Tissue-specific *in vivo* protein–DNA interactions at the promoter region of the *Xenopus* 63 kDa keratin gene during metamorphosis

David Warshawsky and Leo Miller*

Laboratory for Molecular Biology, Department of Biological Sciences, University of Illinois at Chicago, 845 West Taylor Street, Chicago, IL 60607, USA

Received May 30, 1995; Revised and Accepted October 6, 1995

ABSTRACT

The *Xenopus* 63 kDa keratin gene is developmentally regulated and is expressed only in the epidermis. Full activation of the 63 kDa keratin gene requires two regulatory steps, the first independent and the second dependent on the thyroid hormone triiodothyronine (T_3). Sequence analysis of a genomic clone of the 63 kDa keratin gene identified potential AP2 and SP1 binding sites upstream of the transcription initiation site. Electrophoretic mobility shift assays using purified or enriched proteins, as well as HeLa nuclear extract in conjunction with AP2- and SP1-specific antibodies, have been used to demonstrate that human AP2 and SP1 bind elements upstream of the transcription initiation site. *In vivo* footprinting with ligation mediated PCR revealed several footprints, within 350 bp upstream of the transcription initiation site, including those at the AP2 and SP1 sites, that are unique to epidermal cells which express the keratin gene. These footprints were absent in blood cells and XL177 cells which do not express the gene. Comparison of footprints between cells which express the 63 kDa keratin gene at low or high levels showed that the same binding sites are occupied, indicating that these sites are required for basal as well as T_3 -induced expression of the 63 kDa keratin gene.

INTRODUCTION

During amphibian metamorphosis, the thyroid hormone triiodothyronine (T_3) induces dramatic changes in virtually every larval organ, mediating the change from a larval program of differentiation to an adult program of differentiation. Various larval tissues are pre-programmed to respond differently to T_3 . Such pre-programming is responsible for the tissue and temporal specificity of T_3 action during metamorphosis. Our study of metamorphosis has focused on the T_3 -induced changes which regulate the expression of the 63 kDa keratin gene during the transition of the larval epidermis to the adult epidermis (reviewed in 1). The 63 kDa keratin gene is developmentally regulated and is expressed only in the epidermis. Its full activation requires two

regulatory steps, the first independent and the second dependent on T_3 . Low level expression of the 63 kDa keratin gene is first detected at stage 48 and continues for ~1 month, until stage 56. Following the appearance of T_3 in the blood and the onset of metamorphosis at stage 55, the 63 kDa keratin gene is induced to high level expression in head skin at stage 57, and its product becomes the most abundant protein in the adult epidermis. In pre-metamorphic larvae, transcription of the keratin gene is upregulated 10–20-fold by exogenous T_3 , but changes in keratin mRNA levels are not detected until 48 h after the addition of T_3 . Therefore, the 63 kDa keratin gene belongs to a group of late response genes which are upregulated during readout of the adult program of differentiation.

Using a subtractive hybridization technique, Brown and colleagues isolated two groups of genes which are regulated by T_3 in various larval tissues (2–4, reviewed in 5). The first group, the early response genes, includes genes which are regulated directly by T_3 , presumably via the binding of the hormone-receptor complex to T_3 response elements (6). The second class of genes, which are upregulated or downregulated 16–24 h after the addition of T_3 , are called delayed response genes. Several early response and delayed response genes have been identified including typical structural genes such as stromelysin 3 and collagenase, genes which are required for rapid cell growth and putative transcription factors. T_3 also upregulates within 12 h, the expression of transcription factors AP2 and CEBP in the skin and liver, respectively (7,8). Like the keratin gene, a large number of the direct response and delayed response genes are expressed at low levels prior to their upregulation by T_3 . Therefore, many of the transcription factors involved in the expression of these genes are likely to be expressed prior to the appearance of T_3 . Based on the low level expression of the 63 kDa keratin gene and other genes, prior to T_3 -dependent upregulation, we have proposed that the main action of T_3 during metamorphosis is to upregulate genes which are already expressed at low levels (1,7,9,10).

The transcription of the 63 kDa keratin gene is induced by T_3 after a 2 day lag period and therefore it serves as a good model to analyze the molecular mechanisms by which T_3 activates late response genes during metamorphosis. Since the upregulation occurs after a long period of low level expression it will be necessary to determine which transcription factors are associated with the keratin gene prior to T_3 action and then determine what

* To whom correspondence should be addressed

changes occur during T₃-induced upregulation. In this study, two approaches were taken to address these questions. An *in vitro* approach was used to identify transcription factors which bind to elements in the region of the 63 kDa keratin gene promoter and to identify their binding sites. An *in vivo* approach was used to determine if the *in vitro* characterized binding sites, as well as other binding sites, are occupied by transcription factors in epidermal cells and whether these sites are bound during low and/or high level expression of the 63 kDa keratin gene.

MATERIALS AND METHODS

Electrophoretic mobility shift assays (EMSA)

