

# Transcriptional Regulation of the Differentiation-linked Human K4 Promoter Is Dependent upon Esophageal-specific Nuclear Factors\*

(Received for publication, May 1, 1998)

Oliver G. Opitz‡, Timothy D. Jenkins‡, and Anil K. Rustgi‡§¶

From the ‡Gastrointestinal Unit and §Hematology-Oncology Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114

**The stratified squamous epithelium comprises actively proliferating basal cells that undergo a program of differentiation accompanied by morphological, biochemical, and genetic changes. The transcriptional regulatory signals and the genes that orchestrate this switch from proliferation to differentiation can be studied through the keratin gene family. Given the localization of keratin 4 (K4) to the early differentiated suprabasal compartment and having previously demonstrated that targeted disruption of this gene in murine embryonic stem cells results in impairment of the normal differentiation program in esophageal and corneal epithelial cells, we studied the transcriptional regulation of the human K4 promoter. A panel of K4 promoter deletions were found in transient transfection assays to be predominantly active in esophageal and corneal cell lines. A critical *cis*-regulatory element resides between –163 and –140 bp and contains an inverted CACACCT motif. A site-directed mutated version of this motif within the K4 promoter renders it inactive, whereas the wild-type version is active in a heterologous promoter system. It specifically binds esophageal-specific zinc-dependent transcriptional factors. Our studies demonstrate that regulation of the human K4 promoter is in part mediated through tissue-specific transcriptional factors.**

The esophagus is lined by a stratified squamous epithelium that is composed of proliferating basal cells and differentiated suprabasal, intermediate, and superficial squamous cells. Other sites sharing the stratified squamous epithelium include the cornea, oropharynx, larynx, skin, and anogenital tract. Basal cells undergo an exquisite program of differentiation accompanied by a series of morphological, biochemical, and genetic changes as they migrate to the luminal surface with the eventual production of flattened superficial squamous cells. Squamous cells are sloughed from the surface, which may in part be governed by processes such as senescence and apopto-

sis. There is constant renewal of inner basal cells that differentiate in the suprabasal layer. Insights into the underlying molecular mechanisms can be gained through an appreciation of the genes, as exemplified by the keratins, that govern this process and their transcriptional regulation. For example, keratins 5 and 14 are present in proliferating basal cells. Early differentiation genes in suprabasal cells include keratins 1 and 10, whereas late differentiation genes in superficial squamous cells include loricrin, profillagrin, and transglutaminase (1).

The keratins are classified as the type I and type II intermediate filaments (2). The type I keratins are acidic, range from 40 to 60 kDa and consist of at least 17 distinct proteins: 12 epithelial type I keratins (K9<sup>1</sup> through K20), and 5 hair keratins (3). The type II keratins are neutral to basic, range from 50 to 70 kDa, and consist of at least 15 distinct proteins: 10 epithelial (K1 through K8; with multiple forms of several of these) and 5 hair keratins (3). The type I keratins in humans are found in a cluster on chromosome 17, and the type II keratins are on chromosome 12. A similar situation exists in mouse with the type I keratins on chromosome 11 and the type II on chromosome 15.

Keratins are obligate heterodimers of one type I molecule and one type II molecule (4). *In vivo* specific pairing is apparently due to the tissue-specific expression patterns of the keratins (3). For example, K5 and K14 heterodimerize in basal cells and K1 and K10 heterodimerize in differentiating suprabasal cells. Although keratins are critical for providing structural strength for epithelial cells, further clues about keratin function have been gained through the discovery of mutations associated with diseases transmitted in an autosomal dominant fashion which have indicated that keratins are important in cellular organization and growth (5). For example, epidermal bullosa simplex (EBS) is characterized by intraepidermal blistering due to cell lysis within the basal layer (6). Mutations have been identified in K5 and K14 genes that act in a dominant-negative fashion to cause EBS (6). Epidermolytic hyperkeratosis (EH) involves blistering of suprabasal layers of the skin due to cell lysis and degeneration, basal cell hyperplasia, and a thickened granular layer and stratum corneum (skin ridges or scales). Mutations in K1 and K10 are found in EH (7, 8). EBS and EH have been recapitulated with murine models in which the keratins are disrupted by homologous recombination in embryonic stem cells (9, 10) or are aberrantly expressed in transgenic mice (11–13).

K4 and K13 are found in suprabasal cells of nonkeratinizing

\* This work was supported by American Cancer Society Faculty Research Award JFRA-649 (to A. R.); National Institutes of Health Grants R01-DK53377 (to A. R.) and 1P01 DE12467-01A1 (to A. R.); Deutsche Krebshilfe Grant D/96/17197 (to O. O.); a Glaxo Wellcome Institute for Digestive Health basic research award (to T. J.); and Center for the Study of Inflammatory Bowel Disease Award 5P30DK43357-08 (to T. J.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF066051.

¶ To whom correspondence should be addressed: Gastrointestinal Unit, Jackson 904, Massachusetts General Hospital, 50 Blossom St., Boston, MA 02114. Tel.: 617-724-3740; Fax: 617-726-3673; E-mail: rustgi@helix.mgh.harvard.edu.

<sup>1</sup> The abbreviations used are: K, keratin; EBS, epidermal bullosa simplex; EH, epidermolytic hyperkeratosis; bp, base pair(s); EBV, Epstein-Barr virus; PCR, polymerase chain reaction; UTR, untranslated region; EMSA, electrophoretic mobility shift assay; Tricine, *N*-tris(hydroxymethyl)methylglycine; KSF, keratinocyte-specific factor; GKLF, Gut-enriched Krüppel-like factor; CBP, cAMP-response element-binding protein.

TABLE I

Oligonucleotide primer sequences for PCR-based cloning of the human K4 promoter (A), construction of K4 promoter 5' deletion constructs (B), and K4-163 promoter mutant constructs (C)

A, gene-specific primers 1 and 2 (GSP 1 and 2) are designed as non-overlapping, nested PCR primers at the 5' end of the human keratin 4 cDNA. Adapter ligated primer 1 and 2 (AP1 and 2) were provided by the manufacturer. B, the PCR primers for 5' deletion constructs with the corresponding nucleotide position within the K4 promoter are indicated. A *SalI* restriction enzyme site at the 5' end of each construct and a *BglIII* restriction enzyme site at position +59 were added to facilitate cloning. C, the overlapping PCR primers for the K4-163 mutant constructs were introduced using the overlap extension site directed mutagenesis technique (30). A *SalI* site was introduced at position -163 bp to facilitate cloning.

A. Cloning the human keratin 4 promoter	
GSP1	5'-AGGCACCTCTCTTGCCACCGCCTACAAT-3'
GSP2	5'-CTCGGACACACTGCTGTCTGGCAATCAT-3'
B. Human keratin 4 promoter plasmid constructs	
K4-1040	5'-GCGTCGACAAAAACATTAGCTCTGAGAA-3'
K4-940	5'-GCGTCGACACCATTACAGAAAACCAGAA-3'
K4-540	5'-GCGTCGACTACAGAGCACTCCATTAAATC-3'
K4-340	5'-GCGTCGACGTGATGGGGTGTCTTTTCTG-3'
K4-163	5'-GCGTCGACCAGATGACTTCTGTAGGTGT-3'
K4-140	5'-GCGTCGACTAGTGACATGCTCAACGGGT-3'
K4+59	5'-GAAGATCTGGCTGCAGAGAGCGAGCTG-3'
C. Human keratin 4 mutant promoter constructs	
K4-163 K <sub>MT1</sub> S	5'-TGTGCCCTCGTCGACCAGAGTTGAAAGGTAGGTGTG-3'
K4-163 K <sub>MT1</sub> AS	5'-GAGCATGTCACCTAAACACACCTACCTTTCAACTCTG-3'
K4-163 K <sub>MT2</sub> S	5'-GCCCTCGTCGACCAGATGACTTCTGTCTAACATTTTAGT-3'
K4-163 K <sub>MT2</sub> AS	5'-CCCGTTGAGCATGTCACTAAAATGTTAGACAGAAGTCAT-3'

stratified squamous epithelia (14). K4 is a type II keratin (59 kDa) the type I partner of which is K13. K4 has highest expression in the esophagus and cornea, but to a lesser extent is also found in the tongue, pharynx, larynx, and the anus (15). It appears during fetal development while these tissues differentiate. White sponge nevus is a benign autosomal dominant disorder characterized by thickened, white opalescent spongy-fold mucosa, primarily in the mouth but also in the esophageal epithelium, and has been shown recently to be due to a mutation in K4 (16). Recently, we have developed a model in which the mouse K4 gene has been disrupted by homologous recombination in embryonic stem cells (17). K4 homozygous null mice have a phenotype largely restricted to the esophagus and cornea. There are severe disturbances in esophageal epithelial differentiation manifest by loss of the maturation sequence from the basal to the superficial squamous layers (17). There is also evidence of basal cell hyperplasia in the cornea (17). Thus, K4 appears important in maintaining the differentiation phenotype in a highly tissue-specific fashion, and these findings provide additional basis for studying the molecular mechanisms underlying the transcriptional regulation of the K4 promoter.

It appears that a particular combination of nuclear factors in a tissue that leads to the transcription of specific keratins. Illustrative of the regulation of basal cell promoters, AP2 binds the KER1 element in the K14 promoter (18, 19) and a non-phorbol ester-responsive AP1 regulates the bovine K5 promoter (20). Suprabasal cell promoters that have been studied include the K1 promoter, where a 3' element is keratinocyte-specific and Ca<sup>+</sup> dependent (21). Additionally, the K3 promoter has two *cis*-regulatory elements within 300 bp of the transcription start site that are regulated by Sp-1-related and NF- $\kappa$ B-related factors (22, 23).

