Volume 3 · Number 2 · 1993
GENE EXPRESSION

A cluster of five nuclear proteins regulates keratin gene transcription

Mamitaro Ohtsuki,[1] Shawna Flanagan,[1] Irwin M. Freedberg,[1] and Miroslav Blumenberg[1,2]

[1] Ronald O. Perelman Department of Dermatology and [2] Department of Biochemistry, New York University Medical Center, New York, New York

A common feature of all epithelial cells is the presence of keratin proteins that assemble into an intermediate filament cytoskeletal network. Whereas other cell types often use a specific master transcription factor to coordinate cell type-specific transcription, analysis of transcriptional regulation of keratin genes suggests that specific groupings of widely expressed transcription factors, acting on clusters of recognition elements in the promoter regions, confer epithelia-specific transcription. We define such a cluster of three sites that binds five transcription factors in the human K5 keratin gene. Within this cluster, an unusual Sp1 site binds the Sp1 transcription factor and two additional proteins. Flanking the Sp1 site are an AP2 site and another sequence, Site A; each binds a transcription factor. Similar clusters of recognition sites for the same five transcription factors have been also identified in other keratin genes. Such clusters may play a role in epithelia-specific expression of keratins.

Epithelia are distinct components of most organs in higher organism. They have a separate origin in development, exhibit polarity, and express specific protein markers. The epithelial characteristics are presumably conferred by transcription factors that are not yet defined. In muscle tissues, a relatively large and functionally redundant family of transcription factors related to MyoD is responsible for the distinct pattern of muscle gene expression. In the pituitary, one specific transcription factor, Pit 1, has been described, and in liver or brain cells, a series of transcription factors characteristic of hepatocyte or neuron nuclei is responsible for synthesis of the distinct proteins (Kuo et al., 1992). It is of considerable interest to characterize the tissue type-specific transcription factors, but so far such epithelia-specific regulators have not been found.

The proteins most commonly associated with epithelia are keratins, a large family of approximately 30 proteins that belong to either acidic or basic type and that form the intermediate filament component of the cytoskeleton of all epithelial cells (reviewed by Steinert and Roop, 1988). Indeed, the presence of keratins virtually defines epithelial cells and is used in diagnosing whether some poorly differentiated tumors are epithelia-derived carcinomas or not (Osborn and Weber, 1983). One might expect that a transcription factor common to all keratin genes may be an epithelial master switch, an "EpiD" equivalent of MyoD. However, whereas different types of muscle cells express largely overlapping subsets of muscle proteins, expression of keratin genes tends to be much more exclusive. In the epidermis, for example, the basal cells express only keratins K5 and K14.

Received May 25, 1993; revision accepted July 7, 1993.

Correspondence: Miroslav Blumenberg, Ronald O. Perelman Department of Dermatology, New York University Medical Center, 550 First Avenue, New York, NY 10016 Tel (212) 263-5924 Fax (212) 263-8752

 1052-2166/93/0302/201-13$2.00

At the onset of differentiation, these are replaced by keratins K1 and K10. Pathological conditions that involve keratinocyte activation cause induction of keratins K6 and K16 instead. In suprabasal differentiating layers, stratified epithelia other than epidermis, such as cornea, esophagus, intestine, and hair follicles, express characteristic keratin proteins, whereas simple epithelia make their own keratins (reviewed by Schweizer, 1993).

It may therefore be futile to search for a single master regulator of all keratin genes. Indeed, even comparison of the regulatory sequences upstream from the tightly co-expressed keratins, K5 and K14, failed to identify a unique common element. On the other hand, sequence comparisons suggested that the "Blessing" octamers and the AP2 sites may contribute to epithelial or epidermal specificity of transcription (Blessing et al., 1987; Leask et al., 1991).

In our own studies of transcription factors that regulate keratin genes, we were struck by the fact that within the promoter sequences the binding sites for nuclear proteins commonly occur in tight clusters (Ohtsuki et al., 1992; Bernerd et al., 1993; Magnaldo et al., 1993; Wu et al., 1993). It is conceivable that transcription distinctive for epithelial tissues may derive from assembly of characteristic sets of transcription factors, rather than from a single master regulator.

Specifically, in the promoter of the K5 keratin gene, we find a complex protein binding site that binds multiple transcription factors. We determined that five different proteins bind independently to the complex. The five include Sp1 and AP2 transcription factors. The complex is functional, as shown by point-mutagenesis and transfection assays. Interestingly, some of the same five proteins also bind to complex sites in several other keratin genes. However, while the binding proteins are the same, the sequences and the structures of the complex sites are completely different. It is as if each of these keratin gene promoters assembles a complex transcription regulatory site appropriate for the transcription factors present in epithelial cells.

Materials and methods

Oligonucleotides

All oligonucleotides were synthesized on a Pharmacia Gene-Assembler Plus. They are listed in Tables 1 and 2.

Plasmid constructions

The plasmid pK5CAT, which contains 906 bp of the promoter of the human K5 gene, and

Table 1. Synthetic oligonucleotides used in PCR.

No.	Position	Orientation	Restriction site	Purpose	Scheme	Sequence (5′→3′)
INTERNAL DELETIONS						
25	-543	FORWARD	Kpn I	COMMON		TTTGGTACCGATCACTGGGGAATCG
26	-6	REVERSE	Pst I			TTTCTGCAGCTTGTTCCTGGTGGAG
27	-201	R	Xba I	DEL ALL THREE SITES	201 166	TTTTCTAGACAGCTGTGAGCTCACC
28	-166	F	Xba I			TTTTCTAGACAGCTCGACAGCTCTC
(27)	-201	R	Xba I	DEL AP2, Sp1	201 183	
29	-183	F	Xba I			TTTTCTAGAGCCCACTTAATCATTC
30	-179	R	Xba I	DEL SITE A	179 166	TTTTCTAGATGGGCTGGGCATGCCT
(28)	-166	F	Xba I			
MUTANTS						
31	-796	F	Xba I	COMMON		TTTTCTAGAACCTGGCCTTAAAGAG
32	-6	R	Hind III			TTTAAGCTTCTTGTTCCTGGTGGAG
33	-199	R	Xho I	MUT AP2	199 CCCCCAGGC 191	TTTCTCGAGCGGCAGCTGTGAGCTCACCG
34	-191	F	Xho I		CGCTCGAGC	TTTCTCGAGCATGCCCAGCCCACTTAATC
35	-187	R	Xho I	MUT Sp1	188 GCCCAGCCCA 179	TTTCTCGAGTGCATGCCTGGGGGGCAGCT
36	-179	F	Xho I		GCACTCGAGA	TTTCTCGAGACTTAATCATTCACAGCTCG
37	-176	R	Xho I	MUT SITE A	178 CTTAATCATTCAC 166	TTTCTCGAGAAGTGGGCTGGGCATGCCTG
38	-168	F	Xho I		CTTCTCGAGGCAC	TTTCTCGAGGCACAGCTCGACAGCTCTCT
					CTCGAG : Xho I SITE	

other control plasmids, pSV2CAT and pRSVZ, have been described previously (Ohtsuki et al., 1992). The plasmid DNA was purified through Magic Megapreps DNA purification system (Promega).

