Identification of a minimal promoter element of the mouse epidermal growth factor gene

John C. PASCALL* and Kenneth D. BROWN

Growth Factor Laboratory, The Babraham Institute, Cambridge CB2 4AT, U.K.

We have previously generated a transgenic mouse line (EGF/Tag) in which simian virus 40 (SV40) T-antigen expression is directed by the mouse epidermal growth factor (EGF) gene promoter. In these mice, cellular hyperproliferation is observed in the submaxillary gland associated with SV40 T-antigen expression. In addition, SV40 T-antigen-expressing tumours of prostatic origin are seen. We have now derived immortalized cell lines from these tissues and have used the cells to perform a

functional analysis of the EGF gene promoter. Cells were transfected with EGF promoter/reporter constructs, and an element located between 51 and 35 bases upstream of the EGF mRNA start site required for basal activity of the promoter was identified. Electrophoretic mobility-shift analysis suggests that three proteins bind to this region, one of which is either Sp1 or a closely related protein.

INTRODUCTION

Epidermal growth factor (EGF) is a 53-amino acid polypeptide which is mitogenic for various cell types *in vitro* and *in vivo* [1,2]. It was first purified from male mouse submaxillary gland (SMG) [3], and its expression patterns have been studied in most detail in mice. The gene is expressed at high levels at two major sites, the SMG and the kidney, as well as at low levels in many other tissues including the mammary gland, pancreas, uterus and prostate [4]. In the SMG, but not in the kidney, expression of the gene is regulated by steroids. Androgen administration to female mice increases the level of EGF mRNA [5] and protein [6], whereas castration of males results in a rapid reduction of EGF mRNA [7] and protein [6]. In contrast, oestrogens appear to have an inhibitory effect on expression of the EGF gene, as levels of EGF mRNA [7] and protein [8] increase slowly after ovariectomy. The net result of these effects is that EGF levels are 10–100-fold higher in the SMGs of male mice than the SMGs of female mice [6].

Despite this detailed information on the tissue-specific and hormonal control of EGF gene expression, little is known about the elements of the gene controlling these patterns of expression. We have begun to address this question using a transgenic mouse model. We have previously demonstrated that 5.5 kb of the 5'-flanking region of the mouse EGF gene is sufficient to direct reporter gene expression to the EGF-producing granular convoluted tubule (GCT) cells of the SMG [9]. This indicates that the DNA elements required for SMG-specific expression are located within this 5.5 kb fragment of the EGF gene promoter. In addition, in the mouse line that we derived (termed EGF/Tag), transgene expression was observed in the stomach and prostate gland. By using simian virus 40 (SV40) large T-antigen as the reporter gene, we were able to induce hyperproliferation or, in the case of the prostate gland, tumour development in the T-antigen-expressing tissues [9].

Several groups have used tissue-specific promoters to drive SV40 T-antigen expression to selected tissues in transgenic mice and have subsequently derived cell lines from the targeted tissues that retain many of the differentiated characteristics of the cells

in vivo. This approach has been used to derive differentiated cell lines from, for example, pancreatic β -cells [10], hypothalamic cells [11] and renal proximal tubule cells [12]. A more detailed analysis of the functional elements of the EGF promoter would be facilitated by the availability of cell lines in which the promoter is active. To our knowledge, however, cell lines derived from the EGF-producing cells of the mouse do not exist. We have therefore set out to establish such cell lines from the hyperproliferative/tumour tissues of our EGF/Tag mice. We report here the derivation of immortalized cell lines from the salivary and prostate glands of these mice and describe their use in transfection studies to identify putative regulatory elements within the EGF gene promoter.

EXPERIMENTAL

Derivation of cell lines from the salivary and prostate glands of EGF/Tag mice

A mouse with a palpable prostate tumour was killed and the tumour removed into ice-cold Dulbecco's modified Eagle's medium (DMEM). The tumour was chopped to approx. 2 mm³ pieces using sterile scalpel blades and then incubated in DMEM containing foetal calf serum (5%, v/v), type-I collagenase (1 mg/ml) and DNase I (10 μ g/ml) at 37 °C for 3.5 h with occasional trituration. Cells were sedimented at 400 rev./min for 10 min and resuspended in DMEM containing foetal calf serum (5% v/v) and then allowed to attach to plastic tissue-culture-grade Petri dishes (Nunc). The medium was changed after 24 h and then after a further 48 h. Cells were passaged at confluence using trypsin and were subsequently maintained in DMEM containing foetal calf serum (10%, v/v).