Apart from the regulation of keratin cellular promoters, viral gene expression is mediated through overlapping mechanisms in the stratified squamous epithelium. The human papillomavirus-18 C-enhancer is regulated by an apparent tissue-restricted factor designated KRF-1 (24) as well as the ubiquitous Oct-1 and AP1. Our previous studies have demonstrated that the Epstein-Barr virus (EBV) ED-L2 promoter is active in suprabasal cells of the tongue and esophagus in transgenic mice (25). Furthermore, the basal ED-L2 promoter activity is attributable to a 70-kDa keratinocyte-specific factor (KSF) that is zinc-dependent and binds a CACACCT motif (26). The ED-L2

promoter is also regulated by phorbol ester, a factor that activates the differentiation program in stratified squamous epithelial cells (27). Phorbol ester induces ED-L2 promoter activity in part through USF and Gut-enriched Krüppel-like factor (GKLF) (27, 28).

We have now characterized the human K4 promoter for *cis*-regulatory elements and key *trans*-acting factors. Importantly, there is a region from -163 to -140 bp that contains an inverted CACACCT motif that contributes significantly to K4 promoter activity. When mutated, endogenous K4 promoter activity is abrogated. This *cis*-regulatory element binds esophageal-specific transcriptional factors, indicating that the suprabasal K4 promoter is in part regulated in a highly tissue-restricted fashion.

#### EXPERIMENTAL PROCEDURES

**Cloning of the Human Keratin 4 Promoter**—The human K4 promoter was cloned with a PCR-based method that allows for the identification of genomic sequence from cDNA sequence (CLONTECH promoter finder kit). The human keratin 4 mRNA sequence was obtained from GenBank accession no. X67683. Briefly, a special adapter is ligated to the ends of DNA fragments generated by digestion of human genomic DNA with *EcoRV*, *ScaI*, *DraI*, *PvuII*, and *SspI* separately. Following adapter ligation, a small amount of the DNA from each "library" is used as a template for a primary PCR reaction using an outer adapter primer (AP1) (CLONTECH) and an outer gene-specific primer (GSP1) (Table I, part A). The primary PCR reaction mixture is then diluted and used as a template for a secondary PCR reaction using the nested adapter primer (AP2) (CLONTECH) and a nested gene-specific primer (GSP2) (Table I, part A). GSP1 and GSP2 are designed to anneal at the 5' end of the known keratin 4 cDNA. PCR amplifications were performed using *Tth* polymerase mix (CLONTECH), which includes the *Tth*-Start antibody for automatic hot start. Primary PCR reactions were performed in 50- $\mu$ l volumes containing 1  $\mu$ l of ligated and diluted DNA, 40 mM Tris-HCl, pH 9.3, 85 mM KOAc, 1.1 mM MgOAc, 10  $\mu$ M adapter primer 1 (AP1) and GSP1, 10 mM dNTP, and 1  $\mu$ l of *Tth* polymerase mix. The cycle parameters were as follows: 7 cycles of denaturation at 94 °C for 25 s and annealing/extension at 72 °C for 4 min, followed by 32 cycles of denaturation at 94 °C for 25 s and annealing/extension at 67 °C for 4 min and a final annealing/extension time for 4 min. A secondary PCR reaction was performed with 1  $\mu$ l of a 1:100 dilution of the primary PCR reaction using AP2 and GSP2. The same reaction composition and cycle parameters were used except 5 and 20 thermocycles were performed, respectively. PCR products were examined on 1% agarose gels. PCR products from each library were cloned into a TA type vector (Novagen). DNA sequencing of the 5' and 3' ends was by the dideoxy-mediated chain termination method using the Sequenase version 2.0 DNA sequencing kit (U.S. Biochemical Corp.), vector primers T7 sense and M13 antisense (Novagen), and sequence-specific primers.

Sequence analysis was performed using GenBank.

**Construction of the Human K4 Promoter Reporter Genes and Deletion Constructs**—A 1.1-kilobase pair human K4 5'-UTR fragment was isolated by PCR amplification to remove its putative endogenous translation start site (+59) with denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 1 min for 30 cycles. After confirmatory DNA sequencing, the reaction product was digested with *SalI* and *BglII*, agarose gel-purified, and ligated into the luciferase reporter gene promoterless vector, pXP2 (29), to generate the keratin 4 full-length promoter construct, designated as K4 -1040. This plasmid contained 1100 bp (-1040 bp to +59 bp) of the human keratin 4 promoter. A subsequent series of K4 promoter 5' deletion constructs (K4 -940, K4 -540, K4 -340, K4 -163, and K4 -140) were generated in a similar fashion using K4 -1040 as a template for PCR with sense primers designed at the different positions of the promoter 5' to the putative transcription start site and the antisense primer at position +59 (Table IB). The K4 -740 and K4 -185 constructs were generated by digestion of K4 -1040 with *KpnI* and *BglII* or *BamHI* and *BglII*, respectively, deleting 300 or 855 bp, respectively.

**Construction of Wild-type and Mutant Promoter Constructs**—K4 -163 promoter constructs containing mutant nucleotides spanning region -163 to -140 bp, designated as region K, were generated by site-directed mutagenesis using overlap extension PCR as described by Ho *et al.* (30). Complementary oligonucleotide primers were used to generate two DNA fragments having overlapping ends. These fragments were combined in a subsequent "fusion" reaction, in which the overlapping ends anneal, allowing the 3' overlap of each strand to serve as a primer for the 3' extension of the complementary strand. The resulting fusion product was amplified further by PCR. Specific alterations in the nucleotide sequence were introduced by incorporating nucleotide changes into overlapping oligonucleotide primers. Primer sequences for overlapping mutant primers are shown in Table IC. Primers K4 -540 and K4 +59 were used to amplify the fusion product, which was finally digested using a *SalI* site introduced at position -163 and a *BglIII* site at the 3' end with subsequent subcloning. Two mutant versions of region K were introduced to obtain the promoter constructs K4 -163 K<sub>MT1</sub> and K4 -163 K<sub>MT2</sub>.

Minimal promoter DNA constructs containing wild-type or mutant nucleotides of region K were generated by ligation of kinased double-stranded synthetic oligonucleotides (Table II) into a *BglII* site of the pGL3-promoter vector containing the minimal SV40 promoter fused to the luciferase reporter gene (Promega). All constructs were verified with DNA sequencing (U.S. Biochemical Corp.). Plasmids were purified by a modified alkaline lysis method (Qiagen).

**Tissue Culture Cell Lines and Transient Transfection Studies**—Esophageal squamous carcinoma cell lines TE-12, TE-11, T.Tn, and T.T, pancreatic cancer cell line Panc-1 (ATCC, Rockville, MD), skin cancer cell line SCC-13, cervical cancer cell line HeLa (ATCC), colon cancer cell line HT-29 (ATCC), embryonic kidney cell line 293 (ATCC), and fibroblast cell line SL 68 (ATCC), all of human origin, were cultured under standard conditions with Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% fetal bovine serum (Sigma), 100 units/ml penicillin, 100 mg/ml streptomycin (Sigma), and L-glutamine (Sigma). A rabbit normal corneal cell line SIRC (ATCC) was grown in Eagle's minimal essential medium with 10% serum and antibiotics, as indicated.

Transient transfection of all DNA constructs was carried out using the calcium phosphate precipitation technique (5 Prime → 3 Prime, Inc.). Cells were plated at a density of  $1 \times 10^6$  cells/35-mm well and transfected 24 h later with 2  $\mu$ g of the luciferase reporter plasmid and 2  $\mu$ g of pGreen Lantern-1, a reporter plasmid encoding a green fluorescent protein (Life Technologies, Inc.). Cotransfections were carried out, adding to the luciferase reporter plasmid 0.5 or 1  $\mu$ g of expression plasmids pRC/RSV-empty, pCMV-12S.E.1A, pCMV-12SRG2, CMV $\beta$ -p300, or pRC/RSV-mCBP. The transfectant mixture consisted of a 250- $\mu$ l solution of 125 mM CaCl<sub>2</sub>, 25 mM Hepes, pH 7.05, 0.75 mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM KCl, 140 mM NaCl, and 6 mM glucose. After a 12-h incubation, cells were washed twice with phosphate-buffered saline and fresh 10% serum containing medium was exchanged.

Luciferin assays were performed using luciferin, ATP, and coenzyme A (Promega) with a Monolight luminometer (Analytical Luminescence Laboratory) as described previously (26–28). Cells were harvested 36 h after transfection, washed twice with phosphate-buffered saline, lysed in 200  $\mu$ l of  $1 \times$  cell culture lysis reagent (Promega), and 40  $\mu$ l of the lysate was mixed with 100  $\mu$ l of luciferase assay reagent consisting of 20 mM Tricine, 1.07 mM MgCO<sub>3</sub>, 2.67 mM MgSO<sub>4</sub>, 0.1 mM EDTA, 33.3 mM dithiothreitol, 270  $\mu$ M coenzyme A, 530  $\mu$ M ATP, and 470  $\mu$ M luciferin. Incubations were performed in triplicate, and results were calculated as

the mean  $\pm$  S.E. values for luciferase activity. Values were then expressed as a -fold increase or decrease compared with the control for each set of experiments. Activities were expressed as the mean of at least three independent transfection experiments. In a subset of the transfection experiments, cells were examined under fluorescence microscopy (475 nm excitation peak, 490 nm emission peak) to assess green fluorescent protein production. The percentage of fluorescing cells was determined in each well and found not to vary within a given transfection experiment, indicating that transfection efficiency was uniform.