The K5 complex of the three sites was cloned into an enhancerless vector (pCAT-Promoter, Promega) to determine whether the complex can act independently as an enhancer. We created two constructs containing the complex—one downstream from the CAT gene inserted into the BamH I site, the other upstream into the Bgl II site—to compare their CAT activity with that of enhancer trap-CAT alone (negative control) and of the construct containing both the enhancer and the promoter of simian virus 40 (SV40; positive control).

Internal deletions of the binding sites found in the K5 promoter were constructed by assembling two fragments framing the sequence to be deleted, using two separate rounds of polymerase chain reaction (PCR; see Table 1). In the first round, two fragments were amplified using pK5CAT as a template. The 5′ fragment was amplified using a Kpn I site-containing distal forward primer that starts at position −543, together with a Xba I site-containing reverse primer that starts immediately upstream of the designed deletion. Also in the first round, the 3′ fragment was amplified using a Xba I site-containing forward primer that starts immediately downstream of the designed deletion, and a Pst I site-containing proximal reverse primer—the most proximal one, which starts at position −6. The two amplified fragments, 5′ and 3′, were restricted with Xba I and ligated in a 10 μl reaction mixture according to a standard protocol. One μl of the ligation product was then used as a template in the second round of PCR, primed from the forward Kpn I site-containing distal primer and the Pst I site-containing proximal reverse primer. The second round of PCR amplifies across the Xba I site that replaces the deleted sequences. The

Table 2. Synthetic oligonucleotides used in gel retardation assays.

No.	Description	Binding sites (wild-type) included	Position	Sequence (5′→3′ / 3′→5′)
1	K5 COMPLEX	AP2, Sp1, SITE A	-203/-162	GGGCTGCCCCCCAGGCATGCCCAGCCCACTTAATCATTCACAGCT GACGGGGGGTCCGTACGGGTCGGGTGAATTAGTAAGTGTCGAGGG
2	K5 AP2	AP2	-207/-188	GGGACAGCTGCCCCCCAGGCATGGATC TGTCGACGGGGGGTCCGTACCTAGGGG
3	K5 AP2 CORE	AP2	-203/-188	GGGCTGCCCCCCAGGCATGGATC GACGGGGGGTCCGTACCTAGGGG
4	K5 Sp1	Sp1	-192/-177	GGGGATCGCATGCCCAGCCCACTGATC CTAGCGTACGGGTCGGGTGACTAGGGG
5	K5 SITE A	SITE A	-181/-158	GGGCCACTTAATCATTCACAGCTCGAC GGTGAATTAGTAAGTGTCGAGCTGGGG
6	K5 AP2 MUT	Sp1, SITE A	-203/-162	GGGCTGCCCTTCATGCATGCCCAGCCCACTTAATCATTCACAGCT GACGGGAAGTACGTACGGGTCGGGTGAATTAGTAAGTGTCGAGGG
7	K5 Sp1 MUT	AP2, SITE A	-203/-162	GGGCTGCCCCCCAGGCATGTTCAGACCACTTAATCATTCACAGCT GACGGGGGGTCCGTACAAGTCTGGTGAATTAGTAAGTGTCGAGGG
8	K5 SITE A MUT	AP2, Sp 1	-203/-162	GGGCTGCCCCCCAGGCATGCCCAGCCCACTTAGGCAGTCACAGCT GACGGGGGGTCCGTACGGGTCGGGTGAATCCGTCAGTGTCGAGGG
9	K5 DBL MUT	SITE A	-203/-162	GGGCTGCCCTTCATGCATGTTCAGACCACTTAATCATTCACAGCT GACGGGAAGTACGTACAAGTCTGGTGAATTAGTAAGTGTCGAGGG
10	K16 Sp1	Sp 1		GGGGAGGGCCCCGCCTTCCCCAGCTGCTATAAAGGTCT CTCCCGGGGCGGAAGGGGTCGACGATATTTCCAGAGGG
11	K16 AP2-LIKE	AP2 (?)		GGGAGCTCCTTCCCCAGCTGCTATAAAGGTC TCGAGGAAGGGGTCGACGATATTTCCAGGGG
12	K16 SITE A-LIKE	SITE A (?)		GGGAGTACCAGCCAGTTAGAGGGCAGCT TCATGGTCGGTCAATCTCCCGTCGAGGG
13	K17 AP2-LIKE	AP2 (?)		GGGAGCTGAAAGCATAGCTGGAGGGCAGCT TCGACTTTCGTATCGACCTCCCGTCGAGGG
14	K17 SITE A-LIKE	SITE A (?)		GGGCTTCTCCCATATTAGGTCATGGGAAAGCAT GAAGAGGGTATAATCCAGTACCCTTTCGTAGGG
15	Mtlla AP2	AP2		GGGAACTGACCGCCCGCGGCCCGTGTGC TTGACTGGCGGGCGCCGGGCACACGGGG
16	SV40 Sp1	Sp1		GGGTCCCGCCCCTAACTCCGCCCA AGGGCGGGGATTGAGGCGGGTGGG
17	HTLV Sp1	Sp1		GGGTCGAGGCCACGCCTCCGTCGA AGCTCCGGTGCGGAGGCAGCTGGG
18	RARE	RARE 8		AGCTTAAGGGTTCACCGAAAGTTCACTCGCAT TCGAATTCCCAAGTGGCTTTCAAGTGAGCGTA
19	TRE	TRE		TCAGGTCATGACCTGA AGTCCAGTACTGGACT
20	AP 1	AP1		CTAGTGATGAGTCAGCCGGATC GATCACTACTCAGTCGGCCTAG
21	AP 2	AP2		GATCGAACTGACCGCCCGCGGCCCGT CTAGCTTGACTGGCGGGCGCCGGGCA
22	AP 3	AP3		CTAGTGGGACTTTCCACAGATC GATCACCCTGAAAGGTGTCTAG
23	Sp1	Sp1		GATCGATCGGGGCGGGGCGATC CTAGCTAGCCCCGCCCCGCTAG
24	NF1/CTF	NF1/CTF		ATTTTGGCTTGAAGCCAATATG TAAAACCGAACTTCGGTTATAC

purified product of the second round of PCR was restricted with Kpn I and Pst I and inserted, in the correct orientation, into the similarly restricted pGCAT C plasmid (Bernerd et al., 1993).

Constructs with point mutations, as opposed to deletions of the binding sites, were created in a similar manner, except that the primers used in the first round of PCR contained an Xba I site instead of a Kpn I site for the most distal, a Hind III site instead of a Pst I site for the most proximal, and an Xho I site instead of an Xba I site for internal ligation after the first round of PCR. They also differ in size from those with internal deletions (see Fig. 9). These constructs were designed to have mutations such that several base pairs within the protein binding sites were replaced to create an Xho I site (plus one more point mutation right next to the Xho I site; see Table 1) without making a single deletion or insertion that might distort the promoter DNA or disrupt interaction with transcription factors that bind further upstream. All PCR-amplified recombinant DNA was checked by restriction and DNA sequencing.

DNA probes and competitors for gel retardation assays

Double-stranded synthetic DNA was obtained by slow annealing of complementary oligonucleotides (see Table 2), after 5 minutes of boiling in 10 mM TrisHCl, pH 7.0, 1 mM EDTA, and 150 mM NaCl. The oligonucleotides used as radioactive probes were designed to contain recessed 3′ ends. They were end-labeled with 40 μCi [α^{32}P]dCTP and the Klenow fragment of DNA polymerase (Stratagene). Radiolabeled probes were purified by gel filtration using Sephadex G50 columns (Excellulose GF5, Pierce).