SMG was obtained from an 8-month-old transgenic male mouse and the gland was dissected free of other tissue. The gland was chopped to approx. 2 mm³ pieces and incubated in DMEM containing foetal calf serum (5%, v/v), testosterone (10⁻⁷ M), collagenase type II (1 mg/ml) and DNase I (10 μ g/ml) for 2 h at 37 °C with occasional trituration. The cells were recovered by

Abbreviations used: CAT, chloramphenicol acetyltransferase; DMEM, Dulbecco's modified Eagle's medium; EGF, epidermal growth factor; EMSA, electrophoretic mobility-shift analysis; GCT, granular convoluted tubule; SMG, submaxillary gland; SV40, simian virus 40.

* To whom correspondence should be addressed.

centrifugation at 400 rev./min for 10 min and resuspended in 10 ml of DMEM containing foetal calf serum (5%, v/v) and testosterone (10^{-7} M). To enrich for epithelial cells which attach more slowly to the dish than fibroblasts, cells were allowed to attach to a tissue-culture dish for 2 h at 37 °C, and unattached cells were transferred to a fresh dish and left overnight at 37 °C. Unattached cells were then transferred to a fresh dish. Cells that attached to this dish were allowed to grow to confluence and then passaged using trypsin. These cells were routinely maintained in DMEM containing foetal calf serum (10%, v/v).

Northern-blot analysis of RNA samples

RNA was isolated from mouse salivary glands or various cell lines as described previously [9]. Northern-blot analysis was performed as described [7] except that the RNA was blotted to Hybond-N membrane (Amersham International). Blots were probed using linearized total SV40 DNA or a fragment of mouse EGF cDNA spanning nucleotides 3066–3639 [13]. Probes were radiolabelled by a random-priming method [14].

Preparation of EGF promoter/chloramphenicol acetyltransferase (CAT) DNA constructs

A previously reported plasmid containing a 2.0 kb *HindIII*–*XhoI* fragment spanning the mRNA start site of the mouse EGF gene [15] was used to derive the longest EGF promoter/CAT gene construct. The plasmid was digested with *EcoRI*, and the 1200 bp fragment liberated was ligated into the *EcoRI* site of the plasmid Bluescribe M13+. This recombinant was digested with *PstI* and the insert ligated into the *PstI* site of plasmid pCAT Basic, to derive plasmid EGF5'[896]CAT. This plasmid contained 896 bp of sequence upstream of the EGF transcriptional start site, together with 180 bases of DNA encoding a part of EGF untranslated mRNA, inserted upstream of the CAT structural gene. To derive smaller constructs, a PCR strategy was used. Upstream oligonucleotides were designed with a 5'-terminal *HindIII* site and were paired with a fixed downstream oligonucleotide GGCCTGCAGGTGACTGATTTC which spans the *PstI* site in the 5'-untranslated region of the EGF gene. The oligonucleotides were used to derive fragments by PCR from a plasmid containing the cloned genomic fragment λ 6 [15]. The upstream oligonucleotides used were: CCAAGCTTACAAAT-TCCACTATA (491), CCAAGCTTGCAAAGAGGCTTGTCT (96), GCGAAGCTTGAGGTGGAGCCAGAAGAG (51) and GCGAAGCTTTAAAAGGAGAGCAG (35). The numbers in parentheses indicate the number of bases upstream from the EGF mRNA start site that is amplified using each oligonucleotide. In addition, a further amplification was performed with an upstream oligonucleotide of sequence GCGAAGCTTAAAATAGAGCCAGAAGAG (51M) in which four dG residues from oligonucleotide (51) above were replaced with dA residues. Amplified products were purified from low-melting-point agarose gels, digested with *HindIII* and *PstI*, and ligated into *HindIII*–*PstI*-digested pCAT Basic to derive the indicated plasmids. All plasmids therefore contained 180 bases of DNA encoding untranslated EGF mRNA, as well as various numbers of bases upstream of the EGF transcriptional start site. Plasmids were termed EGF5'[n]CAT, where the EGF 5'-flanking sequence begins *n* bp upstream of the transcription start site. The plasmid derived using the upstream oligonucleotide containing the dG to dA substitutions was termed EGF5'[51M]CAT. To eliminate errors due to PCR artifacts, all cloned inserts were sequenced. All plasmids were transformed into and maintained in *Escherichia coli* strain DH5, and plasmids were purified using plasmid isolation kits from Qiagen.