Four overlapping DNA fragments (A–D, Fig. 1A) that range in size from 191 to 251 base pairs (bp), extending from 630 bp upstream to 124 bp downstream of the 63 kDa keratin gene proximal transcription initiation site (7), were prepared by PCR using gene specific primers and *Taq* DNA polymerase (Perkin Elmer). The PCR products were run on agarose gels, the expected bands excised from the gel and the DNA fragments purified using the GeneClean II kit (Bio 101). Subfragments of B and C which contain putative AP2 and SP1 binding sites (Fig. 1A) were prepared by slow annealing of oligonucleotides (Operon) in 100 mM NaCl, after heating at 95°C for 3 min. The subfragment containing the putative AP2 binding site is 25 bp long and has the sequence: (–307) 5'-TTAAACCCCCGAAGGCTTGTTC-CA-3' (–283). The subfragment containing the putative SP1 binding sites is 40 bp long and has the sequence: (–122) 5'-GGATGGGTGCCTGCCAACTATGCCAAGCCCACTATTGCTG-3' (–83). The putative AP2 (11) and SP1 (12) binding sites are underlined, and the position of the fragments relative to the 63 kDa keratin gene transcription initiation site is indicated. Competitor DNA (purchased from Promega) has the following sequences: AP2 (5'-GATCGAACTGACCGCCCGCGGCCCGT-3'), SP1 (5'-ATTCGATCGGGCGGGCGGAGC-3'), CREB (5'-AGAGATTGCCTGACGTCAGAGAGCTAG-3'), Glucocorticoid Responsive Element (5'-TCGACTGTACAGGATGTTCTAGCTACT-3'), OCT1 (5'-TGTCGAATGCAATCACTAGAA-3') and TFIID (5'-GCAGAGCATATAAGGTGAGGTAGGA-3'). Binding reactions contained either 10 µg HeLa nuclear extract from phorbol ester-induced cells (Santa Cruz Biotechnology), 2 µg *E. coli* extract containing cloned human AP2, or 50 ng purified human SP1 protein (both purchased from Promega). The protein was preincubated on ice for 10–15 min, with or without a 100-fold molar excess of unlabeled competitor DNA, in the presence of 1–2.5 µg poly (dI–dC) in a final volume of 10–20 µl containing: 10 mM Tris–HCl, pH 7.5, 7–10% glycerol, 50 mM NaCl, 1 mM MgCl₂, 0.5 mM DTT and 0.5 mM EDTA. Approximately 0.5 ng of ³²P-end labeled fragments were then added, and incubation continued for an additional 20 min at 23°C. Antibody specific to human SP1, or human AP2 (Santa Cruz Biotechnology) was preincubated with the HeLa nuclear extract as described above for 20 min or for 1 h, respectively. Eliminating the preincubation step or changing it to 1 h had no effect on the mobility shift produced with the SP1-specific antibody. The entire reaction volumes were loaded on 4% non-denaturing polyacrylamide gels (40:1 acrylamide:bis-acrylamide) and run at 130 V for 2–3 h. Gels were dried and exposed to X-ray film (Kodak) at –70°C between intensifying screens, for 3–16 h.

Deoxyribonuclease I (DNase I) and micrococcal nuclease (MNase) digestion of permeabilized cells and free *Xenopus* genomic DNA

Epidermal cells from body skin and tail skin of stage 54/55 tadpoles, stage 55 tadpoles treated with 10^{–8} T₃ for 4 days and stage 62 tadpoles were obtained as previously described (13). Purified epidermal cells, blood cells from adult frogs and XL177 cells [a *Xenopus* epithelial cell line (14)] were washed with phosphate buffered saline three times and resuspended in 220 µl nuclease buffer (300 mM sucrose, 15 mM Tris–HCl pH 7.5, 50 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 1 mM CaCl₂, 0.5 mM 2-mercaptoethanol). DNase I (Worthington) was added to a concentration of 18, 60 or 200 U/ml. MNase (Worthington) was added to a concentration of 30, 75 or 250 U/ml. The cells were then permeabilized by adding NP-40 to a final concentration of 0.2%. After 5 min at 23°C the reaction was stopped by adding 75 µl of stop solution (2.5% SDS, 50 mM EDTA, 100 mM Tris–HCl pH 8, 50 mM NaCl, 2 mg/ml Proteinase K) and the cellular proteins digested overnight at 37°C. The nicked genomic DNA was purified by two phenol/chloroform extractions, one chloroform extraction and ethanol precipitation and resuspended in 100 µl 10 mM Tris–HCl pH 7.4, 1 mM EDTA. RNA was digested with 50 µg/ml of RNase A (Sigma) at 65°C for 20 min followed by an additional 30 min at 37°C. The DNA was repurified as described above and resuspended in H₂O.

Protein free *Xenopus* genomic DNA (30 µg in 200 µl) was digested with DNase I or MNase. DNase I was used at 2 or 5 U/ml for 5 min, at 23°C in: 40 mM Tris–HCl, pH 7.8, 10 mM NaCl, 5 mM MgCl₂. MNase was used at 20 or 50 U/ml for 5 min, at 23°C in: 15 mM Tris–HCl, pH 7.4, 60 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 1 mM CaCl₂. The reaction was stopped by adding EDTA to a concentration of 10 mM. The DNA was purified as above and resuspended in 5 mM Tris–HCl pH 7.4, 0.5 mM EDTA at a concentration of ~0.5 µg/µl. For production of guanosine-specific nicking protein-free *Xenopus* genomic DNA was treated with DMS and piperidine as previously described (15).

Ligation mediated PCR (LMPCR)

The enzymatic DNA nicking patterns were revealed by ligation mediated PCR (LMPCR) (15,16) with magnetic extension product capture, essentially as described previously (17). The following modifications were used: the enzyme used for the primer extension was Vent exo– (New England Biolabs); the primer extended DNA was desalted and ligated to the common linker as previously described (18); and 20–26 cycles of PCR were performed using the 'hot start' technique with Ampli wax PCR gems 100 (Perkin Elmer), as recommended by the manufacturer. Since small, unknown quantities of genomic DNA were used the number of cycles had to be determined empirically for each DNA preparation. The labeling reaction consisted of six PCR cycles in a final volume of 20 µl using 10% of the first PCR reaction (5 µl) with 1–2 pmol ³²P-end labeled primer C (19). The primers used to obtain the pattern shown in Figure 3 were (+124) 5'-GCTAACAGTGTGCCGCTTCTT-3' (+104), annealed at 61°C, (+109) 5'-CTTCTTGATGGAGACTTGGGAGAAA-3' (+85), annealed at 62°C and (+103) 5'-GATGGAGACTTGGGAGAAAGACATG AT-3' (+77), annealed at 68°C. The primers used to obtain the pattern shown in Figure 4 were (–298) 5'-CCGAAGGCTTGTTCAGTAAC-3' (–278), annealed at 62°C, (–284) 5'-CAGTAACAGAGTACATTGGTAGCAG-3'