Nuclear extracts

Nuclear extracts from HeLa cells and human epidermal keratinocytes (HEK) were prepared essentially according to the method of Dignam et al. (1983), with the following modification for HEK. Pelleted cells (800 × g, 5 minutes, 4°C) were resuspended in 2 volumes of cell-homogenization buffer containing 10 mM Hepes, pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, and broken in a glass Dounce homogenizer by between 20 and 30 strokes with a B pestle. The yield of intact nuclei and the efficiency of cell breakage were frequently monitored during the homogenization by microscope. The resulting nuclei were pelleted by centrifugation at 25,000 × g for 20 minutes at 4°C in an SW41 rotor. They were resuspended in 3 ml of homogenization buffer containing 20 mM Hepes, 420 mM NaCl, 225 mM EDTA, 1.5 mM $MgCl_2$, 25% glycerol, 0.5 mM PMSF, 0.5 mM DTT, 1 mM benzamidine, 1 μg/ml pepstatin A, and 0.5 μg/ml leupeptin. Nuclei were broken with 15 to 20 additional strokes with a B pestle and stirred gently for 30 minutes at 4°C. Cell debris was removed by centrifugation at 25,000 × g for 20 minutes at 4°C in an SW41 rotor. Clear supernatant was dialyzed for 5 hours at 4°C against 1 liter of buffer containing 20 mM HEPES, 100 mM KCl, 200 mM EDTA, 20% glycerol, 0.5 mM DTT, and the same concentrations of protease inhibitors as above.

Gel retardation assays

Approximately 5 μg of HeLa or HEK nuclear extract protein were first incubated for 15 minutes on ice, with or without a 100-fold molar excess of unlabeled competitor DNA, in the presence of 2 μg of poly(dI-dC) in a final volume of 25 μl. Binding reactions also contained 20 mM Tris-Cl, pH 8, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 10% glycerol, 2% polyvinyl alcohol, and 0.1 mM EDTA (final concentration). 80,000 cpm of a synthetic oligonucleotide probe was added, and incubation continued for an additional 20 minutes on ice. The probe bound to nuclear protein was then resolved from the free probe through a 5% nondenaturing polyacrylamide gel (30:0.8 acrylamide:bis-acrylamide), at 120 V for 1.5 to 2 hours. After drying, gels were autoradiographed overnight at −70°C on XAR 5 film (Kodak) with screen intensifiers.

For supershift assays, the Sp1-specific antibody (Santa Cruz Biotechnology) was added prior to the addition of a radioactive probe and incubated in the binding reaction described above for 20 minutes. Between 5 and 500 ng of antibody was used in the supershift.

Cell transfection and enzyme assays

HeLa cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS). SCC cells were maintained in DMEM supplemented with 10% FCS and 0.5 ng/ml hydrocortisone. Normal HEK from the foreskin were a gift from Dr. M. Simon. They were grown in serum-free keratinocyte medium (SFM) supplemented with epidermal

growth factor and bovine pituitary extract (Gibco). Before transfection, keratinocytes were expanded through two 1:4 passages.

HeLa cells were transfected using Chen and Okayama's procedure (1987), slightly modified according to Jiang et al. (1991). HEK cultures (80% confluent) were transfected using the polybrene-DMSO procedure as previously described (Jiang et al., 1992). Forty-eight hours after transfection, cells were washed twice with PBS and harvested by scraping. Cells were disrupted by freeze-thaw cycles, three for HeLa and SCC cells, and five for HEK. β-galactosidase and CAT assays have also been described (Jiang et al., 1992). Relative CAT activity values represent the promoter activities normalized for the efficiency of transfection.

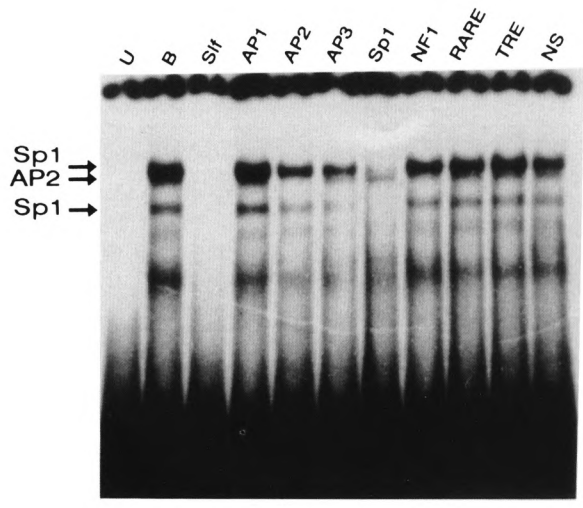

Figure 1. The oligonucleotide probe that contains the K5 complex binds a large number of nuclear proteins. The first two lanes, U and B, contain the DNA probe unbound and bound to nuclear proteins, respectively, without any competitor. Competitors (100 ng each) include the probe itself (Slf), commercially available oligonucleotides containing recognition elements for various transcription factors, and synthetic recognition elements for the retinoic acid receptor (RARE) and the thyroid hormone receptor (TRE). NS, the nonspecific competitor, contains K5 promoter site B, located further upstream from the complex site (Ohtsuki et al., 1992).

Results

The cluster of adjacent protein binding sites includes those for AP2 and Sp1 transcription factors

In our previous work (Ohtsuki et al., 1992), we found within the promoter of the K5 keratin gene a segment that is essential for the promoter function. Several proteins bind to this segment, resulting in a complex gel shift pattern with several different bands (Fig. 1). To analyze the proteins that interact with this complex sequence, we used as competitors the oligonucleotides containing consensus binding sites for transcription factors AP1, AP2, AP3, Sp1, NF1/CTF, and retinoic acid receptor, as well as oligonucleotides containing the more distal sequences from the K5 gene promoter. We found that AP2 and Sp1 oligonucleotides, but no other competitor, interfered with the binding (Fig. 1). Furthermore, whereas the AP2 oligonucleotide competitor eliminated a single band, the one presumably due to the AP2 protein, the Sp1 oligonucleotide eliminated several of the retarded bands. The band with the fastest mobility was unaffected by any oligonucleotide.

Characterization of the SP1 site

The binding competition with the Sp1 consensus oligonucleotide is surprising, because the K5 gene promoter sequence does not contain the consensus Sp1 site. Therefore we used increasing concentrations of the Sp1 consensus oligonucleotide as competitors and compared their efficiency with that of the K5 promoter Sp1 site itself. We found that the Sp1 consensus oligonucleotide is approximately as effective a competitor as the probe itself (Fig. 2). This means that protein binding to the Sp1 site of the K5 gene DNA is as strong as to the consensus Sp1 site, and therefore that the K5 promoter contains a bona fide Sp1 site. The sequence of this site is GCCCAGCCCA.

To examine the relationship of these DNA-binding proteins with the Sp1 transcription factor, we used Sp1-specific antibody in "supershift" experiments (Fig. 3). The top band contains two proteins, one of which is the Sp1 transcription factor, because this band was the only one specifically supershifted with the Sp1 protein-specific antibody. The other half of the top band and the bottom band remained unmoved. This means that three different proteins, one of them true Sp1, bind to the same site in the K5 keratin gene promoter.