**Electrophoretic Mobility Shift Assays (EMSAs)**—Nuclear extracts from different cell lines were prepared as described previously (26–28). The protein concentration was determined by a colorimetric method (Bio-Rad protein assay).  $\alpha$ -<sup>32</sup>P-labeled oligonucleotide DNA probes were constructed with 5 pmol of double-stranded oligonucleotides (Table II), synthesized by the phosphoramidite procedure (Applied Biosystems), and purified by gel electrophoresis. Radiolabeling was done by a Klenow fill-in reaction in a buffer consisting of 10 mM Tris-HCl, pH 7.5, 5 mM MgCl<sub>2</sub>, 7.5 mM dithiothreitol, 33  $\mu$ M dATP, 33  $\mu$ M dGTP, 33  $\mu$ M dTTP, 0.33  $\mu$ M [ $\alpha$ -<sup>32</sup>P]dCTP (NEN Life Science Products), 1 unit of DNA polymerase I Klenow fragment (Amersham Pharmacia Biotech), and followed by polyacrylamide gel-purification.

EMSAs were carried out by incubating 5  $\mu$ g of nuclear extract with 5 fmol of the  $\alpha$ -<sup>32</sup>P-labeled oligonucleotide DNA probe (20,000 cpm) in a 20- $\mu$ l binding reaction containing 10 mM Tris-HCl, 50 mM NaCl, 1 mM dithiothreitol, 1 mM EDTA, 10% glycerol, and 1.0  $\mu$ g of poly(dA-dT) (Amersham Pharmacia Biotech). After incubation at room temperature for 15 min, the samples were loaded on a 6% polyacrylamide, 0.25 $\times$  Tris borate gel and electrophoresed at 10 V/cm for 2 h. The gels were dried and exposed to x-ray film (Kodak X-AR) at -80 °C for 12 h.

**Competitor Oligonucleotides and Antibodies used in EMSAs and Immune Supershift Reactions**—For competition experiments, the nuclear extract was preincubated with 100-fold excess of unlabeled double-stranded oligonucleotides (Table II) prior to the addition of the  $\alpha$ -<sup>32</sup>P-labeled oligonucleotide probe. All oligonucleotides were synthesized by the phosphoramidite procedure (Applied Biosystems) and purified by gel electrophoresis. Immune supershift assays were performed using a polyclonal anti-AP2 antibody (AP2 $\alpha$  (C-18)) and polyclonal anti-CBP antibodies (CBP1 (451), CBP2 (A-22)) (all from Santa Cruz Biotechnology). The antibody was preincubated with the nuclear extract at 4 °C for 1–12 h prior to the addition of the  $\alpha$ -<sup>32</sup>P-labeled oligonucleotide DNA probe. Other conditions for EMSAs are described above.

**Ultraviolet Light-induced Cross-linking**—The B and B<sub>MT2</sub> single-stranded oligonucleotides corresponding to the wild-type and mutant sequences of the human keratin 4 promoter (Table II) were labeled by annealing corresponding 7-mers (oligonucleotides) followed by the Klenow fill-in reaction, as described above, except that 5-bromodeoxyuridine was substituted for dTTP. Binding reactions were performed in an identical fashion, except that the reaction was for 30 min at 4 °C to inhibit protein degradation. Samples were exposed to a medium wave (312 nm) UV transilluminator (UVP Inc.) for 30 min on ice at a distance of 3 cm. Samples were then mixed with 2 $\times$  SDS-sample buffer and boiled for 5 min. Electrophoresis was carried out on a 10% SDS-polyacrylamide gel. Gels were then dried and exposed to x-ray film (Kodak X-AR) at -80 °C for 12 h.

## RESULTS

**Cloning of the Human K4 Promoter**—To obtain the 5'-untranslated region of the human K4 gene, a cloning strategy was selected to obtain sequence from genomic DNA based upon the known human K4 open reading frame sequence (GenBank accession no. X67683). This nested PCR-based cloning strategy contains five adapter ligated genomic DNA libraries and produced a single PCR product in four of five libraries (data not shown). The longest PCR fragment corresponded to 1.1 kilobase pairs of the 5'-UTR of the human K4 gene and was subcloned in a TA-type cloning vector, and several clones were subjected to DNA sequencing bidirectionally (Fig. 1). DNA sequence analysis demonstrated a 1103-bp sequence, a TATAA box at position -27, a putative transcription start site at position zero and a putative translation start site at position +59 (Fig. 1). This sequence was greater than 99% identical to a recently submitted K4 gene 5'-flanking region sequence (X97566), except our strategy yielded an additional 91 bp at the

```

-1040 AAAAACATTA GCTCTGAGAA GGGATCCATA GGTTCATCA TTCTGCCAAA
      GGGATCCCCA GCAGCAAAAA AGTTATGAAC TCCTGGTACC TGAATCTGTC
-940 ACCATPACAG AAAACCCAGAA CACACCPCAG AAACAACCTGG ACAGCAAAATG
      GAGCAGCAGA ATGCGACACCA CCAGACCCCG CTGGAGAACG CCCATCTAGA
-840 ATPTGTCAAAG CACAGATGGA AGGGAGGCCCT GCGGGGCAGC ATGGGTACCC
      AGGGCTGAGA AAGACAGTCA CCGTTTCTCT TGGGAGGGTG TTTTGTCTTC
-740 CCCATCTCC TTAGTACCTA GCTCCATGCC TGGCAGTTTC TCAGTAGCTA
      TTGGCCAAGT TGAAGTGAAT CTGGATCTAG GTTGAAGGC TAAGGCTGGA
-640 ATACAGAAAT TTTCCCTTTT TTATAGGTAC ATGGAGCCAA GAACATGTGG
      AGGAATAAGA ATGCTAAGAA TGACAATAT ACATGTGGCT TTTGACAGTT
-540 TACAGAGCAC TCCATTAACT ACTAGCTCAC TCTATCCTGA GAATGGCAA
      GGCAATAATT TGTATTACTT CCAATTTACA GGGAAAGCAA CTGGCCTCAG
-440 AGGGTGACTT GCTCCGTCGG GTTTGCAGAG CTAGGAGGTG ACAGGCTGTG
      CGCTCAAACCT CAGGCTGTCT AACTCCACAT TCTGTGGGGT GAGAGGATGG
-340 GTGATGGGGT GTCTTTTCTG GAGGAGGGAG GTGCTGTGAG CCTAGCGAGA
      TGGAGTAGCA PTGGGTGTGG GCCTGGAGCG CTGGGCCAG GCAGGGGCTT
-240 CTGATTAGGA AGCCCTGGGG CACCAGTTCA GGTTCCTCCA GAGAGTAGA
      TGATGGGATC CAGTAACCTG TGCCCTCCAG ATGACTTCTG TAGGTGTGTT
-140 TAGTGACATG CTTCAACGGGT GCGGGAAGGA TGGCTGTGTC CAAGGCCAAG
      CCAGAGATG TTTAGATTTT TTCCCTTTAT GCCTCTGCAA CCAAGCCCTG
-40 CTGCTCCAGC ACATATAAGA GACCAAGGCT GAGGGCTCCA GCACTCACCG
      GCCTGGGCC TGTCACTTCT CTGATAGCTC CCAGCTCGCT CTCTGCAGCC
+60 ATG

```

FIG. 1. **The human keratin 4 promoter sequence.** The sequence was obtained by applying a PCR-based cloning strategy. A 1.1-kilobase pair PCR product was sequenced bidirectionally. The obtained sequence was 1103 bp with a putative TATAA box at position -27, a putative transcription start site at position zero, and a putative translation start site at position +59 as indicated. Sequence analysis using GenBank revealed over 99% identity to a recently submitted K4 gene 5'-flanking region sequence (X97566) but with an additional 91 bp at the extreme 5' end.

extreme 5' end.

**The Human K4 Promoter Is Active in Esophageal and Corneal Squamous Epithelial Cells**—To assess transcriptional activity of the human K4 5'-UTR, the full-length 1.1-kilobase pair fragment was PCR-amplified to remove its ATG, and subcloned in-frame within the pXP2 luciferase reporter gene. Given that the highest expression of K4 is in the esophagus and cornea (15) coupled with an impaired differentiation phenotype in these tissues of K4 homozygous null mice (17), we assessed K4 promoter activity in corneal and esophageal cells. A prominent transcriptional activity of the K4 promoter was observed in esophageal squamous carcinoma cell lines, TE-12 (Fig. 2A), TE-11 (Fig. 2B) and T.Tn (data not shown), as well as in rabbit corneal cell line, SIRC (Fig. 2C), with transient transfection of the full-length K4 -1040 construct. The transcriptional activity observed was as strong as the potent pRSV-luc promoter (data not shown). Activity was also observed, albeit attenuated, in cell lines of epithelial origin, namely HeLa, 293, SCC-13, and Panc-1. However, the human K4 promoter was not active in T.T, HT-29, and SL-68 cell lines, the latter of nonepithelial origin (Fig. 2D). Mock-transfected cells and cells transfected with pXP2 alone showed no luciferase activity.