(-260), annealed at 64°C and (-275) 5'-AGTACATTGGTAGC-AGGTTGAAGGAGA-3' (-249), annealed at 66°C. The primers used to obtain the pattern shown in Figure 5 were (-102) 5'-ATAGTTGGCAGGCACCCATC-3' (-121), annealed at 62°C, (-121) 5'-CCACTCAGTCTATTACATTGATCAC-3' (-145), annealed at 64°C and (-128) 5'-GTCTATTACATTGATCACACCTATGGGA-3' (-155), annealed at 66°C. The primers used to obtain the pattern shown in Figure 6 were (-443) 5'-CCAGCTGACTCTTTCAGTTGA-3' (-423), annealed at 62°C, (-422) 5'-GCTATTTTCATCATATAGCCAGCATGTC-3' (-396), annealed at 64°C and (-414) 5'-ATCATATAGCCAGCATGTCTATTACCTG-3' (-386), annealed at 66°C. For each primer set the approximate annealing temperature was first estimated using the formula $T_m = 4(G+C) + 2(A+T)$. Then, the optimal annealing temperatures were determined empirically using a downstream primer with an annealing temperature of 68°C or higher, with plasmid and genomic DNA as templates. The Mg concentration was 2 mM for all primers. Since MNase treatment creates DNA fragments without 5' phosphates, the DNA was phosphorylated *in vitro* using T4 polynucleotide kinase (New England Biolabs) following the primer extension step.

RESULTS

In vitro protein-DNA interactions

Sequence analysis of the *Xenopus* 63 kDa keratin gene and its upstream region revealed the presence of potential SP1 and AP2 sites (7). Electrophoretic mobility shift assays (EMSA) were used to determine if the regions flanking the 63 kDa keratin gene transcription initiation site contain elements which bind transcription factors SP1 and AP2. Four overlapping DNA fragments, which range in size from 191 to 251 bp, were prepared from the region extending from 630 bp upstream to 124 bp downstream of the transcription initiation site (Fig. 1A). Each of the four fragments was incubated with purified human SP1 protein, or with a bacterial extract containing cloned human AP2. A 250 bp fragment (fragment C, Fig. 1A) which contained two potential SP1 sites, one with 90% homology to the SP1 site upstream of the transcription initiation site of the human keratin gene, K5 (20), was shifted when incubated with purified SP1. A 250 bp fragment which contains a site with high homology to the consensus AP2 binding site (fragment B, Fig. 1A) had a shifted band when incubated with AP2 extract (data not shown). To narrow down the location of the binding sites for SP1 and AP2, synthetic oligonucleotides which contain the potential SP1 and AP2 sites with short flanking sequences of fragments C or B, respectively, were obtained. A 40 bp subfragment from fragment C (-83 to -122 bp), which contains the two potential SP1 sites was shifted when incubated with purified human SP1 (Fig. 1B, lane 2). This shift was eliminated when the fragment was preincubated with purified SP1 in the presence of a 100-fold molar excess of an oligonucleotide which contains a consensus SP1 binding site (Fig. 1B, lane 3), but not when preincubated with a 100-fold molar excess of an oligonucleotide which contains a consensus AP2 binding site (Fig. 1B, lane 4). A 25 bp subfragment from fragment B (-283 to -307 bp) had a shifted band when incubated with *E. coli* extract containing cloned human AP2 (Fig. 1C, lane 2). This shifted band was eliminated when the fragment was preincubated with the AP2 extract in the presence of a 100-fold molar excess of an oligonucleotide which

Figure 1. Electrophoretic mobility shift assays with human AP2 and SP1 proteins. (A) Schematic diagram of the 63 kDa keratin gene promoter region and the DNA fragments used for EMSA. The approximate positions of the putative TATA box and the sites with sequence homology to the SP1 and AP2 recognition elements are indicated. (B) A 40 bp subfragment [-83 to -122 bp, see (A)] which contains two potential SP1 binding sites (see Materials and Methods for sequence) was end-labeled and incubated in binding buffer in the absence of protein (lane 1), with purified human SP1 (lane 2), with purified SP1 in the presence of a 100-fold molar excess of an oligonucleotide which contains a consensus SP1 binding site (lane 3) and with purified SP1 in the presence of a 100-fold molar excess of an oligonucleotide which contains a consensus AP2 binding site (lane 4). (C) A 25 bp [-283 to -307 bp, see (A)] subfragment which contains a potential AP2 binding site (see Materials and Methods for sequence) was end-labeled and incubated in binding buffer in the absence of protein (lane 1), with an *E. coli* extract containing human AP2 (lane 2), with an *E. coli* extract containing human AP2 in the presence of a 100-fold molar excess of an oligonucleotide which contains a consensus AP2 binding site (lane 3) and with an *E. coli* extract containing human AP2 in the presence of a 100-fold molar excess of a non-competing oligonucleotide which can bind TFIID (lane 4).

contains a consensus AP2 binding site (Fig. 1C, lane 3), but not when preincubated with a 100-fold molar excess of an oligonucleotide which binds TFIID (Fig. 1C, lane 4). The sequences of the short fragments from the 63 kDa keratin gene promoter region, including the potential SP1 and AP2 core elements, are given in Materials and Methods.