The identification of the unusual Sp1 binding site in the K5 promoter, GCCCAGCCCA, prompted us to compare it to the consensus Sp1 binding site, CCCCGCC, of the K16 kera-

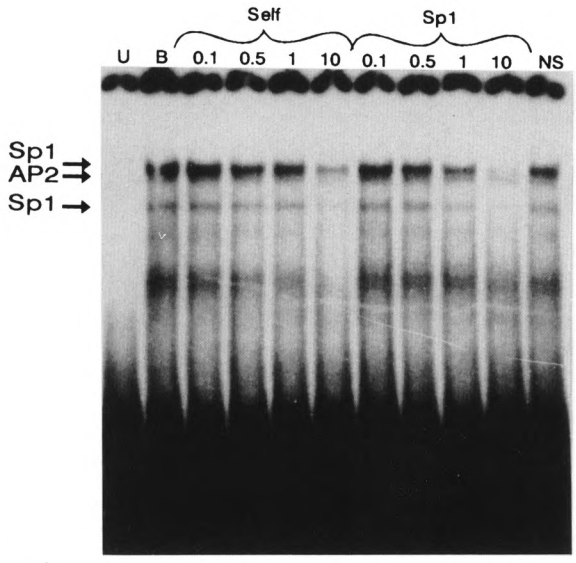

Figure 2. The consensus Sp1 binding sequence is approximately as effective as the probe itself in competition for the binding of the transcription factor. This means that the protein binds to the unusual K5 gene sequence as well as to the consensus Sp1 sequence. The amounts of competitors range from 0.1 to 10 ng; 100 ng of the same nonspecific DNA competitor as in Figure 1 were added to lane NS.

tin promoter (Magnaldo et al., 1993). Therefore we synthesized two sets of shorter oligonucleotides that contain only the Sp1 sites of the K5 or the K16 genes and used them in gel shift assays. The two sites show identical protein binding patterns: three bands at the same positions (Fig. 3). Furthermore, the K5 and K16 Sp1 sites compete with each other reciprocally. Therefore the same three proteins that bind the Sp1 site of the K5 keratin gene also bind the Sp1 site of the K16 keratin gene. This finding was the first indication that similar groupings of transcription factors may be responsible for the common features of expression of different keratin genes.

Competition with the Sp1 consensus oligonucleotide simultaneously removed three retarded bands. The three bands are consistently found in the same molar ratios in different nuclear protein preparations (not shown) and therefore most likely represent three different independent proteins, and not proteolysis artifacts. Furthermore, HeLa cells and HEK express exactly the same gel shift pattern of the three proteins (Fig. 4). A similar pattern has been observed with another promoter in another cell type (Anderson and Freytag, 1991).

The AP2 and Sp1 sites are separable

The K5 gene is somewhat different from the human metallothionein gene, where the AP2 and Sp1 sites actually overlap (Lee et al., 1987; Williams and Tjian, 1991a). In the K5 gene, their sequences are adjacent but not overlapping. To investigate the interactions at the AP2 and the Sp1 sites separately, we prepared a set of two probes containing either the AP2 or the Sp1

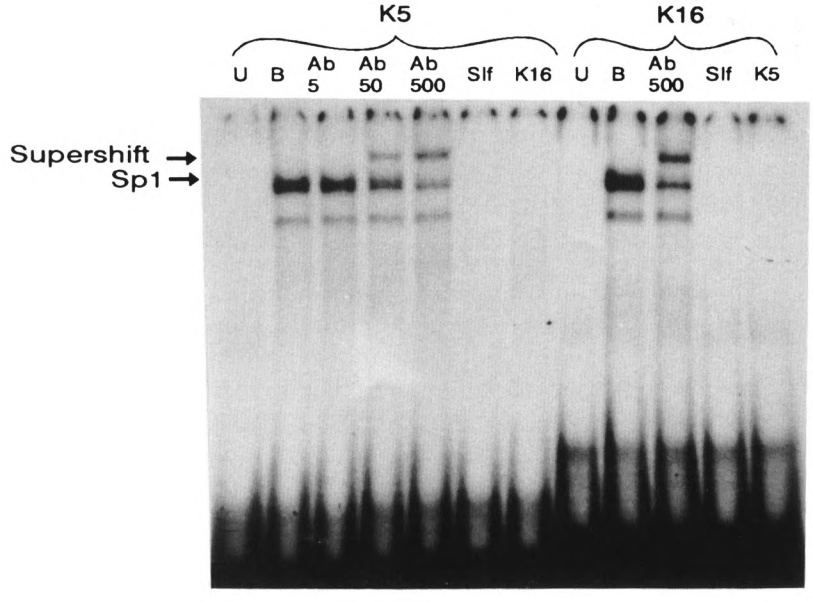

Figure 3. Supershift experiment using Sp1-specific antibodies. The probes contained the Sp1 site from the K5 (left) or K16 gene (right). The lanes marked Ab contain the antibody, 5 to 500 ng, as indicated. Beside self-competition (Slf), each of the two probes competes with the other one as well (lanes marked K16 and K5).

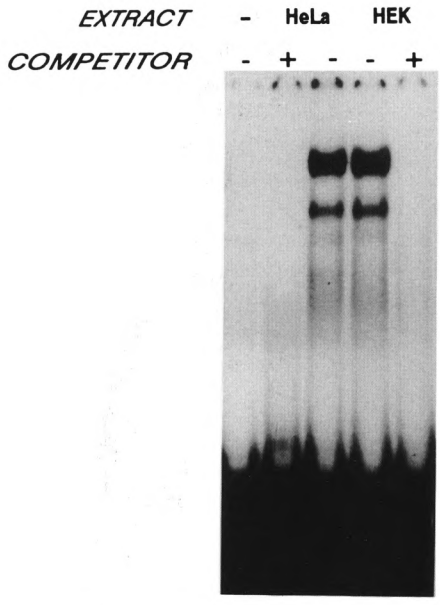

Figure 4. The same three proteins that bind the Sp1 site are found both in HeLa cells and in primary human epidermal keratinocytes (HEK). Note that similar levels, both relative and absolute, of the bands are found in both nuclear extracts.

site. Binding to the AP2 site alone results in a single band with the same mobility as that of the AP2-specific band seen with the larger probe, which contains all the sites (compare Figure 5, second track, with tracks 2 and 5 in Figure 1). Binding resulting in this band is competed with oligonucleotides containing the core of the K5 gene AP2 site or the AP2 site from the human metallothionein IIa gene, but not by the oligonucleotide containing the Sp1 site (Fig. 5).

Conversely, binding to the oligonucleotide containing the Sp1 site is competed by the oligonucleotides containing the Sp1 sites of the SV40 or human T-cell leukemia virus (HTLV) promoters, but not by a mutated Sp1 site (Fig. 5). Therefore, in the K5 promoter, AP2 and Sp1 bind independently to adjacent sites.

