 1984. Remarkable conservation among intermediate filament genes. *Cell.* 39:491-498.
- Marchuk, D., S. McCrohon, and E. Fuchs. 1985. Complete sequence of a type I human keratin gene: presence of enhancer-like elements in the regulatory region of the gene. *Proc. Natl. Acad. Sci. USA.* 82:1609-1613.
- Miesfeld, R., S. Rusconi, P. J. Godowski, B. A. Maler, S. Okret, A.-C. Wilkstrom, J.-A. Gustafsson, and K. R. Yamamoto. 1986. Genetic complementation of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. *Cell.* 46:389-399.
- 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.

- Munro, S., and H. R. B. Pelham. 1984. Use of peptide tagging to detect proteins expressed from cloned genes: deletion mapping functional domains of *Drosophila* hsp 70. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:3087-3093.
- O'Farrell, P. Z., H. M. Goodman, and P. H. O'Farrell. 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell.* 12:1133-1142.
- Oshima, R. G., W. E. Howe, F. G. Klier, E. D. Adamson, and L. H. Shevinsky. 1983. Intermediate filament protein synthesis in preimplantation murine embryos. *Dev. Biol.* 99:447-455.
- Parker, B. A., and G. R. Stark. 1979. Regulation of simian virus 40 transcription: sensitive analysis of the RNA species present early in infections by virus or viral DNA. *J. Virol.* 31:360-369.
- Petkovich, M., N. J. Brand, A. Krust, and P. Chambon. 1987. A human retinoic acid receptor which belongs to the family of nuclear receptors. *Nature (Lond.)* 330:444-450.
- Quarless, S. A., and G. Heinrich. 1986. The use of complementary RNA and S1 nuclease for the detection and quantitation of low abundance mRNA transcripts. *Biotechniques.* 4:434-438.
- Ramaekers, F., A. Huysmans, G. Schaart, O. Moesker, and P. Vooijs. 1987. Tissue distribution of keratin 7 as monitored by a monoclonal antibody. *Exp. Cell Res.* 170:235-249.
- Rauscher, F. J., L. C. Sambucetti, T. Curran, R. J. Distel, and B. M. Spiegelman. 1988. Common DNA binding site for fos protein complexes and transcription factor AP-1. *Cell.* 52:471-480.
- Rheinwald, J. G. 1980. Serial cultivation of normal human epidermal keratinocytes. *Methods Cell Biol.* 21A:229-254.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
- Steinert, P. M., W. W. Idler, and S. B. Zimmerman. 1976. Self-assembly of bovine epidermal keratin filaments in vitro. *J. Mol. Biol.* 108:547-567.
- Steinert, P. M., A. C. Steven, and D. R. Roop. 1985. The molecular biology of intermediate filaments. *Cell.* 42:411-419.
- Strahle, U., G. Klock, and G. Schutz. 1987. A DNA sequence of 15 base pairs is sufficient to mediate both glucocorticoid and progesterone induction of gene expression. *Proc. Natl. Acad. Sci. USA.* 84:7871-7875.
- Sundelin, J., H. Anundi, L. Tragardh, U. Eriksson, P. Lind, H. Ronne, P. A. Peterson, and L. Rask. 1985. The primary structure of rat liver cellular retinol-binding protein. *J. Biol. Chem.* 260:6488-6493.
- Tyner, A. L., M. J. 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.
- Wu, Y.-J., L. M. Parker, N. E. Binder, M. A. Beckett, J. H. Sinard, C. T. Griffiths, and J. G. Rheinwald. 1982. The mesothelial keratins: a new family of cytoskeleton proteins identified in cultured mesothelial cells and nonkeratinizing epithelia. *Cell.* 31:693-703.
- Zaret, K. S., and K. R. Yamamoto. 1984. Reversible and persistent changes in chromatin structure accompany activation of a glucocorticoid-dependent enhancer element. *Cell.* 38:29-38.