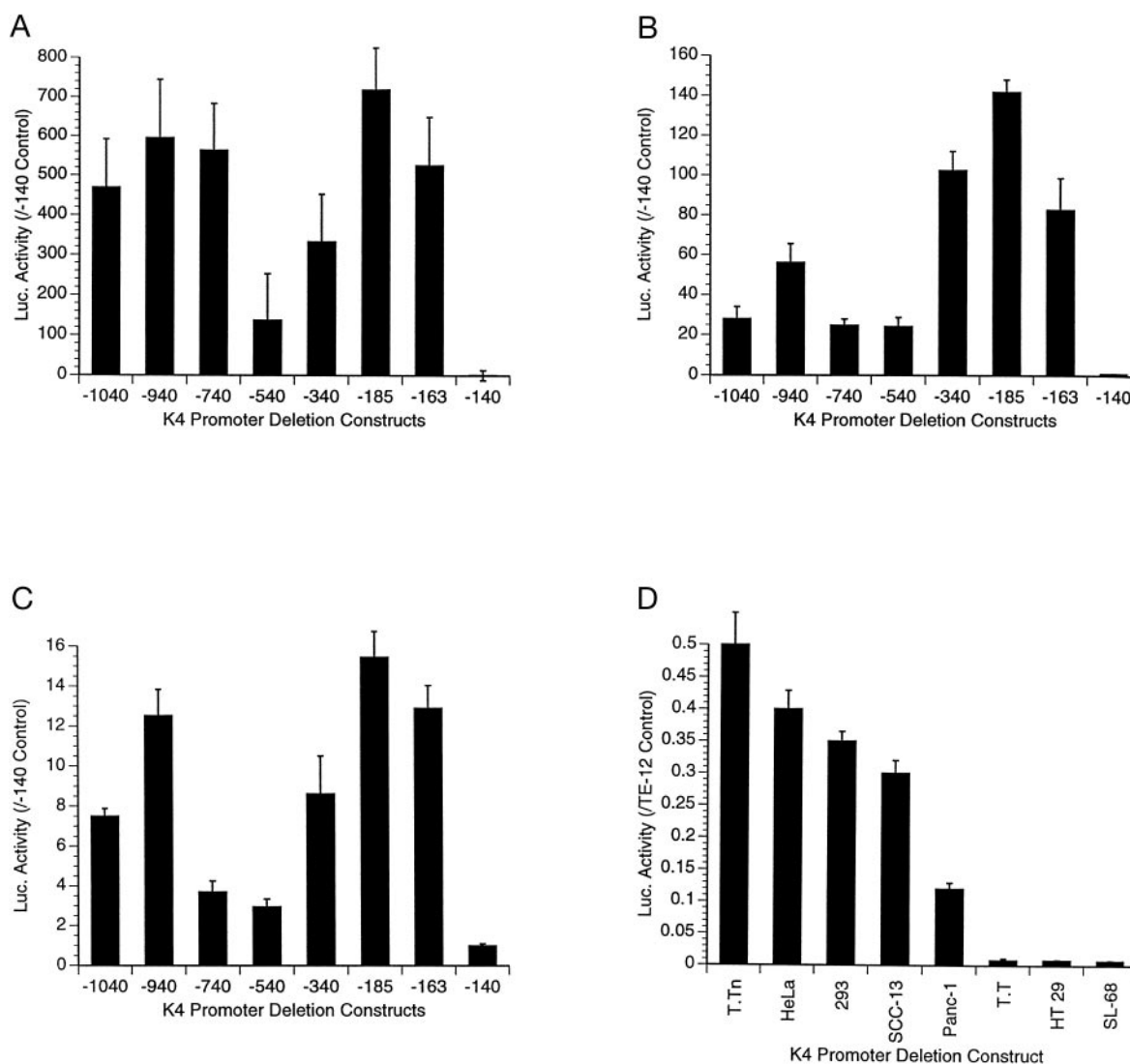
**Functional Dissection of the Human K4 Promoter**—Since 1100 bp of the human K4 promoter proved to be sufficient in achieving gene expression, this region was subjected to functional dissection analysis by serial 5' deletions. Constructs containing 940, 740, 540, 340, 185, 163, and 140 of flanking DNA sequence 5' to the putative transcription start site were used in a series of transient transfection experiments. The esophageal TE-12 and TE-11 cell lines and the corneal SIRC cell line were selected for deletion analysis (Fig. 2, A-C). This analysis revealed an increase in activity with the K4 -940 construct in TE-11 and SIRC cell lines, suggesting the presence of a negative *cis*-regulatory element at the extreme 5' end of the promoter. Promoter activity declined or did not significantly change from K4 -740 to K4 -540 in the different cell lines. An increase in activity was observed with the K4 -340 and the K4 -185 constructs, but declined again with the K4 -163 promoter constructs, suggesting the presence of different negative and positive *cis*-acting elements throughout the K4 promoter, one of which at position -281 was recently demonstrated by us to be transactivated by GKLf (28). However, the most obvious loss in promoter activity in the esophageal and corneal cells was observed between K4 -163 and K4 -140,

indicating that the intervening sequence contains *cis*-regulatory element(s), which account for over 90% K4 promoter basal activity (Fig. 2, A-C). The K4 -163 promoter had approximately 600-fold greater activity in TE-12 cells, 100-fold greater activity in TE-11 cells, and 14-fold greater activity in SIRC cells than a construct containing only the first 140 base pairs of the K4 promoter. We concentrated our analysis on *cis*-regulatory element(s) between K4 -163 and K4 -140 since the greatest reduction in promoter activity occurred in this region, strongly indicating that positive *cis*-regulatory element(s) reside within -163 to -140bp. At the same time, it can be appreciated important positive and negative *cis*-regulatory elements reside further upstream.

**Cell Type-specific K4 Promoter Activity Is Associated with a CACACCT *cis*-Regulatory Element**—The region between nucleotide positions K4 -163 and K4 -140, designated as region K, contains three sequences that resemble known *cis*-regulatory elements, namely an E-box motif (CAGATG), an AP1-like motif (TGACTTCT), and a CACACCT motif in reverse orientation (AGGTGTG), the latter demonstrated previously by us to reside within the EBV ED-L2 promoter and to bind KSF in a tissue- and cell type-specific fashion (26). Mutational analysis of the sequence between K4 -163 and K4 -140 was carried out to determine the functional consequences of these motifs in the K4 promoter. Site-directed mutagenesis using the overlap extension PCR technique was performed to introduce two mutations in the K4 -163 promoter with one eliminating the E-box motif and the AP1-like motif since they have overlapping sequences (K4 -163  $K_{MT1}$ ) and the other abolishing the inverted CACACCT motif (K4 -163  $K_{MT2}$ ) (Table IC). The wild-type and mutant constructs were transfected in TE-12 and SIRC cells. In TE-12 cells promoter activity declined only partially with the K4 -163  $K_{MT1}$  construct with a mutation in the E-box/AP1-like sequence but was almost entirely eliminated with the K4 -163  $K_{MT2}$  construct bearing the disrupted CACACCT element (Fig. 3A). In SIRC cells both K4 -163 mutant constructs retained significant promoter activity comparable to the wild-type K4 -163 promoter (Fig. 3B).

Furthermore, double-stranded oligonucleotides containing mutation of different nucleotides within region K were ligated into the heterologous promoter system pGL3-promoter, which contains a minimal SV40 promoter. The same two region K mutants were constructed (Table II), one eliminating the E-box/AP1-like motif ( $K_{MT1}$ ) and the other abolishing the inverted CACACCT motif ( $K_{MT2}$ ). The transcriptional activity of the wild-type region K in the pGL3-promoter vector was enhanced 5-fold in TE-12 cells and 2-fold in SIRC cells (Fig. 3, C and D), indicating that this region is active in a heterologous promoter system. A similar increase was observed in TE-11 cells (data not shown). Importantly, mutation in the CACACCT motif ( $K_{MT2}$ ) reduced promoter activity 3-4-fold in TE-12 cells, whereas the E-box/AP1 mutation ( $K_{MT1}$ ) had no effect. Both mutations had little effect in SIRC cells as observed in K4 -163 mutant constructs. These data suggest the critical role of the inverted CACACCT *cis*-acting element in the basal transcription of the human K4 promoter within esophageal cells, forming the basis for further biochemical analysis.

**Distinct Nuclear Transcription Factors from Esophageal Squamous Epithelial Cells Bind Region K**—In order to characterize the nuclear *trans*-acting factors that bind region K within the K4 promoter, EMSAs were performed using nuclear extracts from TE-12 and SIRC cells. These assays used as probes  $\alpha$ -<sup>32</sup>P-labeled double-stranded oligonucleotides representing the wild-type and mutant versions of region K (Table II). Two mutant versions of region K were used,  $K_{MT1}$  eliminating the overlapping E-box motif and AP1-like motif and



**FIG. 2. The human K4 promoter is active in esophageal and corneal squamous cells with *cis*-regulatory elements between  $-163$  and  $-140$  bp.** The K4 luciferase reporter gene constructs (K4  $-1040$ , K4  $-940$ , K4  $-740$ , K4  $-540$ , K4  $-340$ , K4  $-185$ , K4  $-163$ , and K4  $-140$ ) were transiently transfected into the human esophageal squamous cancer cell lines TE-12 (A) and TE-11 (B) as well the normal rabbit corneal cell line SIRC (C). Luciferase activity is expressed as -fold increase relative to the K4  $-140$  construct (mean  $\pm$  S.E. values) calculated from three independent transfections. Although not shown, the full-length K4  $-1040$  promoter activity in TE-12, TE-11, and SIRC cells was equivalent to RSV-luc activity. D, the K4  $-1040$  promoter was transfected into additional cell lines, including T.Tn, T.T, HeLa, 293, SCC-13, Panc-1, HT-29, and SL-68. Here, luciferase activity is expressed as a ratio of activity of each cell line to TE-12 cells and represents the mean of three independent transfections.