The binding of AP2 is specific. Oligonucleotides containing various other consensus binding sequences do not compete for binding, but those containing the characterized AP2 sites do (Fig. 6). Furthermore, a mutated AP2 site from the metallothionein IIa gene that was shown not to bind AP2 transcription factor (Williams and Tjian, 1991a), also failed to compete for binding (Fig. 6B), which confirms that the sequence in the K5 promoter is a true AP2 transcription site. We found no evidence for binding of a larger transcription factor, KTP-1, which binds to AP2-like sequences (Tamai et al., 1993). However, at this point we have not unambiguously proven, e.g., by supershifts, that the binding protein is the AP2 transcription factor itself.

Strong binding to the Sp1 site obscures binding to an additional site

The abundance and binding strength of the Sp1 protein may obscure the binding of other nuclear proteins. Therefore we synthesized a mutant oligonucleotide in which the Sp1 binding site was altered. This allowed us to observe binding of a nuclear protein to an additional site,

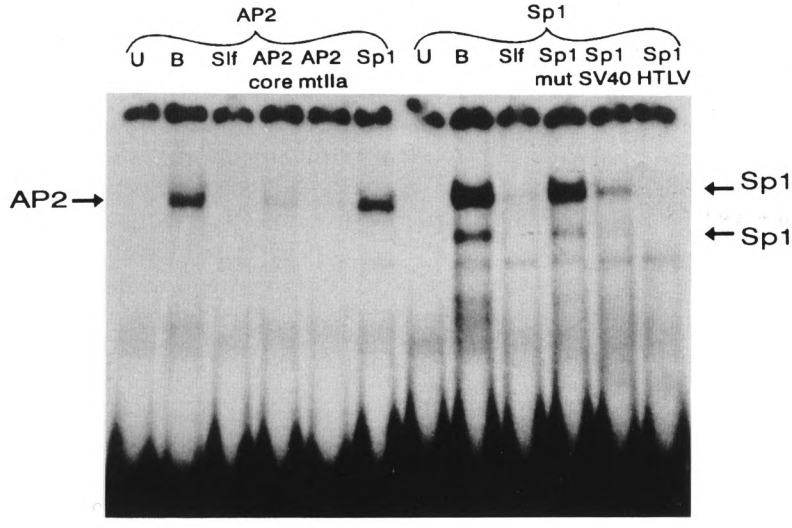

Figure 5. AP2 and Sp1 sites bind protein independently. The binding to the AP2 site, left, is abolished by the probe itself (Slf), i.e., the AP2 site of the K5 gene; by the core element of the AP2 site (core); and by the human metallothionein IIa gene promoter AP2 site (mtIIa); but not by the Sp1 consensus sequence. Conversely, binding to the Sp1 site, right, is completed by the probe (Slf), i.e., the Sp1 site of the K5 gene, and also by the Sp1 sites from the SV40 and HTLV-1 promoters, but not by the mutated Sp1 site. Lanes U and B contain unbound and protein-bound probes without competitor DNA.

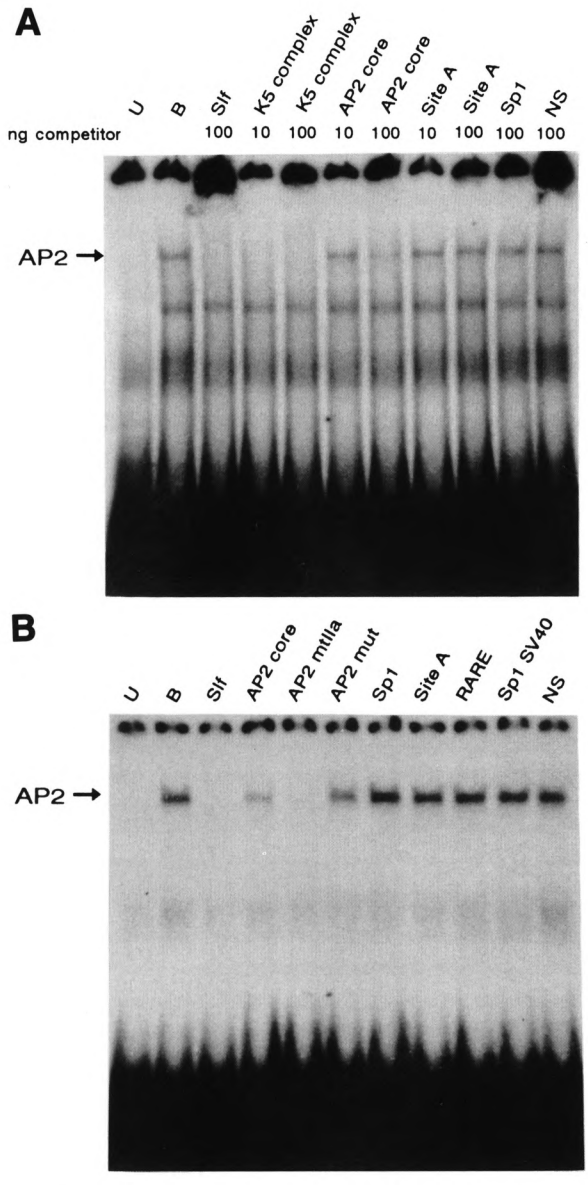

Figure 6. Characterization of the AP2 site. **A.** The competitors for binding include the probe itself (Slf), as well as the three-site complex and—less efficiently—the K5 AP2 core element. Neither Site A- nor Sp1-containing sequences competed for binding. **B.** The Mtlla gene AP2 site is as effective a competitor as the probe itself (Slf), i.e., the K5 gene AP2 site, whereas a mutated AP2 site, Sp1, Site A, RARE, and nonspecific DNA are not.

provisionally named "Site A." As shown in Figure 7, the altered sequence still binds protein at the AP2 site, which confirms that AP2 and Sp1 bind their respective sites independently (note specific competitions with the oligonucleotides containing the wild-type AP2 site). The mutant oligonucleotide no longer binds the three Sp1 site-specific proteins; instead, it reveals an additional band. The binding that produces this band is competed by both the altered and the original sequence, which means that the Site A protein binds the original sequence, i.e., that we have not accidentally created a new, irrelevant protein binding site by mutagenesis. Oligonucleotides containing the wild-type Site A compete for binding, but those containing only the AP2 or the Sp1 site do not, thus localizing Site A and confirming the specificity of the binding.

A short oligonucleotide that contains only Site A, without flanking DNA, did not bind protein (data not shown), which means that the protein needs adjacent DNA for interaction with the sequence of Site A. Therefore, we created a longer oligonucleotide with both the AP2 and the Sp1 sites altered. This oligonucleotide bound protein specifically at Site A. Oligonucleotides containing AP2, Sp1, or a mutated version of Site A did not compete, but all those containing the wild-type Site A competed for binding (Fig. 8A). This means that Site A is independent not only of the Sp1 site but also of the AP2 site. Furthermore, an oligonucleotide containing Site A and additional sequences downstream, i.e., to the side opposite to the Sp1 and AP2 sites, bound the same nuclear "protein A" (Fig. 8B).

Functional analysis of the three sites in the K5 promoter

All three sites described are essential for the function of the K5 promoter. We constructed three deletion mutants: in one Site A is missing, in another sites AP2 and Sp1, and in the third all three sites are missing. These constructs had no detectable promoter activity when tested by transfection into HeLa cells, SCC-F12 cell line derived from human squamous cell carcinoma (Rheinwald and Beckett, 1981), or primary cultures of human epidermal keratinocytes (Fig. 9).