To determine if these short fragments can also bind the SP1 and AP2 protein present in HeLa nuclear extract we used EMSA in conjunction with antibodies specific to human AP2 and human SP1. Incubation of the 25 bp fragment, which contains the putative AP2 binding site, with HeLa nuclear extract produced two shifted bands (Fig. 2A, lane 2). Both shifted bands were eliminated by preincubation of the extract with a 100-fold molar excess of the non-labeled 25 bp fragment (Fig. 2A, lane 5). However, only the top band was eliminated by cold competition

Figure 2. Electrophoretic mobility shift assays with HeLa nuclear extract. (A) A 25 bp (–283 to –307 bp, Fig. 1A) subfragment which contains a potential AP2 binding site (see Materials and Methods for sequence) was end-labeled and incubated in binding buffer in the absence of protein (lane 1), with 10 μ g of HeLa nuclear extract (lane 2), with HeLa extract in the presence of 1 μ g of an AP2-specific antibody (lane 3), with HeLa extract in the presence of a 100-fold molar excess of an oligonucleotide which contains a consensus AP2 binding site (lane 4), with HeLa extract in the presence of a 100-fold molar excess of the unlabeled 25 bp fragment (lane 5), with HeLa extract in the presence of 1 μ g of an AP2-specific antibody and a 100-fold molar excess of an oligonucleotide which contains a CREB site (lane 6), with HeLa extract in the presence of 1 μ g of an SP1-specific antibody (lane 7). (B) A 40 bp subfragment (–83 to –122 bp, Fig. 1A) which contains two potential SP1 binding sites (see Materials and Methods for sequence) was end-labeled and incubated in binding buffer in the absence of protein (lane 1), with 10 μ g of HeLa nuclear extract (lane 2), with HeLa extract in the presence of 1 μ g of an SP1-specific antibody (lane 3), with HeLa extract in the presence of 3 μ g of an SP1-specific antibody (lane 4), with HeLa extract in the presence of a 100-fold molar excess of an oligonucleotide which contains a consensus SP1 binding site (lane 5), with HeLa extract and a 100-fold molar excess of the unlabeled 40 bp fragment (lane 6), with HeLa extract in the presence of 1 μ g of an SP1-specific antibody and a 100-fold molar excess of an oligonucleotide which contains a CREB site (lane 7), with HeLa extract in the presence of 100-fold molar excess of an oligonucleotide which contains a CREB site (lane 8), with HeLa extract in the presence of 1 μ g of an AP2-specific antibody (lane 9).

with an oligonucleotide which contains the consensus AP2 binding site (Fig. 2A, lane 4). Oligonucleotides containing a CREB site or TFIID binding site did not compete out either of the two bands (data not shown). Incubation of the extract with an antibody specific to human AP2 resulted in the disappearance of

the top band and the formation of a 'supershift', while the lower band was unaffected (Fig. 2A, lane 3). An oligonucleotide containing the CREB site did not compete out the 'supershift' (Fig. 2A, lane 6), and an antibody specific to SP1 did not produce a 'supershift' (Fig. 2A, lane 7). These data suggest that at least two different proteins in the HeLa extract, one of them AP2, can bind to the region between –283 and –307.

Incubation of the 40 bp fragment, which contains the putative SP1 binding sites, with HeLa nuclear extract produced a rather large mobility shift which appears as a 'smear' with several distinct bands (Fig. 2B, lane 2, vertical bar and arrows). These mobility shifts result from specific protein–DNA interactions as they are eliminated by a 100-fold molar excess of the unlabeled 40 bp fragment (Fig. 2B, lane 6), but not by the same molar ratio of oligonucleotides which contain a CREB site (Fig. 2B, lane 8) or consensus binding sites for OCT1, TFIID, or the glucocorticoid receptor (data not shown). When preincubated with a 100-fold molar excess of an oligonucleotide which contains a consensus SP1 binding site, two bands (labeled SP1 and '2') are eliminated, but the 'smear' and the other bands remain unaffected (Fig. 2B, compare lane 5 to lanes 2, 8 and 9). When the extract was incubated with 1 or 3 μ g of the SP1 specific antibody, only the band labeled SP1 was eliminated (Fig. 2B, compare lanes 3 and 4 with lanes 2, 8 and 9). Addition of a 100-fold molar excess of an oligonucleotide which contains the CREB site had no effect on the ability of the SP1 antibody to eliminate the SP1 band (Fig. 2B, compare lane 7 with lanes 3 and 4). The AP2-specific antibody did not eliminate the SP1 band or any other band (Fig. 2B, lane 9). These results clearly indicate that SP1 binds to the 40 bp fragment. Nevertheless, the inability of the SP1-specific antibody or the oligonucleotide which contains the consensus SP1 binding site to eliminate several bands or the 'smear' also indicates that other proteins can bind the 40 bp fragment.

While the SP1-specific antibody eliminated one of the shifted bands it did not produce a 'supershift'. However, using the HeLa nuclear extract with an oligonucleotide which contains the SP1 consensus binding site, rather than the 40 bp fragment of the 63 kDa keratin gene promoter, a 'supershift' was obtained (data not shown). Interestingly, three specific shifted bands are detected with the 25 bp oligonucleotide which contains the SP1 consensus binding site, but only one is 'supershifted'. Exactly the same results were obtained by Ohtsuki *et al.* (20) with a 25 bp fragment from the promoter region of the human keratin gene, K5. These results indicate that proteins in the HeLa nuclear extract, other than SP1, can bind the SP1 recognition sequence. The multiplicity of proteins which are capable of binding the 40 bp fragment of the 63 kDa keratin gene promoter may explain why the SP1-specific antibody eliminates one of the bands but does not produce a 'supershift'. That is, the antibody–SP1 complex may not be able to compete with the other proteins for binding to the 40 bp fragment.