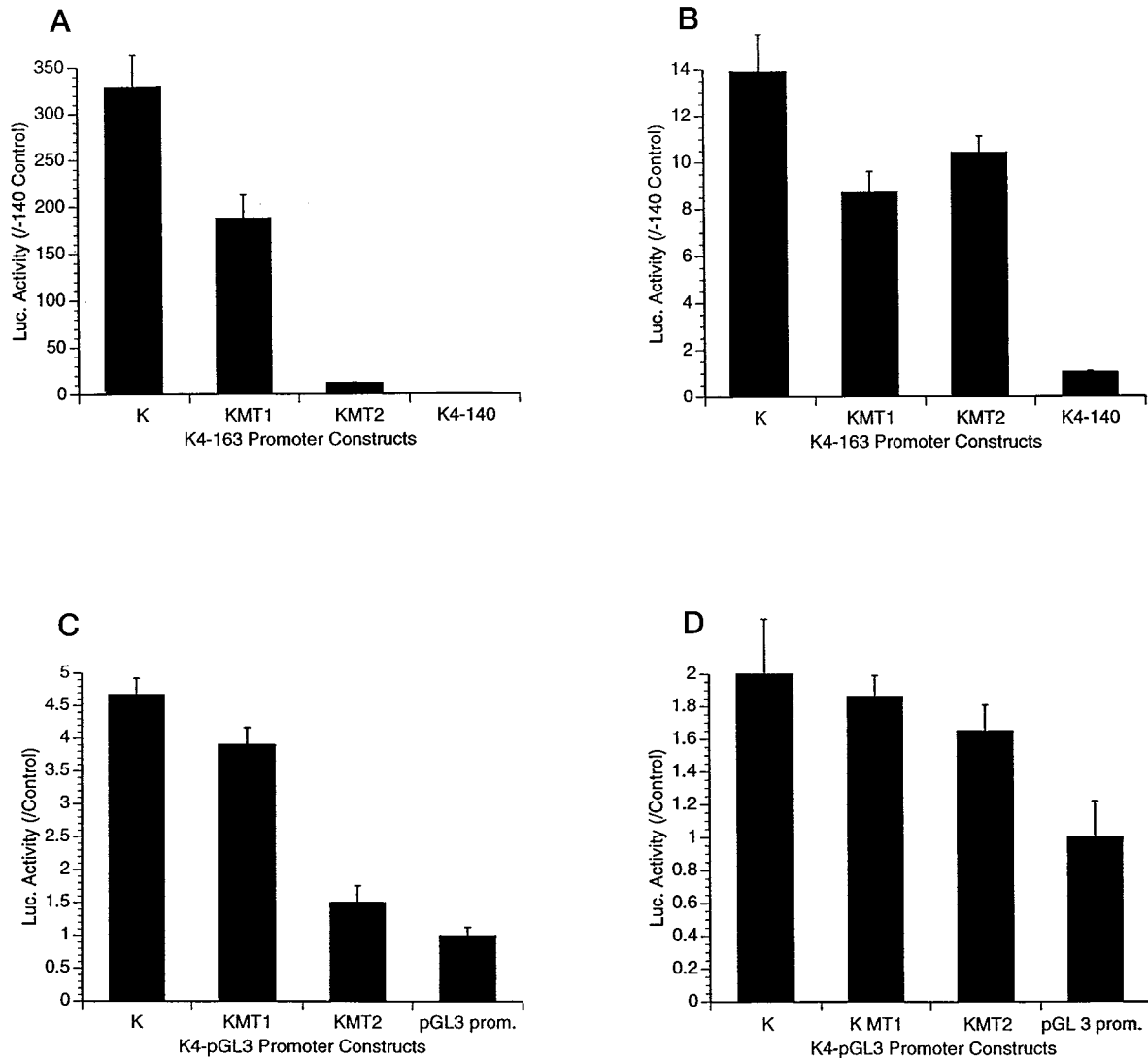
$K_{MT2}$  abolishing the inverted CACACCT motif. The wild-type K probe bound several nuclear proteins from TE-12 cells and two proteins from SIRC cells (Fig. 4A). Although no specific complex could be observed in SIRC nuclear extracts with the mutants tested comparing wild-type probe K with either mutant probe, three distinct complexes, designated as complexes a, b, and c bound the inverted CACACCT motif when comparing wild-type probe K or  $K_{MT1}$  with mutant version  $K_{MT2}$  (Fig. 4A). Competition experiments confirmed the specificity of complexes a, b, and c in TE-12 nuclear extracts (Fig. 4B). K and  $K_{MT1}$ , both bearing the wild-type CACACCT motif, competed complexes a, b, and c, whereas  $K_{MT2}$  harboring the mutated CACACCT motif did not, although there was minimal attenuation of some nonspecific complexes as well.

To determine the tissue distribution of complexes a–c, EMSAs with nuclear extracts from different cell lines of epithelial origin were performed. As shown in Fig. 4C, complexes a, b, and c were present variably in esophageal squamous cancer cell lines but not in a limited number of cell lines tested that were of other epithelial origin. Significantly, the intensity of the

complexes correlated with K4 promoter activity (Fig. 2D).

**Different Transcription Factors from TE-12 and TE-11 Cell Lines Bind the Inverted CACACCT Motif**—To further characterize the binding specificity of nuclear proteins to probe K, minimal-length oligonucleotide probes were synthesized and mutated, designated as probes B and  $B_{MT2}$  (Table II). Three specific nuclear proteins bind the inverted CACACCT motif within probe B, when comparing the wild-type and mutant versions of probe B. These three complexes are consistent with complexes a, b, and c (Fig. 5A) under similar electrophoretic mobility shift assay conditions. It should be noted that using different sizes of probes caused the pattern of nonspecific complexes to change.

As we have demonstrated previously that the CACACCT motif is a critical DNA binding site for the nearly 70-kDa KSF and that this is responsible for the majority of basal EBV ED-L2 promoter activity (26), the role of this motif within the human K4 promoter was compared with the ED-L2 promoter. A slower migrating complex corresponding to KSF could be specifically reconstituted in TE-11 nuclear extracts using



**FIG. 3. Cell type-specific K4 promoter activity is associated with a CACACCT *cis*-regulatory element.** TE-12 cells and SIRC cells were transiently transfected with wild-type and mutant constructs of region K (–163 to –140 bp) within the endogenous K4 –163 promoter deletion construct (A and B) or in a heterologous promoter system as K4-pGL3 wild-type and promoter mutant constructs (C and D). The  $K_{MT1}$  mutant constructs abolish the overlapping E-box- and AP1-like motif, and  $K_{MT2}$  mutant constructs eliminate the inverted CACACCT motif (see Tables IC and II for sequences of K,  $K_{MT1}$ , and  $K_{MT2}$ ). Luciferase activity is expressed as -fold increase relative to K4 –140 construct (A and B) or pGL3-promoter control (C and D), respectively (mean  $\pm$  S.E. values) calculated from three independent transfections.

TABLE II

Sense strand sequences of double-stranded oligonucleotides of the human keratin 4 promoter used for EMSAs and minimal promoter constructs

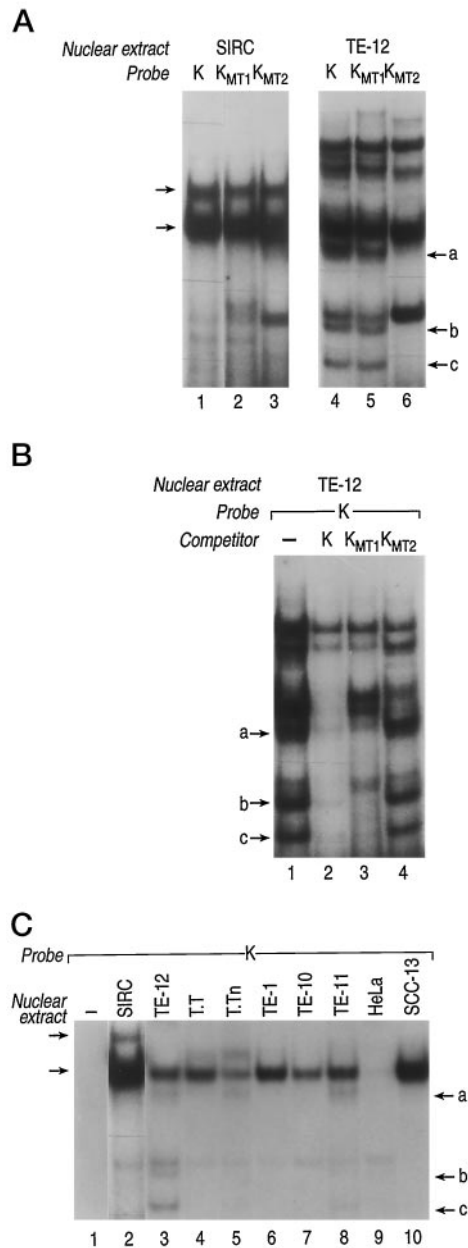
Nucleotide positions in the human keratin 4 promoter are –163 to –140 bp for probe K and –151 to –140 bp for probe B. Nucleotide position of probe F is –215 to –204 bp within the EBV ED-L2 promoter. At the 5' end of each K4 oligonucleotide, *Bgl*II (probes K) or *Hind*III (probes B) restriction enzyme sites were added to facilitate Klenow fill-in labeling and cloning. Sequences of interest are depicted in bold, and mutations are underlined.