Deletions of the sites not only removed the binding site for the corresponding transcription factor, but may also have disturbed the spacing between other protein binding sites. We prepared a set of constructs containing less disruptive point mutations in the binding sites. The point mutations abolish binding to the transcription factors, because the mutated sequences

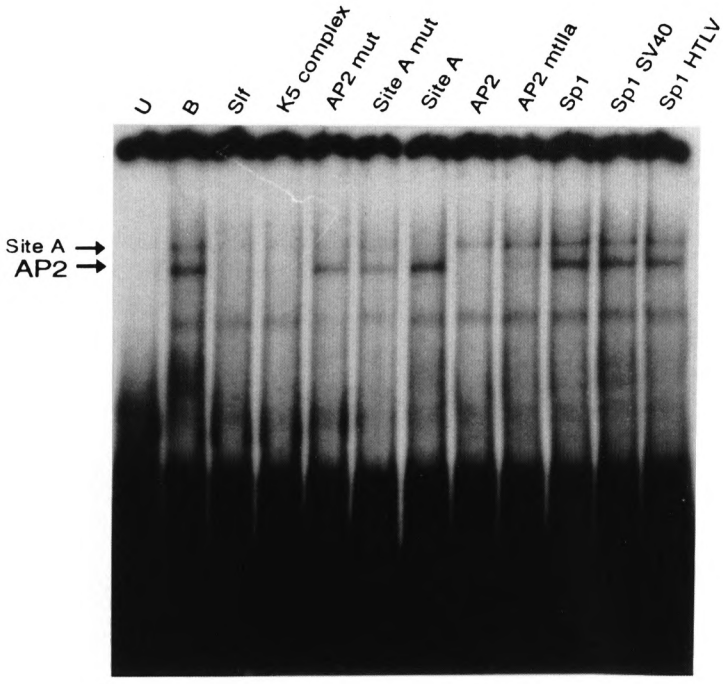

Figure 7. Binding to Sp1 obscures another element, Site A. However, with a probe containing a mutated Sp1 site, binding to Site A becomes clear. Both the mutant (Slf) and the wild-type K5 complex DNAs compete for binding. Various competitors specifically compete for binding either to Site A or to the AP2 site, whereas Sp1 site-containing oligonucleotides do not.

no longer compete for the binding in gel shift assays (not shown). Mutations in each of the three sites reduce the promoter activity (Fig. 9). Interestingly, the effects of the point mutations were much milder than the effects of deletions. We cannot exclude the possibility that the transcription factors bind to DNA in spite of the mutations, although with much weaker affinity. Alternatively, the spacing between regulatory sites in the K5 promoter DNA is very important, and mutations that preserve the correct spacing have less drastic effect on the promoter function than deletions.

The cluster of sites interacts with additional regulatory elements

The severe effects of deletions, when compared with the effects of point mutations, suggested that the complex of three sites and the proteins that bind it interact with additional regulatory elements in the K5 gene promoter. If so, the complex should not be able to act independently, i.e., outside the context of the K5 gene promoter. We tested this possibility by cloning an oligonucleotide containing the entire complex into an enhancer trap vector. Similar experiments with a regulatory site in the K6 gene showed that the K6 gene contains an independent enhancer (Blessing et al., 1989; Bernerd et al., 1993). However, the complex from the K5 gene did not function as an enhancer (Fig. 10). This is somewhat surprising, since the complex contains an Sp1 and an AP2 site, but supports the notion that these transcription factors specifically interact with other regulatory elements within the context of the K5 gene promoter.

Clusters of AP2, Sp1, and Site A sequences in the promoter regions of keratin genes

The importance of the AP2 site of the K14 keratin gene has been demonstrated before, and AP2-like elements have been found in K1, K6, and K16 gene promoters as well (Leask et al., 1991; Magnaldo et al., 1993). We therefore searched for additional AP2 sites in the promoters of keratin genes. Interestingly, in the K17 keratin gene promoter we found sequences related to the AP2 site.

To determine whether the K17 gene promoter sequence competes for binding to the AP2 site, we synthesized an oligonucleotide containing this site and used it as a competitor of the K5 gene AP2 site probe. Indeed, the K17 gene sequence competes, indicating that a common transcription factor binds the K5 and K17 keratin gene promoters (Fig. 11).

Interestingly, immediately adjacent to this site in the K17 gene promoter is a consensus

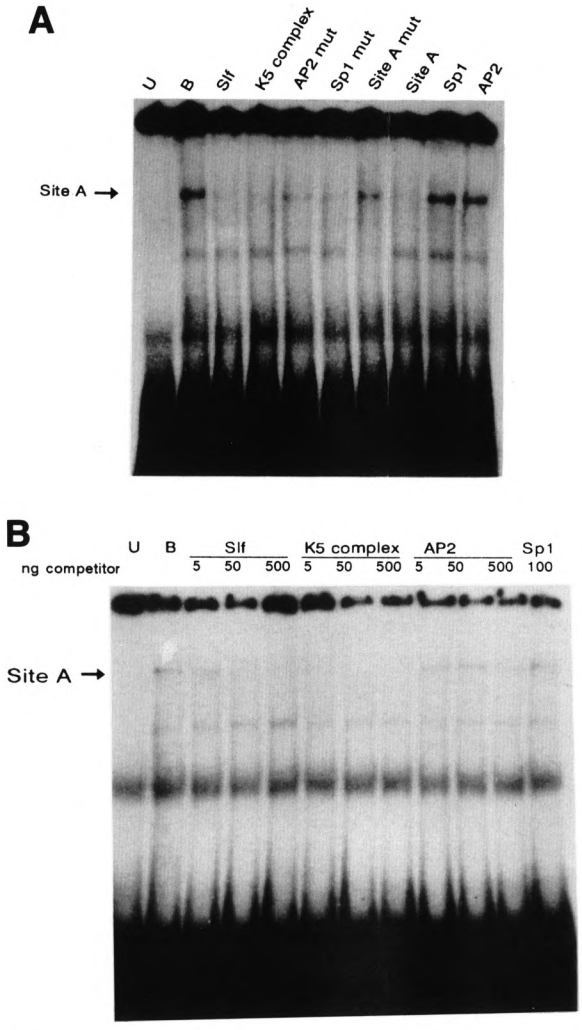

Figure 8. Binding to Site A is independent of the AP2 site. **A.** The probe contains point mutations both in the Sp1 and in the AP2 sites, leaving Site A as the only specific protein binding site. Note that only competitors containing an intact Site A compete for binding, whereas all other oligonucleotides fail to do so. **B.** The probe containing Site A and additional sequences downstream binds the same nuclear protein. Increasing amounts of the K5 complex compete for binding, but neither the AP2 nor the Sp1 site does. Oligonucleotides used as probes and competitors are listed in Tables 1 and 2.