Experimental strategy for analysis of *in vivo* protein–DNA interactions

In order to determine if the AP2 and SP1 sites are bound *in vivo* and to identify other protein–DNA interactions, we carried out LMPCR *in vivo* footprinting in the 350 bp immediately upstream of the transcription initiation site. Two structural probes, DNase I and MNase were used. These probes detect different types of DNA–protein interactions and therefore complement each other.

DNase I binds to the minor groove of DNA and nicks the phosphate backbone (21,22). It can detect protected bases within the minor groove. DNase I also preferably nicks bases at DNA helix deformations such as bends and regions with an abnormal minor groove width (22,23). Such irregularities may be caused by binding of proteins or result from interactions between proteins which are bound at different *cis* elements (24,25). MNase binds and cleaves the phosphodiester backbone of DNA (24,26). It is used to provide data on accessibility of the backbone and can detect proteins which contact the backbone. *In vivo* nuclease nicking patterns were compared between epidermal cells which express the 63 kDa keratin gene at low levels or high levels, and non-epidermal cells in which the gene is inactive. The body skin of stage 54–55 tadpoles and the tail skin of stage 55 tadpoles provided epidermal cells which express the 63 kDa keratin gene at low basal levels. Although the tail skin is reabsorbed along with the tail during metamorphosis, it nonetheless activates the 63 kDa keratin gene to low levels (1,27). Body skin of stage 62 tadpoles and T₃-treated stage 55 tadpoles provided epidermal cells which express the 63 kDa keratin gene at high levels. Free DNA was used as a negative control. Blood cells from adult frogs and XL177 cells which do not express the 63 kDa keratin gene were examined to verify the tissue specificity of footprints. All reported footprints were highly reproducible and were present in epidermal cells from larvae of different developmental stages. The experiments indicate that the entire promoter region of the 63 kDa keratin gene is at least 10-fold more sensitive to enzymatic cleavage in epidermal cells, in comparison to XL177 or blood cells (estimated from the number of cells and PCR cycles used).

The SP1 binding region is footprinted *in vivo* in epidermal cells which express the 63 kDa keratin gene at a low or high level

DNase I footprints were identified on both strands at the putative SP1 binding sites, in epidermal cells from body skin of stage 55 tadpoles, stage 62 tadpoles and T₃-treated stage 55 tadpoles (Fig. 3, lanes 4–10, Fig. 4, lanes 4–7). The coding strand of the SP1 site contains several protected nucleotides, in the region between nucleotides –85 and –120, in all epidermal cells (Fig. 3, vertical bar, lanes 4–11). On the non-coding strand, the protected region was not as large and two regions of protection were observed, approximately between nucleotides –85 to –98 and –112 to –122 (Fig. 4, vertical bars). Two protected nucleotides were also detected at positions –105 and –110 of the non-coding strand (Fig. 4, open circles). The large size of the footprinted region, especially on the coding strand, is in agreement with the EMSA data, which showed that in addition to SP1 several unknown proteins present in HeLa nuclear extract can bind the region between nucleotides –83 and –122. In epidermal cells of tail skin from stage 55 tadpoles, the footprints were somewhat weaker in comparison to those from body skin (Fig. 3, compare lane 11 to lanes 4–10). The footprints are specific to epidermal cells as they were not only absent in free DNA (Fig. 3, lanes 1 and 2, Fig. 4, lane 2) but also in XL177 cells (Fig. 3, lane 3) and blood cells (Fig. 4, lane 3).

Figure 3. Epidermal specific *in vivo* DNase I footprints on the coding strand, in the region upstream of the transcription initiation site. Cells from the following sources were permeabilized and treated with DNase I: lane 3, XL177 cells (X); lanes 4 and 5, body skin of stage 62 tadpoles (62); lane 6, body skin of stage 54 tadpoles (54); lanes 7 and 8, body skin of stage 55 tadpoles (55); lanes 9 and 10, body skin of stage 55 tadpoles treated with T₃ for 4 days, and lane 11, tail skin of stage 55 tadpoles (T). Lanes 1 and 2, protein-free DNA (Free) was treated with DNase I *in vitro*. After purification the DNA samples were subjected to LMPCR, the labeled products were separated by electrophoresis through sequencing gels which were dried and exposed to X-ray film. When two or more lanes for each DNA source are shown, increasing concentrations of DNase I were used from left to right. The DNA source is indicated on the upper part of the figure. Numbers indicate nucleotide positions relative to the 63 kDa keratin gene transcription initiation site (arrow). The position of the putative TATA box (–27 to –34 bp) and the SP1 site (–92 to –101 bp) are indicated by the TATA and SP1 labels. Vertical bars indicate maximal regions of protection from cleavage in epidermal cells. Filled-in circles indicate the positions of nucleotides which are hypersensitive to nicking.

The AP2 site is footprinted *in vivo* in epidermal cells which express the 63 kDa keratin gene at low or high levels

A DNase I footprint was identified on the coding strand, at the site which was shown to bind AP2 *in vitro*, in epidermal cells from stage 55 tadpoles, stage 62 tadpoles, and T₃-treated stage 55 tadpoles (Fig. 5, lanes 6–10). Like the footprints at the SP1 site, the footprint at the AP2 site was tissue-specific. In protein-free DNA and DNA from XL177 cells and blood cells, several bands were present within, or close to the putative core sequence of the AP2 site which resides between nucleotides –291 and –299 (Fig. 5, lanes 2–5). These bands were absent or much weaker in DNA from epidermal cells (Fig. 5, lanes 6–11). As observed for the SP1

Figure 4. Epidermal specific *in vivo* DNase I footprints on the non-coding strand, at the SP1 binding region. Abbreviations, symbols and details are described in Figure 3 except for the abbreviation for protein-free DNA (F), the addition of blood cells (B), a guanosine sequence ladder (G) and open circles, which indicate the positions of single nucleotides which are protected in epidermal cells. The regions of protection from DNase I cleavage were observed in epidermal cells which express the 63 kDa keratin gene at low or high levels. The approximate position of the putative SP1 binding site at -92 to -101 bp relative to the transcription initiation site is shown on the right hand side (see lanes 4-8).