K4-K (wild-type)	5'–GATCT <b>CAGATGACTTCTGTAGGTGTGTTA</b> –3'
K4- $K_{MT1}$	5'–GATCTCAGAGTTGAAAGGTAGGTGTGTTA–3'
K4- $K_{MT2}$	5'–GATCTCAGATGACTTCTGTCTAACATTTA–3'
K4-B (wild-type)	5'–AGCTTGT <b>AGGTGTGTTTC</b> –3'
K4- $B_{MT1}$	5'–AGCTTGTAGGTG <b>CGTTC</b> –3'
K4- $B_{MT2}$	5'–AGCTTGTAGGTGTT <b>TTTC</b> –3'
K4- $B_{MT3}$	5'–AGCTTGTATATGTGTTTC–3'
ED-L2-F (wild-type)	5'–AGCC <b>CACACCTAA</b> –3'
ED-L2-F <sub>MT</sub>	5'–AGCA <b>ACACCTAA</b> –3'

probes (probe F and F<sub>MT</sub>) (Table II) containing the CACACCT motif within the ED-L2 promoter as shown previously (26). The K4 probe B also binds KSF in TE-11 nuclear extracts. In

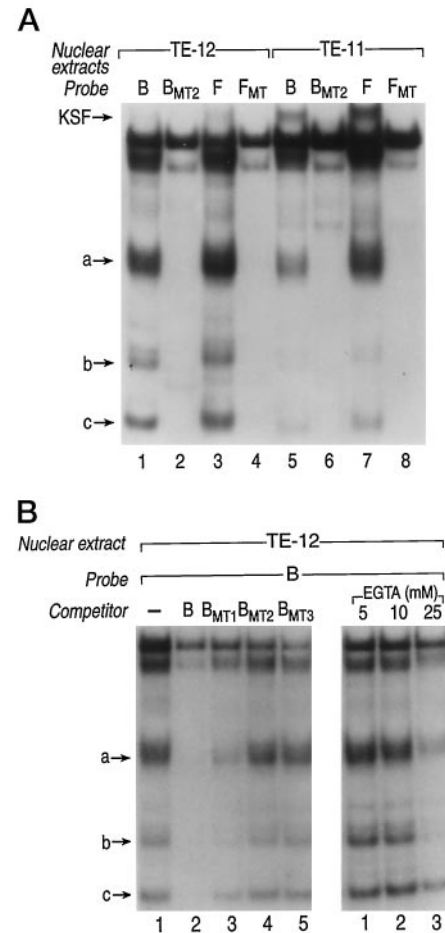
addition, the ED-L2 probe F recognizes complexes a, b, and c in the TE-12 nuclear extract, indicating that the CACACCT motif within both promoters act as a transcription factor binding site regardless of sequence orientation. In contrast, the KSF complex could only be detected faintly with either probes B or F in TE-12 nuclear extracts. Complex a was the most abundant in both TE-12 and TE-11 nuclear extracts using either probe B or F. These results demonstrate that different esophageal-specific nuclear transcription factors can bind the same motif, depending upon the cell line tested.

To further delineate the critical nucleotides necessary in the formation of complexes a, b, and c, competition experiments were done with 100-fold excess of wild-type B competitor and its serially mutated sequences. Taking advantage of the observation that the four cytosine residues within the CACACCT motif play an critical role in KSF complex formation in TE-11 cells (26), three representative mutated sequences ( $B_{MT1-3}$ ) were chosen for the inverted CACACCT motif (Table II). One contained a single base pair mutation in the first thymidine of the GTGTGGA motif ( $B_{MT1}$ ), one a mutation in the initial guanine residue ( $B_{MT2}$ ), and one had a two-base pair mutation



**FIG. 4. Distinct nuclear transcription factors from esophageal squamous epithelial cells bind region K.** The double-stranded DNA probes corresponding to region K of the human K4 promoter (see Table II for sequence) were labeled with [ $\alpha$ - $^{32}$ P]dCTP and used in EMSAs. Crude nuclear extracts were prepared from SIRC, TE-12, T.T, T.Tn, TE-1, TE-10, TE-11, HeLa, and SCC-13 cells. *A*, wild-type and mutant probes were used with K<sub>MT1</sub> abolishing the overlapping E-box and AP1-like motifs and K<sub>MT2</sub> eliminating the CACACCT motif. DNA binding patterns of the nuclear transcription factors with the three different radioactively labeled DNA probes are compared. *Labeled arrows* indicate complexes specifically bound to the CACACCT motif within region K and are referred to in subsequent figures. The *arrows* also indicate nonspecific complexes in SIRC nuclear extracts. *B*, competition assays using radioactively labeled probe K. 100-fold excess unlabeled competitor double-stranded oligonucleotides are incubated with radioactively labeled probe K. Complexes a, b, and c are markedly attenuated by competitors K and K<sub>MT1</sub> but not by K<sub>MT2</sub>. *C*, the radioactively labeled DNA probe K was incubated with crude nuclear extract prepared from cells of epithelial origin as indicated above, and *labeled arrows* indicate complexes a, b, and c. Nonspecific complexes in SIRC nuclear extracts are indicated by *unlabeled arrows*, and nonspecific complexes in all other nuclear extracts are not marked.

in the final two guanine residues (B<sub>MT3</sub>) (Table II). Mutated competitor oligonucleotides with substitutions at the cytosine/guanine positions in the inverted CACACCT motif did not

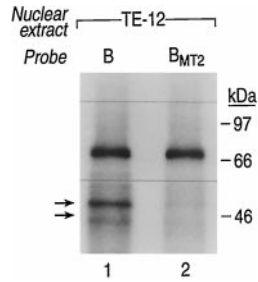


**FIG. 5. Different transcription factors from TE-12 and TE-11 cell lines bind the CACACCT motif.** Double-stranded DNA probes corresponding to region B within the K4 promoter and region F within the EBV ED-L2 promoter (see Table II for sequences of probes B, B<sub>MT2</sub>, F, and F<sub>MT</sub>) were labeled with [ $\alpha$ - $^{32}$ P]dCTP and used in EMSAs. *A*, DNA binding patterns of the nuclear transcription factors with the two different radioactively labeled probes (probes B and F) and their mutant versions (probes B<sub>MT2</sub> and F<sub>MT</sub>) incubated with TE-12 and TE-11 nuclear extracts are compared. Specific complexes are marked with *labeled arrows*. This reveals that there is a cell line-dependent complex formation with complexes a, b, and c in the TE-12 nuclear extract and complexes a and KSF in the TE-11 nuclear extract. *B*, the *right panel* shows competition assays using radioactively labeled probe B incubated with 100-fold excess of unlabeled competitor double-stranded oligonucleotides B, B<sub>MT1</sub>, B<sub>MT2</sub>, and B<sub>MT3</sub> (see Table II for sequences) and the TE-12 nuclear extract. Complexes a, b, and c are markedly attenuated by competitors B and B<sub>MT1</sub> but not by B<sub>MT2</sub> and B<sub>MT3</sub>. *B*, the *left panel* shows increasing concentrations of EGTA (5, 10, and 25 mM) with variable attenuation of complexes a, b, and c, indicating these complexes may be Zn<sup>2+</sup>-dependent.

interfere with the formation of complexes a, b, and c (Fig. 5*B*, *left panel*), although mutations in the other nucleotides did interfere. These data are consistent with what was observed previously with the CACACCT motif within the ED-L2 promoter. Overall, these competition experiments indicate that the guanine nucleotides within the inverted CACACCT motif contribute to the formation of complexes a, b, and c.

**Zn<sup>2+</sup> Chelation and Ultraviolet Light-induced Cross-linking Reveal Additional Characteristics of the DNA-binding Nuclear Proteins**—Additional EMSAs were performed utilizing EGTA to identify which factors may require Zn<sup>2+</sup> for DNA binding (31). These data revealed that complexes a and b, and to a lesser extent complex c, were attenuated (Fig. 5*C*, *right panel*), suggesting that all complexes a, b, and c may require zinc.

To estimate the molecular masses of complexes a, b, and c,



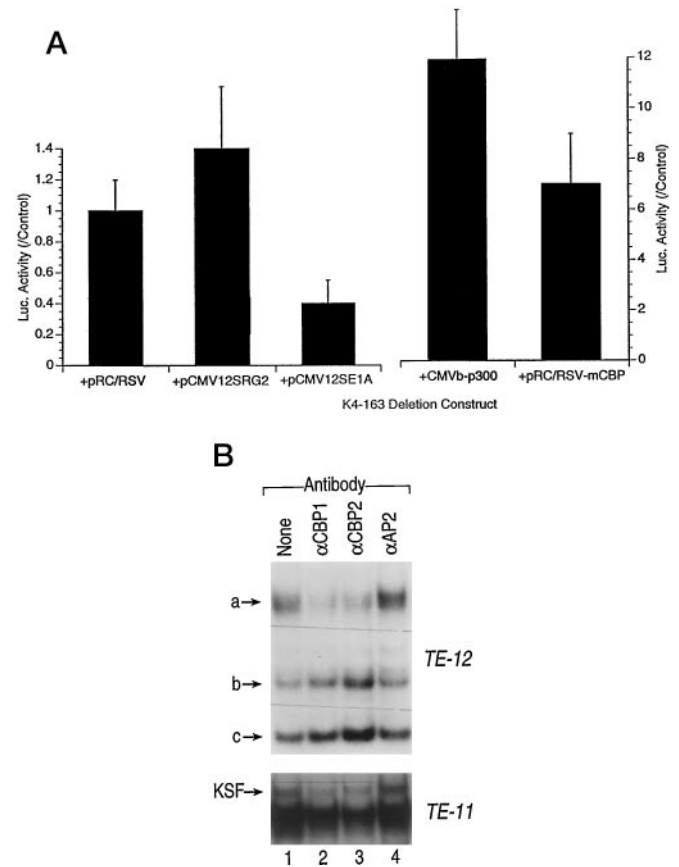
**FIG. 6. UV light-induced cross-linking experiment with radioactively labeled probe B corresponding to the CACACCT motif within the human K4 promoter.** DNA probes B and BMT2 were labeled with [ $\alpha$ - $^{32}$ P]dGTP as described under "Experimental Procedures." EMSA reactions were the same except for the following modifications: DNA probes were Klenow filled-in by annealing corresponding 7-mers, 5-bromodeoxyuridine was substituted for dTTP, and reactions were 30 min at 4 °C to inhibit protein degradation. Samples were exposed to medium wave (312 nm) UV-transilluminator for 30 min on ice at a distance of 3 cm and separated with 10% SDS-polyacrylamide gel electrophoresis. Arrows indicate the major protein complexes unique to the wild-type DNA probe B but not seen in with the corresponding mutant DNA probe BMT2. Molecular mass size markers are expressed in kilodaltons (kDa).

ultraviolet light-induced cross-linking experiments were performed. This technique takes advantage of the binding specificity of nuclear factors to wild-type sequences, but not to mutant sequences (27). EMSA reactions were performed in an identical fashion with  $\alpha$ - $^{32}$ P-labeled wild-type and mutant probes, followed by UV-induced cross-linking with separation by SDS-PAGE. Probe B cross-links to one dominant specific protein of approximately 48 kDa and to a second less abundant specific protein of 44 kDa (Fig. 6).