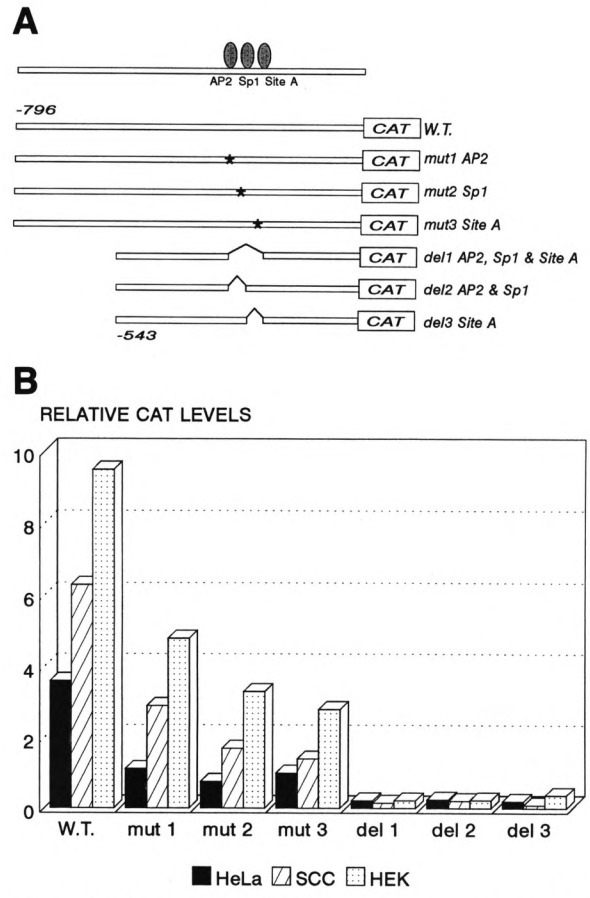

Figure 9. Each of the three sites is important for the promoter function. **A.** Structures of the DNA constructs. Asterisks indicate point mutations. **B.** The constructs were transfected into three cell types, HeLa, squamous cell carcinoma line SCC-F12, and primary cultures of human epidermal keratinocytes. Each of the point mutations reduces the transcription levels two- to threefold in all three cell types. Deletions reduce CAT levels at least tenfold, to virtually unmeasurable.

Sp1 site. Also adjacent to this site, but on the opposite side to the Sp1 site, we found a sequence resembling Site A. We synthesized an oligonucleotide containing this sequence of the K17 gene and used it as a competitor against the Site A-containing probe of the K5 gene. The K17 gene promoter sequence competed for the binding (Fig. 11).

The consensus Sp1 site of the K16 promoter described above (Fig. 3) is flanked on one side by a sequence resembling the AP2 site of the K5 gene, and on the other by a sequence resembling Site A. The former did not compete for binding to the AP2 site. However, the latter did compete for binding to the Site A of the K5 gene (Fig. 11). This result indicates that the K16 gene also contains, adjacent to its Sp1 site, a functional Site A-like sequence that binds a transcription factor in common with the K5 and K17 genes.

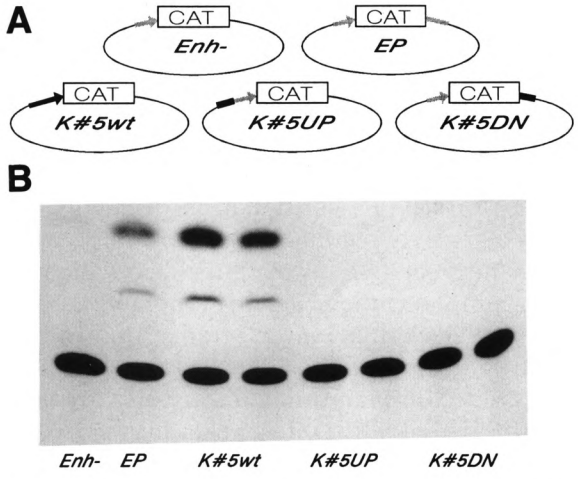

Figure 10. The three-site complex is not an enhancer. **A.** Structures of the transfected DNAs, including the control construct containing both the enhancer and the promoter of the SV40 virus (EP), the construct containing 796 bp of the K5 promoter (K5WT), the enhancer trap vector (Enh-), and the two constructs containing the three-site complex either downstream from the CAT gene inserted into the BamH I site (K5DN), or upstream into the Bgl II site (K5UP). **B.** Results of the CAT assays. Duplicate transfections of constructs containing the K5 gene DNA are shown. Note that neither the K5UP nor the K5DN construct produces CAT activity higher than the parent enhancer trap vector.

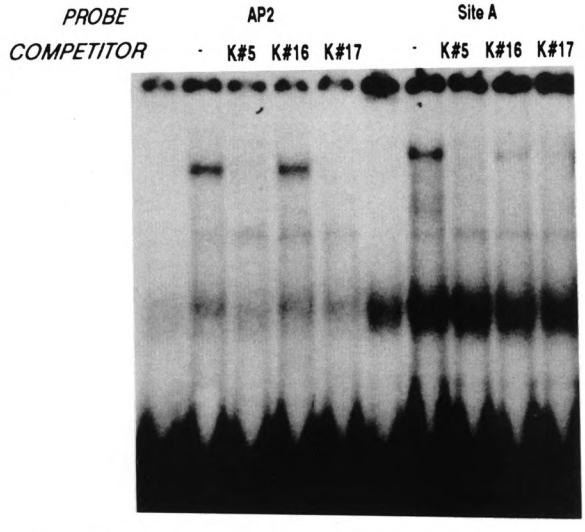

Figure 11. K16 and K17 DNA both compete for binding at Site A, but only K17 at the AP2 site. The probe used on the left is the same as in Figure 6B; the one on the right is the longer Site A probe presented in Figure 8B. In both cases, the first lane contains no nuclear protein.

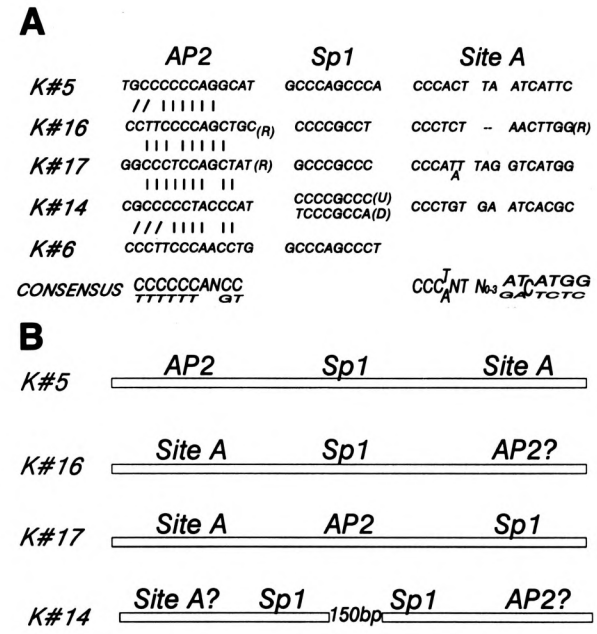

Figure 12. Common protein binding sites in promoters of keratin genes. **A.** Sequences of the sites. R marks sequences found in the reverse orientation. Vertical lines in the AP2 segments indicate identical nucleotides; slanted lines mark the additional runs of C residues. Note two different Sp1 sequences. U and D mark the upstream and the downstream consensus Sp1 sites in the K14 gene promoter. Allowing various numbers of bases within Site A gives improved alignment of sequences. **B.** Arrangement of sites within the promoters. The tightest arrangement is in the K5 gene. In the K14 gene, the two areas are separated by approximately 150 bp. Question marks indicate that these sites have been identified by sequence comparison only, and not yet by gel shift or mutagenesis experiments.