Figure 5. Epidermal specific *in vivo* DNase I footprints on the coding strand, in the region of the AP2 binding site. Abbreviations, symbols and details are described in Figure 3 except for the addition of blood cells (B) and a guanosine sequence ladder (G). The approximate position of the AP2 site at -291 to -299 bp relative to the transcription initiation site is shown on the right hand side.

footprint, the AP2 footprint was stronger in body epidermal cells than in tail epidermal cells (Fig. 5, compare lanes 6-10 to lane 11). The nucleotide at position -294, which is located within the putative core AP2 binding sequence, near its 3' end, was slightly hypersensitive to DNase I nicking in all epidermal cells (Fig. 5, lanes 6-11). Absence of DNase I nicking at the core of the AP2 site made it difficult to discern a footprint on the non-coding strand (data not shown). Similar results were reported for the

Figure 6. Epidermal specific *in vivo* MNase footprints on the non-coding strand, between nucleotides -225 and -260 relative to the transcription initiation site. Abbreviations, symbols and details are described in Figure 3 except for the addition of blood cells (B) and a guanosine sequence ladder (G).

mouse muscle creatine kinase enhancer by Mueller and Wold (16) who observed protection only on the coding strand of the AP2 site and a hypersensitive nucleotide within the putative core AP2 binding sequence, near its 3' end. Several additional nucleotides located downstream of the AP2 site, including the nucleotides at positions -266 and -259, were hypersensitive to DNase I nicking in all cells which express the 63 kDa keratin gene (Fig. 5, filled circles, lanes 6-11). Similar DNase I hypersensitive nucleotides have been detected between sites which bind transcription factors upstream of the *Drosophila* alcohol dehydrogenase gene (24). The authors suggested that such hypersensitive nucleotides may be due to interactions between transcription factors which result in DNA helix distortions. The hypersensitive nucleotides downstream of the AP2 site might also represent DNA bending or some other helix distortion, resulting from interactions between AP2 and a protein bound at the site of the MNase footprint between -225 and -260 (Fig. 6, discussed below).

Tissue-specific protein-DNA interactions in the region between the putative TATA box and the transcription initiation site

As seen with the SP1 and AP2 sites, the region immediately downstream of the putative TATA box was protected from DNase I nicking in all cells which express the 63 kDa keratin gene (Fig. 3, lanes 4-11). This footprint was tissue-specific since there was no protection in XL177 cells (Fig. 3, lane 3). The footprint was not centered over the putative TATA box (positioned between nucleotides -34 and -27); rather, it extended downstream from the 3' end of the TATA box with strong protection at nucleotides -19 and -21 (Fig. 3, lanes 4-11). The nucleotides at positions -4 and -6 were hypersensitive to nicking by DNase I in all the examined epidermal cells (Fig. 3, filled circles, compare lanes 4-11 to lanes 1-3). Similar hypersensitive nucleotides were detected immediately upstream of the transcription initiation site in cells which express the mouse metallothionein gene (28) and the *Drosophila* alcohol dehydrogenase gene (24). The general interpretation of these hypersensitive nucleotides is that they result from protein-induced conformational changes near the transcription start site (24).

Other protein–DNA interactions common to epidermal cells which express the 63 kDa keratin gene at low or high levels

Several footprints were identified at sites other than those which were shown by EMSA to bind SP1 and AP2. Epidermal-specific DNase I footprints were identified between nucleotides –60 and –67 and between nucleotides –69 and –78 of the coding strand (Fig. 3, lanes 4–11). The epidermal-specific hypersensitive nucleotide immediately flanking the protected region, at position –58 (Fig. 3, lanes 4–11), may also indicate the binding of a transcription factor. Several nucleotides on the non-coding strand, between positions –225 and –260, were protected from MNase nicking in epidermal cells (Fig. 6, compare lanes 2–4 to lanes 5–12). The protection pattern at this MNase footprint appears to be slightly different between stage 62 and stage 55 epidermal cells (Fig. 6, compare lanes 5–7 to lanes 8 and 9). It is possible that this difference is related to transcriptional upregulation (29); however, there is no difference in the footprint between stage 55 epidermal cells and T₃-treated stage 55 epidermal cells (Fig. 6, compare lanes 8 and 9 to lane 10).

While the identity of the factors which bind to these sites is unknown, they contain sequences which are positioned in the same 5'–3' orientation that are conserved between the *Xenopus* 63 kDa keratin gene (7) and the human 67 kDa keratin gene, K1 (30). The *Xenopus* keratin gene sequence between nucleotides –66 and –74, 5'-TCCTCTGAC-3', is identical in 8/9 nucleotides to the sequence, 5'-TCCGCTGAC-3', centered ~80 bp upstream of the K1 transcription initiation site. The MNase footprint contains two blocks which are identical to sequence blocks upstream of the K1 promoter. The first block, between nucleotides –239 and –248, 5'-TTATATTTAT-3', is identical to a 10 bp sequence ~550 bp upstream of the K1 transcription initiation site. The second sequence block, 5'-TTGAAGGA-3', located between nucleotides –251 and –258 of the *Xenopus* keratin gene is centered at virtually the same position, 250 bp upstream of the K1 transcription initiation site.