**Interaction With CBP/p300 Further Characterizes Esophageal-specific Nuclear Factors**—An emerging recent theme is the linkage of the basal transcriptional machinery to factors binding upstream enhancer elements through the CBP/p300 co-activators (32, 33). To further characterize the esophageal-specific factors in TE-12 cells and the previously described KSF in TE-11 cells, cotransfection studies with the K4 -163 promoter construct and CBP/p300 expression constructs were performed. Both CBP (pRC/RSV-mCBP) and p300 (CMV $\beta$ -p300) when cotransfected with the K4 -163 deletion construct or with the region K in pGL3 (data not shown) resulted in a 7–12-fold stimulation of transcriptional activity compared with basal K4 promoter activity alone (Fig. 7A). E1A (pCMV-12S.E.1A) cotransfection disrupted the K4 -163 promoter activity whereas the E1A mutant version 12SRG2 (pCMV-12SRG2) did not (Fig. 7A). To determine whether CBP/p300 may contribute to the formation of complexes a, b, or c in TE-12 cells or to KSF complex formation in TE-11 cells, EMSAs were performed using probe B (Table II) and antibodies specific to CBP/p300 in immune supershift assays. Anti-CBP/p300 antibody 451 eliminated complex a, but not b and c, in TE-12 cells. In addition, KSF was eliminated with the same antibody in TE-11 cells. The CBP antibody A-22 only partially abolished all these complexes. The former antibody recognizes amino acids 451–720 within the cAMP-responsive element-binding protein binding domain of CBP and is specific for CBP and p300. Anti-CBP antibody A-22 recognizes CBP at its amino terminus, but does not cross-react with p300. An unrelated anti-AP2 antibody serving as control did not disrupt any complexes.

#### DISCUSSION

Studies aimed at elucidating the molecular regulation of proliferation and differentiation in the stratified squamous epithelium have inherent advantages in ultimately defining the normal program that switches cells from a proliferative to



**FIG. 7. CBP/p300 transactivates the human K4 promoter possibly through interaction with esophageal keratinocyte-specific transcription factors.** A, TE-12 cells were transiently transfected with the K4 -163 promoter deletion luciferase reporter construct and co-transfected with pRC/RSV-empty, pCMV-12S.E.1A, pCMV-12SRG2, CMV $\beta$ -p300, or pRC/RSV-mCBP constructs for 24 h, followed by harvesting for the luciferase assay. Luciferase activity is expressed as -fold increase relative to the empty vector pRC/RSV cotransfection (mean  $\pm$  standard error values) and is calculated from three independent transfections. B, double-stranded oligonucleotides corresponding to region B were labeled with [ $\alpha$ - $^{32}$ P]dCTP and used in the EMSA along with antibodies to CBP/p300 (CBP1 (451), CBP2 (A22)) and AP2 (Ap2 $\alpha$  (C-18)). Antibodies were incubated with the EMSA reaction components prior to addition of the radioactively labeled DNA probe B. Arrows indicate complexes a–c and KSF binding probe B in TE-12 and TE-11 nuclear extracts, respectively. Complexes a and KSF are markedly attenuated when antibodies to CBP are used. No effect on complex pattern is observed with the AP2 antibody.

a differentiated state, facilitating identification of possible progenitor stem cells (34), characterizing epithelial response to injurious agents, and understanding how the epithelium undergoes a dedifferentiated phenotype during malignant transformation. Investigations of the keratin genes form a paradigm for elucidating underlying molecular mechanisms as they are critically linked to cellular organization and growth within the stratified squamous epithelium (35, 36).

The suprabasal compartment within the epithelium harbors cells that undergo a switch from a proliferative to an early differentiated state. It is important to delineate both the genes responsible for this switch and the transcription factors that orchestrate the pattern of gene expression in this process. In this context, there are ubiquitous keratins, such as K1 and K10, which form heterodimers, and tissue-restricted ones such as K4 and K13, which also heterodimerize. Taking advantage of K4's relatively restricted expression pattern in the esophagus and cornea (15), we have generated a model in which this gene is disrupted through homologous recombination in murine embryonic stem cells (17). K4 disruption resulted in an

impairment of the normal differentiation sequence, especially in the esophageal but also in the corneal epithelia, in young homozygous null mice but not in age-matched heterozygous or wild-type mice. This suggested to us that the transcriptional regulation of K4 might be important in understanding the commitment to early differentiation in these tissues.

After cloning and sequencing the human K4 promoter, serial deletions fused to the luciferase reporter gene were generated and transiently transfected in corneal and esophageal cell lines, among others. Recognizing that the K4 promoter has multiple positive and negative *cis*-regulatory elements, reminiscent of other keratin promoters, we nonetheless concentrated upon a key region between -163 to -140 bp as deletion of this region renders the K4 promoter functionally inactive. Known elements in this 23-bp region include an E-box, an AP1-like motif, and an inverted CACACCT motif, the latter previously shown by us to be critical in basal transcription of the suprabasal EBV ED-L2 promoter (26). Introduction of mutations in the overlapping E-box and AP1-like motifs or the CACACCT motif in the endogenous K4 promoter revealed that the latter motif is especially critical for K4 promoter activity. This was further corroborated by inserting the wild-type 23-bp region and the same mutant versions in a heterologous promoter system. Importantly, complementary biochemical analyses demonstrated that three esophageal-specific transcriptional factors bind the inverted CACACCT motif within the human K4 promoter in a sequence-specific fashion. Interestingly enough, the abundance of these three factors varies among different esophageal cell lines and correlates with K4 promoter activity within these cell lines. Indeed, the three factors are most prominent in the TE-12 cell line, but only one of these (designated "a") is abundant in TE-11. The latter cell line contains a previously described tissue- and cell-specific factor, KSF (26), whereas TE-12 contains barely detectable levels of KSF. It is tempting to speculate that esophageal suprabasal keratin as well as viral (human papillomavirus, EBV) promoters are regulated differentially by tissue-specific transcriptional factors, which may in turn respond to signals that trigger differentiation. At the same time, regulation of the K4 promoter in corneal cells is in large measure also mediated through this 23-bp region, although different factors (perhaps corneal specific) are responsible.

An intriguing finding in our studies is that co-transactivation of the K4 promoter involves some interplay between the esophageal-specific transcriptional factors in TE-11 and TE-12 cells and the CBP/p300 factors. An emerging theme has been the link of CBP/p300 to differentiation programs in different cell types, such as muscle, B-cells, and erythroid cells (37, 38). Furthermore, as is also evident in our model system, there appears to be a preferential co-activation by p300 based upon transfection studies and immune supershift assays. A differential co-transactivation by CBP/p300 has been recently described in HTLV-1 Tax protein-mediated gene expression (39) and in retinoic acid induced F9 cell differentiation (40).

Promoter analyses and transgenic mouse studies in other models of transcriptional regulation have revealed that for keratin genes, there are key sequences involved in regulating their expression (41, 42). Among the candidates implicated in orchestrating epidermal gene expression is the sequence 5'-GCCTGCAGGC-3' in the 5'-UTR of the K14 gene (18). For the K14 gene, the sequence acts in synergy with a distal element to regulate transcription in keratinocytes. Epidermal nuclear extracts contain a protein(s) that binds to this sequence and cross-reacts with antibodies against the transcription factor AP2 (19, 43). AP2-binding sites have since been found in the promoters of most epidermal and some hair-specific genes, and

where tested, they are functionally important for gene expression. Studies indicate that AP2 plays a role in K5 and K14 activation in basal cells *in vivo* (13, 44, 45). At the same time, during the conduction of our studies, limited analysis of the K4 promoter revealed that AP2 contributes to K4 transcriptional regulation in cultured HaCaT skin keratinocytes (46). However, our studies indicate more complex regulatory mechanisms are present in a tissue-specific fashion, consistent with the high expression pattern of K4 in the cornea and esophagus.

Members of the AP1 family of transcription factors are also found in the epidermis of skin. These include JunB and c-Fos, which reside primarily in the differentiating layers of the epidermis, implicating these members of the AP1 family in regulating differentiating specific functions in skin (21). Consistent with this notion is the finding that promoters active in terminally differentiating keratinocytes contain AP1 sites. In addition, the bovine K5 promoter appears to be regulated in part by a "keratinocyte"-specific AP1 factor, one that is not responsive to phorbol ester (19). Based upon our studies within the -163 to -140 *cis*-regulatory region of the K4 promoter, it does not appear that AP1 is contributing to transactivation. Another factor that appears important in the basal cell compartment is basonuclin, a zinc-finger protein (47). Basonuclin persists in cells that have withdrawn from the cell cycle, but it is absent in terminally differentiating cells (48).