Transfections and reporter enzyme assays

Cell lines were grown to confluence and the medium changed on the day before transfection. On the day of transfection, cells were cleared from the dishes with trypsin, pooled and spun at 1000 rev./min for 5 min. They were then washed twice by resuspension in 20 ml of DMEM and resedimentation. Finally the cells were resuspended to a density of 3.3×10^7 – 4.4×10^7 /ml in DMEM and transfected by electroporation as described previously [16], except that the DNA added comprised 20 μ g of the appropriate EGF promoter/CAT construct and 10 μ g pRSVlac. The dishes were incubated at 37 °C for 8 h, washed twice with PBS_a (138 mM NaCl/2.8 mM KCl/8 mM Na₂HPO₄/1.5 mM KH₂PO₄) and then treated for 15 min with 400 μ l of reporter lysis buffer (Promega). Cells were scraped from the dishes, subjected to one freeze-thaw cycle and centrifuged at 8000 g (10000 rev./min) for 2 min. Supernatants were removed and stored at –70 °C until used in the enzyme assays. CAT activity was assayed as described previously [16] except that extracts were heated to 65 °C before assay and the reaction volume was increased to 130 μ l. β -Galactosidase assays were performed as follows. Cell extract (50–150 μ l) plus reporter lysis buffer to a final volume of 150 μ l was mixed with an equal volume of a solution containing 120 mM Na₂HPO₄, 80 mM NaH₂PO₄, 2 mM MgCl₂, 100 mM 2-mercaptoethanol and 4.4 mM *o*-nitrophenyl β -D-galactopyranoside. Samples were incubated for 1 h at 37 °C and then heated to 65 °C for 10 min to terminate the reaction. Water (500 μ l) was then added and the A_{420} of the solution measured.

Electrophoretic mobility-shift assay (EMSA)

Total cell extracts for EMSAs were prepared from confluent 9 cm dishes of cell lines as described previously [16]. The oligonucleotide pair for radiolabelling (oligonucleotide A) was synthesized (Babraham Institute Microchemical Facility) as one full-length oligonucleotide (5'-GCGAAGCTTGAGGTGGAG-CCAGAAGAG-3') and one of 12 bases (5'-CTCTTCTGGC-TC-3') designed to anneal to the 3' end of the full-length oligonucleotide. The oligonucleotide pair was annealed by heating equimolar amounts to 95 °C and allowing the mixture to cool slowly to room temperature. Annealed oligonucleotide (1 pmol) was made fully double-stranded in an incubation mixture containing 50 mM Tris/HCl, pH 8, 10 mM MgCl₂, 100 μ M dATP, 100 μ M dGTP, 100 μ M dTTP, 25 μ Ci of [α -³²P]dCTP (3000 Ci/mmol) and one unit of the Klenow fragment of DNA polymerase I in a final volume of 25 μ l. The reaction mixture was incubated at 37 °C for 2 h, 2.5 μ l of 1 mM dCTP was added and the reaction continued for a further 10 min. Radiolabelled double-stranded oligonucleotide was separated from non-incorporated radioactive nucleotide using NucleoTrap columns (Stratagene), and EMSAs were performed as described previously [16]. Except for the Sp1 oligonucleotide pair (oligonucleotide E), which was purchased from Promega, all competitor oligonucleotide pairs were synthesized by the Babraham Institute Microchemical Facility and were annealed by heating equimolar amounts in 50 μ l of water to 95 °C and cooling slowly to room temperature. The sequences of the competitor double-stranded oligonucleotides (upper strand only shown) were:

- (A) 5'-GCGAAGCTTGAGGTGGAGCCAGAAGAG-3'
- (B) 5'-GAGGTGGAGCCAGAAGAG-3'
- (C) 5'-AAAGAAGAGGAGGTGGAGCCAGAAGAG-3'
- (D) 5'-GCGAAGCTTAAAATAGAGCCAGAAGAG-3'
- (E) 5'-ATTCGATCGGGGCGGGGCGAGC-3'.