DISCUSSION

In this report a comparison was made of protein–DNA interactions, in the regions flanking the 63 kDa keratin gene transcription initiation site, in cells which transcribe the gene at different levels. The rationale was to gain insight into the molecular mechanisms by which T₃ induces the 63 kDa keratin gene to change its expression from a basal level to a high level, during metamorphosis. Using EMSA with purified SP1, AP2 and HeLa nuclear extract, human SP1 and AP2 proteins were shown to bind to putative elements upstream of the 63 kDa keratin gene transcription initiation site. In addition, it was shown that unknown proteins present in HeLa nuclear extract are able to bind in close proximity to the SP1 binding site. LMPCR *in vivo* footprinting was used to determine if the SP1 and AP2 binding sites are occupied *in vivo* and to determine if these binding sites or others change during T₃-induced upregulation of the keratin gene.

SP1, AP2 and other transcription factors which bind in the 350 bp upstream region appear to be required for both low and high level expression of the 63 kDa keratin gene

LMPCR *in vivo* footprinting detected numerous epidermal specific footprints including footprints at the putative core SP1 and AP2 sites (summarized in Fig. 7). Clear differences in footprint patterns

Figure 7. Summary of *in vivo* protein–DNA interactions at the 63 kDa keratin gene promoter region during metamorphosis. All of the protein–DNA interactions indicated were observed in cells which express the gene at low, basal levels and at high levels. The (?) symbol indicates that the identity of the transcription factor bound at the site is unknown.

were not observed among epidermal cells which express the gene at low, basal levels (epidermal cells from the body and tail of pre-metamorphic larvae), cells which express the gene at high levels (epidermal cells from stage 62 tadpoles) and cells in which the gene is induced to high level expression by T₃ treatment of pre-metamorphic larvae. All the observed footprints were specific to epidermal cells since they were not found in cells which do not express the 63 kDa keratin gene (blood cells or an epithelial cell line). Since the detected binding sites for transcription factors, including AP2 and SP1, in the 350 bp upstream of the 63 kDa keratin gene transcription initiation site do not change during T₃-induced upregulation, it is likely that they are required for basal expression, as well as high level expression.

It has been suggested that the assembly of an array of transcription factors upstream of the promoter is a coordinated, cooperative process (24). According to this model, the absence of a transcription factor, or its inability to interact with other transcription factors, or to bind its recognition site would preclude the entire array from binding. It will be interesting to determine if all the footprints upstream of the keratin gene promoter appear simultaneously. Experiments to detect the initial formation of each footprint will be carried out using cells from earlier developmental stages.

Conservation in regulatory mechanisms between the *Xenopus* 63 kDa keratin gene and mammalian keratin genes

The human keratin gene K1 and the *Xenopus* 63 kDa keratin gene are regulated in a similar manner. Both K1 and the 63 kDa keratin gene are upregulated during terminal differentiation of epidermal cells, and glucocorticoids and calcium are involved in their regulation (31–34). This study strongly suggests that SP1 and AP2, which are involved in regulating the expression of K1 (35,36), are also involved in regulating the *Xenopus* 63 kDa keratin gene. Previous work has shown that the transcriptional regulation of many mammalian keratin genes is controlled by SP1 and AP2 (12,20,37). AP2 is also involved in regulation of a *Xenopus* embryonic keratin (38). The conservation of regulatory signals controlling the 63 kDa keratin gene and human K1 is also indicated by sequence similarities in the promoter regions of the two genes. The region of the MNase footprints between nucleotides –225 and –260 and the region of the DNase I footprints between nucleotides –60 and –78 contain striking similarities to sequences upstream of the K1 promoter. Two sequence blocks between nucleotides –225 and –260 of the 63 kDa keratin gene are identical to sequences upstream of the human keratin gene, K1. One of these blocks is an 8 bp sequence which is found in virtually identical positions, ~250 bp upstream of the transcription start sites of the two genes. A sequence found

at the DNase I footprint, between -66 and -74 relative to the 63 kDa keratin gene transcription initiation site, is almost identical (8/9 bases) to a sequence centered at -80 relative to the K1 transcription initiation site. The conservation of these sequences and their relative positions in the promoters suggest that the same transcription factors are involved in regulating the expression of the two keratin genes at these sites.

Comparison of protein-DNA interactions at the promoter regions of the 63 kDa keratin gene and other inducible eucaryotic genes

The promoter regions of several inducible genes were previously analyzed by *in vivo* footprinting. The *Xenopus* liver-specific vitellogenin gene is induced by estrogen in hepatocytes of females but is not expressed in males. However, no difference in footprints was found in the 450 bp upstream of the transcription initiation site between hepatocytes of males and female frogs (39). The human IL-2 gene promoter region contains nine elements which are occupied prior to IL-2 induction (29). Following induction, three more sites become occupied and several activation specific changes in footprints become apparent in stimulated T-cells. The mouse metallothionein I gene is expressed at basal levels prior to induction by heavy metals. When metallothionein I transcription is induced, six metal responsive elements in the promoter region which previously lacked footprints become occupied (28). These examples indicate that transcriptional induction may involve a number of obvious changes, or very subtle, if any, changes in the binding sites of different promoters.

In conclusion, this study indicates that transcription factors SP1 and AP2 are required for basal as well as T₃-induced high level expression of the 63 kDa keratin gene. In addition, other sites, for unknown transcription factors were shown to be occupied during basal and T₃-induced expression. If T₃ induces new DNA-binding transcription factors which are responsible for upregulating late response genes, such as the keratin gene, they might bind to *cis*-elements which are located outside of the region screened in this study, or their binding might not be detected by the footprinting methods used. Alternatively, T₃-induction of the keratin gene does not require new DNA-binding factors. Rather, T₃-induction could be mediated by an alteration in the pre-existing array of bound transcription factors, possibly by a secondary modification or by induction of an activator protein which does not bind directly to DNA. A functional analysis will have to be completed to determine the importance of the AP2 and SP1 sites, as well as the other footprinted sites, in the regulation of 63 kDa keratin gene expression. Understanding the roles of these binding sites in keratin gene expression will provide further insight into the molecular mechanism by which T₃ upregulates late response, tissue specific genes during metamorphosis.