Discussion

We have identified in the promoter of the K5 keratin gene a complex protein binding sequence that binds five transcription factors. Two of these are previously identified proteins, AP2 and Sp1. The Sp1 site binds, in addition to the Sp1 protein, two more proteins. Another unidentified protein binds an adjacent sequence, Site A.

The binding sites for these transcription factors are present in other keratin gene promoters. The sequences of these sites are presented in Figure 12A. Previously, several AP2 sites have been characterized in K1, K6, K14, and K16 genes (Leask et al., 1991; Magnaldo et al., 1993). Here we add the K17 gene sequence. We

note, however, that whereas the K5 gene AP2 site closely resembles the AP2 transcription factor consensus binding sequence, the similar sites in other genes do not. K16 promoter DNA binds the bona fide AP2 protein, but at an AP2 site 300 bp distant from the one described here (Magnaldo et al., 1993). The identity of the transcription factor that binds the AP2 sequences listed in Figure 12A has not been confirmed. If it is AP2, then its specificity of binding allows considerable freedom, because the binding sequence can be found in either orientation and in either position relative to the Sp1 site. Importantly, the AP2 transcription factor has a strong dimerization domain that functions in protein–protein interactions (Williams and Tjian, 1991b).

We have identified the Sp1 sites in K5, K6, K14 (two sites), K16, and K17 gene promoters. Furthermore, in two of them, K5 and K6, an unusual Sp1 binding sequence, GCCCAGCCC, has been characterized. The Sp1 sites, unusual and consensus, bind identical sets of three nuclear proteins.

Perhaps more importantly, the K14 gene promoter contains two consensus Sp1 sites separated by approximately 150 bp. One of them is adjacent to a sequence similar to the AP2 site of K5, and the other is adjacent to a sequence similar to the Site A. With known capability of the Sp1 protein to dimerize and consequently bring distant DNA elements to close proximity (Mastrangelo et al., 1991), these two Sp1 sites may also be a part of a transcription complex containing the same five proteins that the K5 promoter complex binds. This complex is particularly important, in view of the fact that K5 and K14 keratins are co-expressed in the basal layer of stratified epithelia.

A novel protein binding sequence, Site A, has been described. This sequence in the K5 gene binds a transcription factor. The homologous sequences from the K16 and K17 genes compete for this binding. The binding to Site A is weaker than to Sp1 or AP2 sites. Nonetheless, Site A is just as important for the promoter function, because point mutations in Site A impair the promoter function as much as the mutations in the AP2 or Sp1 sites. This site can also be found in both orientations and upstream as well as downstream from the Sp1 site.

The fact that these motifs of transcription factor binding sites consistently occur in many keratin genes may explain the epithelial specificity of keratin gene expression. While many of the factors also appear in nonepithelial cells, e.g., Sp1 and AP2 (Briggs et al., 1986; Mitchell et al., 1991), the appropriate constellations may be present only in epithelia.

Why do epithelia control cell type-specific expression using a constellation of factors instead of a single regulator? Perhaps the answer lies in the great diversity of epithelial cell phenotypes. Simple epithelia create monolayers, stratified can have many layers, transitional both. Some epithelia cornify, others secrete, yet others are contractile. Hair and nails are also formed from epithelial cells. Different keratins are associated with the diversity of epithelial cell phenotypes. Unlike single master regulators, constellations of transcription factors may allow various epithelial cell types to choose which keratin genes to express. Such complex constellations may also be easier to fine-tune, and adjust the level of expression of individual keratin genes in response to extracellular stimuli such as hormones, vitamins, growth factors, matrix molecules, and pathological states.

Acknowledgments

We thank Dr. Marcia Simon for the generous gift of keratinocytes. We also thank C.-K. Jiang, J. Filipovska, and R. Vidal for help and comments.

Our research was supported by National Institutes of Health grants AR30682 and AR39176, New York University Skin Disease Research Center Grant AR39749, and the R. L. Baer Foundation. M. Blumenberg is a recipient of the Irma T. Hirschl Career Scientist Award.

The costs of publishing this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC Section 1734 solely to indicate this fact.

Mamitaro Ohtsuki is currently at the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan.

References

G. M. Anderson and S. O. Freytag (1991), Mol Cell Biol 11, 1935–1943.

F. Bernerd, T. Magnaldo, I. M. Freedberg, and M. Blumenberg (1993), Gene Expr 3, 187–199.

M. Blessing, H. Zentgraf, and J. L. Jorcano (1987), EMBO J 6, 567–575.

M. R. Briggs, J. T. Kadonaga, S. P. Bell, and R. Tjian (1986), Science 234, 47–52.

C. Chen and H. Okayama (1987), Mol Cell Biol 7, 2745–2752.

J. D. Dignam, P. L. Martin, B. S. Shastry, and R. G. G. Roeder (1983), Methods Enzymol 101, 582–598.
C.-K. Jiang, D. Connolly, and M. Blumenberg (1991a), J Invest Dermatol 97, 969–973.
C.-K. Jiang, H. S. Epstein, M. Tomic, I. M. Freedberg, and M. Blumenberg (1991b), J Invest Dermatol 96, 162–167.
C.-K. Jiang, T. Magnaldo, M. Ohtsuki, I. M. Freedberg, F. Bernerd, and M. Blumenberg (1993), Proc Natl Acad Sci USA 90, 6786–6790.
C. J. Kuo, P. B. Conley, L. Chen, F. M. Sladek, J. E. Darnell, Jr., and G. R. Crabtree (1992), Nature 355, 457–460.
A. Leask, C. Byrne, and E. Fuchs (1991), Proc Natl Acad Sci USA 88, 7948–7952.
A. Leask, M. Rosenberg, R. Vassar, and E. Fuchs (1990), Genes Dev 4, 1985–1998.
W. Lee, A. Haslinger, M. Karin, and R. Tjian (1987), Nature 325, 368–372.
T. Magnaldo, F. Bernerd, I. M. Freedberg, M. Ohtsuki, and M. Blumenberg (1993), DNA Cell Biol, in press.
I. A. Mastrangelo, A. J. Courey, J. S. Wall, S. P. Jackson, and P. V. C. Hough (1991), Biochemistry 88, 5670–5674.
P. J. Mitchell, P. M. Timmons, J. M. Hébert, P. W. J. Rigby, and R. Tjian (1991), Genes Dev 5, 105–119.
T. Ohtsuki, M. Tomic-Canic, I. M. Freedberg, and M. Blumenberg (1992), J Invest Dermatol 99, 206–215.
M. Osborn and K. Weber (1983), Lab Invest 418, 372–394.
J. G. Rheinwald and M. A. Beckett (1981), Cancer Res 41, 1657–1663.
J. Schweizer (1993), in Molecular Biology of the Skin (M. Darmon and M. Blumenberg, eds.), Academic Press, New York, pp. 46–56.
P. M. Steinert and D. R. Roop (1988), Annu Rev Biochem 57, 593–625.
K. Tamai, K. Li, and J. Uitto (1993), J Invest Dermatol 100, 498.
T. Williams and R. Tjian (1991a), Genes Dev 5, 670–682.
T. Williams and R. Tjian (1991b), Science 251, 1067–1071.
R.-L. Wu, S. Galvin, S.-K. Wu, C. Xu, M. Blumenberg, and T.-T. Sun (1993), J Cell Sci, in press.