Further examples of the interplay of ubiquitous and tissue-restricted factors within the stratified squamous epithelium can be demonstrated in suprabasal promoters. These include the regulation of the K3 promoter by Sp-1-like and NF- $\kappa$ B-like factors (23). Furthermore, there appears to be a coordinated action in the levels of Sp-1 and AP2, namely an increased Sp-1/AP2 ratio regulates K3 gene transcription in differentiating corneal epithelial cells (49). In addition, although some members of the POU family are expressed throughout the stratified epithelium, such as Oct-1, others such as Skn-1a and Tst-1 appear more critical in the suprabasal compartment (50). Interestingly, our previous studies also illustrate that relatively tissue-restricted factors, such as KSF and GSKL, can regulate viral promoters, such as the ED-L2 promoter, and cellular promoters, such as the human K4 promoter in this compartment (26-28).

The intermediate and superficial squamous compartments, where cells have undergone terminal differentiation, are also regulated by complex transcriptional mechanisms. A particularly interesting factor is ESE-1, an epithelial specific member of the Ets family, which localizes to this compartment and transactivates terminally differentiated linked SPRR2A and keratin 8 promoters (51).

Taken together, the stratified squamous epithelium undergoes a carefully orchestrated program of differentiation involving a complex interplay of both ubiquitous and tissues-specific transcription factors. Our novel findings with the human K4 promoter provide a model in which to study the regulation of gene expression during early differentiation.

*Acknowledgments*—We are grateful to Dr. Richard H. Goodman for the CBP/p300 expression vectors, Dr. Andrew B. Leiter for the E1A expression vectors, Dr. Takuya Inomoto for initial help with the cloning of the keratin 4 promoter, and Dr. Hubert E. Blum.

#### REFERENCES

- Byrne, C., Tainsky, M., and Fuchs, E. (1994) *Development* **120**, 2369-2383
- Fuchs, E., and Weber, K. (1994) *Annu. Rev. Biochem.* **63**, 345-382
- Moll, R., Franke, W. W., Schiller, D. L., Geiger, B., and Krepler, R. (1982) *Cell* **31**, 11-24
- Hatzfeld, M., and Weber, K. (1990) *J. Cell Biol.* **110**, 1199-1210
- Fuchs, E. (1994) *J. Cell Biol.* **125**, 511-516
- Bonifas, J. M., Rothman, A. I., and Epstein, E. H. (1991) *Science* **254**, 1202-1205
- Cheng, J., Syder, A. J., Yu, Q. C., Letai, A., Paller, A. S., and Fuchs, E. (1992) *Cell* **70**, 811-819

8. Rothnagel, J. A., Dominey, A. M., Dempsey, L. D., Longley, M. A., Greenhalgh, D. A., Gagne, T. A., Huber, M., Frenk, E., Hohl, D., and Roop, D. R. (1992) *Science* **257**, 1128–1130
9. Chan, Y., Anton-Lamprecht, I., Yu, Q. C., Jackel, A., Zabel, B., Ernst, J. P., and Fuchs, E. (1994) *Genes Dev.* **8**, 2563–2573
10. Rugg, E. L., McLean, W. H. I., Lane, E. B., Pitera, R., McMillan, J. R., Dopping-Hepenstal, P. J. C., Navsaria, H. A., Leigh, I. M., and Eady, R. A. J. (1994) *Genes Dev.* **8**, 1563–1573
11. Vassar, R., Rosenberg, M., Ross, S., Tyner, A., and Fuchs, E. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 1563–1567
12. Fuchs, E., Esteves, R. A., and Coulombe, P. A. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 6906–6910
13. Byrne, C., and Fuchs, E. (1993) *Mol. Cell. Biol.* **13**, 3176–3190
14. Takahashi, T., Shikata, N., Senzaki, H., Shintaku, M., and Tsubura, A. (1995) *Histopathology* **26**, 45–50
15. van Muijen, G. N. P., Ruitter, D. J., Franke, W. W., Achtstatter, T., Saasnoot, W. H. B., Ponc, M., Warnaar, S. O. (1986) *Exp. Cell Res.* **162**, 97–113
16. Rugg, E. L., McLean, W. H., Allison, W. E., Lunny, D. P., Macleod, R. I., Felix, D. H., Lane, E. B., and Mundro, C. S. (1995) *Nat. Genet.* **11**, 450–452
17. Ness, S., Edelmann, W., Jenkins, T., Liedtke, W., Rustgi, A. K., and Kucheralapati, R. (1998) *J. Biol. Chem.* **273**, 23904–23911
18. Leask, A., Rosenberg, M., Vassar, R., and Fuchs, E. (1990) *Genes Dev.* **4**, 1985–1998
19. Leask, A., Byrne, C., and Fuchs, E. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 7948–7952
20. Casatorres, J., Navarro, J. M., Blessing, M., and Jorcano, J. L. (1994) *J. Biol. Chem.* **269**, 20489–20496
21. Lu, B., Rothnagel, J. A., Langley, M. A., Tsai, S., and Roop, D. R. (1994) *J. Biol. Chem.* **269**, 7443–7449
22. Wu, R. L., Galvin, S., Wu, S. K., Xu, C., Blumenberg, M., and Sun, T.-T. (1993) *J. Cell Sci.* **105**, 303–316
23. Wu, R. L., Chen, T. T., and Sun, T. T. (1994) *J. Biol. Chem.* **269**, 28450–28459
24. Mack, D. H., and Laimins, L. A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9102–9106
25. Nakagawa, H., Wang, T., Zukerberg, L., Odze, R., Togawa, K., May, G. W. H., Wilson, J., and Rustgi, A. K. (1997) *Oncogene* **14**, 1185–1190
26. Nakagawa, H., Inomoto, T., and Rustgi, A. K. (1997) *J. Biol. Chem.* **272**, 16688–16699
27. Jenkins, T. D., Nakagawa, H., and Rustgi, A. K. (1997) *J. Biol. Chem.* **272**, 24433–24442
28. Jenkins, T. D., Opitz, O. G., Okano, J.-I., and Rustgi, A. K. (1998) *J. Biol. Chem.* **273**, 10747–10754
29. Nordeen, S. K. (1988) *BioTechniques* **6**, 454–457
30. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene (Amst.)* **77**, 51–59
31. Chung, D. C., Brand, S. J., and Tillotson, L. G. (1995) *J. Biol. Chem.* **270**, 8829–8836
32. Gu, W., Shi, X. L., and Roeder, R. G. (1997) *Nature* **387**, 819–823
33. Lill, N. L., Grossman, S. R., Ginsberg, D., DeCaprio, J., and Livingston, D. M. (1997) *Nature* **387**, 823–8271
34. Jones, P. H., Harper, S., and Watt, F. M. (1995) *Cell* **80**, 83–94
35. Eckert, R. L., Crish, J. F., and Robinson, N. A. (1997) *Physiol. Rev.* **77**, 397–424
36. Fuchs, E. (1994) *J. Cell Biol.* **125**, 511–516
37. Eckner, R., Yao, T.-P., Oldread, E., and Livingston, D. M. (1996) *Genes Dev.* **10**, 2478–2490
38. Blobel, G. A., Nakajima, T., Eckner, R., Montminy, M., and Orkin, S. H. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2061–2066
39. Bex, F., Yin, M.-J., Burny, A., and Gaynor, R. B. (1998) *Mol. Cell. Biol.* **18**, 2392–2405
40. Kawasaki, H., Eckner, R., Yao, T.-P., Taira, K., Chiu, R., Livingston, D. M., and Yokoyama, K. K. (1998) *Nature* **21**, 284–289
41. Blessing, M., Zentgraf, H., and Jorcano, J. L. (1987) *EMBO J.* **6**, 567–575
42. Blessing, M., Jorcano, J. L., and Franke, W. W. (1989) *EMBO J.* **8**, 117–126
43. Williams, T., and Tjian, R. (1991) *Genes Dev.* **5**, 670–682
44. Snape, A. M., Jonas, E. A., and Sargent, T. C. (1990) *Development* **109**, 157–165
45. Snape, A. M., Winning, R. S., and Sargent, T. C. (1991) *Development* **113**, 283–293
46. Wanner, R., Zhang, J., Dorbic, T., Mischke, D., Henz, D. M., Wittig, B., and Rosenbach, T. (1998) *Arch. Dermatol. Res.* **289**, 705–708
47. Tseng, H., and Green, H. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10311–10315
48. Tseng, H., and Green, H. (1994) *J. Cell Biol.* **126**, 495–506
49. Chen, T. T., Wu, R. L., Castro-Munozledo, F., and Sun, T. T. (1997) *Mol. Cell. Biol.* **17**, 3056–3064
50. Anderson, B., Weinberg, W. C., Rennekampff, McEvilly, R. J., Berminham, J. R., Hosshmand, F., Vasilyev, V., Hansbrough, J. F., Pittelkow, M. R., Yuspa, S. H., and Rosenfeld, M. G. (1997) *Genes Dev.* **11**, 1873–1884
51. Oettgen, P., Alani, R. M., Barcinski, M. A., Brown, L., Akbarali, Y., Boltax, J., Kunsch, C., Munger, K., and Libermann, T. A. (1997) *Mol. Cell. Biol.* **17**, 4419–4433