ACKNOWLEDGEMENTS

We thank Drs Ken Zaret and Gerald Pfeifer for their advice on LMPCR and Dr Miroslav Blumenberg for his advice on 'super-

shift' assays. This work was supported by a National Institutes of Health grant (HD 24438).

REFERENCES

- 1 Miller, L. (1995) In L.I. Gilbert, B.G. Atkinson and J.R. Tata (ed.), 'Metamorphosis': Post-embryonic Reprogramming of Gene Expression in Amphibian and Insect Cells. Academic Press, in press.
- 2 Buckbinder, L. and Brown, D. (1992) *J. Biol. Chem.*, **267**, 25 786-25 791.
- 3 Shi, Y.B., Brown, D. (1993) *J. Biol. Chem.*, **268**, 20 312-20 317.
- 4 Wang, Z. and Brown, D. (1993) *J. Biol. Chem.*, **268**, 16 270-16 278.
- 5 Shi, Y.B. (1994) *Trends Endocr. Metab.*, **5**, 14-20.
- 6 Ranjan, M., Wong, J. and Shi, Y.B. (1994) *J. Biol. Chem.*, **269**, 24699-24705.
- 7 French, R.P., Warshawsky, D., Tybor, L., Mylniczenko, N.D. and Miller, L. (1994) *Dev. Genet.*, **15**, 356-365.
- 8 Xu, Q. and Tata, J.R. (1992) *Mechanisms Dev.*, **38**, 69-81.
- 9 Mathisen, P.M. and Miller, L. (1989) *Mol. Cell. Biol.*, **9**, 1823-1831.
- 10 Mathisen, P.M. and Miller, L. (1987) *Genes Dev.*, **1**, 1107-1117.
- 11 Williams, T. and Tjian, R. (1991) *Genes Dev.*, **5**, 670-682.
- 12 Blumenberg, M. (1993) In Darmon, M. and Blumenberg, M. (ed.), *Molecular Biology of the Skin: The Keratinocyte*, Academic Press, pp. 1-32.
- 13 Nishikawa, A.K., Shimizu-Nishikawa, K. and Miller, L. (1990) *Cell. Dev. Biol.*, **26**, 1128-1134.
- 14 Ellison, T., Mathisen, P.M. and Miller, L. (1985) *Dev. Biol.*, **112**, 329-337.
- 15 Pfeifer, G.P., Stiegerwald, S.D., Mueller, P.R., Wold, B. and Riggs, A.D. (1989) *Science*, **246**, 810-813.
- 16 Mueller, P.R. and Wold, B. (1989) *Science*, **246**, 780-786.
- 17 Tormanen, V.T., Swidersky, P.M., Kaplan, B.E., Pfeifer, G.P. and Riggs, A.D. (1992) *Nucleic Acids Res.*, **20**, 5487-5488.
- 18 Warshawsky, D. and Miller, L. (1994) *BioTechniques*, **16**, 792-798.
- 19 McPherson, C.E., Shim, E.Y., Friedman, D.S. and Zaret, K.S. (1993) *Cell*, **75**, 387-398.
- 20 Ohtsuki, M., Flanagan, S., Freedberg, I.M. and Blumenberg, M. (1993) *Gene Exp.*, **3**, 201-213.
- 21 Suck, D. and Oefner, C. (1988) *Nature*, **321**, 620-625.
- 22 Hogan, M.E., Robertson, M.W. and Austin, R.H. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9273-9277.
- 23 Zhang, L. and Gralla, J.D. (1989) *Nucleic Acids Res.*, **18**, 1797-1803.
- 24 Jackson, R. J. and Benyajati, C. (1992) *Nucleic Acids Res.*, **20**, 5413-5422.
- 25 Van Der Vliet, P.C. and Verrijzer, C.P. (1993) *BioEssays*, **15**, 25-32.
- 26 Drew, H.R. (1984) *J. Mol. Biol.*, **176**, 535-557.
- 27 Nishikawa, A., Shimizu-Nishikawa, K. and Miller, L. (1992) *Dev. Biol.*, **151**, 145-153.
- 28 Mueller, P.R., Salser, S.J., and Wold, B. (1988) *Genes Dev.*, **2**, 412-427.
- 29 Brunvard, M.W., Krumm, A. and Groudine, M. (1993) *Nucleic Acids Res.*, **21**, 4824-4829.
- 30 Johnson, L.D., Idler, W.W., Zhou, X., Roop, D.R. and Steinert, P.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 896-1900.
- 31 Shimizu-Nishikawa, K. and Miller, L. (1991) *J. Exp. Zool.*, **260**, 165-169.
- 32 Shimizu-Nishikawa, K. and Miller, L. (1992) *Differentiation*, **49**, 77-83.
- 33 Roop, D.R., Huitfeldt, H., Kilkenny, A. and Yupsa, S.H. (1987) *Differentiation*, **35**, 143-150.
- 34 Yupsa, S.H., Kilkenny, A., Steinert, P.M. and Roop, D.R. (1989) *J. Cell. Biol.*, **109**, 1207-1217.
- 35 Leask, A., Rosenberg, M., Vassar, R. and Fuchs, E. (1990) *Genes Dev.*, **4**, 1985-1998.
- 36 Leask, A., Byrne, C. and Fuchs, E. (1990) *Proc. Natl. Acad. Sci. USA*, **88**, 7948-7952.
- 37 Byrne, C., Tainsky, M. and Fuchs, E. (1994) *Development*, **120**, 2369-2383.
- 38 Snape, A.M., Winning, R.S., Sargent, T.D. (1991) *Development*, **13**, 283-293.
- 39 Cardinaux, J.-R., Chapel, S. and Wahli, W. (1994) *J. Biol. Chem.*, **269**, 32 947-32 956.