RESULTS

Establishment and characterization of cell lines

Cell lines were derived by continuous passage of cultures isolated from the submaxillary and prostate glands of EGF/Tag mice as described in the Experimental section. During very early passages, low levels of SV40 T-antigen were detectable by Western blotting, suggesting that, initially, few T-antigen-expressing cells established in culture. However, after three to four passages, T-antigen was easily detectable (results not shown), suggesting that T-antigen-expressing cells had a selective growth advantage over other cell types in the culture. Phase-contrast photographs of cell lines established from the prostate (termed P1) and the salivary gland (SL1) are shown in panels 1 and 2 of Figure 1(A). These cultures had an epithelioid appearance, packing closely together at confluence, and, in particular in the case of the P1 culture, had well-defined cellular boundaries. Panels 3 and 4 of Figure 1(A) show immunocytochemical localizations of T-antigen in the P1 and SL1 cell lines respectively. Nuclear T-antigen staining of essentially all of the cells in both cultures was observed.

In the SMG, we have demonstrated that, despite hyperproliferative changes associated with expression of the oncogenic SV40 T-antigen, EGF continues to be synthesized by the GCT cells [9]. Therefore it was of interest to determine whether the cell lines derived from the EGF/Tag mice tissues also expressed the endogenous EGF gene. Figure 1(B) shows

Figure 1 Characterization of the P1 and SL1 cell lines

(A) Cell lines were grown on plastic Petri dishes and were either photographed using phase-contrast microscopy (panels 1 and 2) or fixed by treatment at room temperature for 10 min with methanol/acetone and stained for SV40 T-antigen expression using the Vectastain ABC kit (Vector Laboratories, Peterborough, Cambs., U.K.) using a 1:1000 dilution of a rabbit polyclonal antiserum to SV40 T-antigen as the primary antibody (panels 3 and 4). Panels 1 and 3, P1 (prostate-derived) cell line; panels 2 and 4, SL1 (SMG-derived) cell line. The scale bar for panels 1 and 2 represents 200 μm and for panels 3 and 4, 50 μm. (B) Total RNA (10 μg) from male SMG (lane 1), NIH3T3 cells (lane 2), SL1 (lane 3) or P1 cells (lane 4) were subjected to Northern-blot analysis using a ³²P-labelled probe for SV40 T-antigen (Tag) mRNA or EGF mRNA.

Figure 2 Identification of a minimal EGF promoter element by deletion analysis

(A) Plasmids containing various lengths of the EGF promoter sequence linked to a CAT reporter gene were transfected, together with a plasmid containing a Rous sarcoma virus-driven β -galactosidase gene, into either the P1 or SL1 cell lines. After 8 h at 37 °C the cells were processed for analysis of CAT and β -galactosidase activity as described in the Experimental section. CAT activity was normalized for transfection efficiency using the β -galactosidase activity. The numbers under each column indicate the number of bases upstream of the EGF mRNA start site, with zero representing pCAT Basic alone. Note that all constructs (except pCAT Basic) also contain bases 1 to 179 of DNA encoding untranslated EGF mRNA. □, P1 cells; ▨, SL1 cells. (B) The DNA sequence of the proximal upstream region of the mouse EGF gene promoter. The putative TATA box is underlined. The start of transcription is indicated by +1. The bases -51 and -35 within the promoter are indicated.

analysis of expression of both transgene (SV40-T-antigen) mRNA and the endogenous EGF mRNA in the SL1 and P1 cell lines compared with their expression in RNA isolated from the mouse NIH3T3 cell line and from the SMG of an EGF/Tag transgenic mouse. T-antigen (Tag) mRNA was detectable in RNA isolated from the mouse SMG (lane 1) and at higher levels in the SL1 (lane 3) and P1 (lane 4) cell lines but not in RNA isolated from the mouse NIH3T3 cell line (lane 2). Since in the SMG of EGF/Tag mice, T-antigen expression occurs only in the EGF-producing GCT cells [9], the expression of T-antigen in the SL1 cell line indicates that these cells were derived from GCT cells. However, whereas high levels of EGF mRNA were detectable in the RNA isolated from the mouse SMG (lane 1), no EGF mRNA was detected in RNA from the SL1 cell line (lane 3). Similarly, EGF mRNA was not detectable by Northern-blot analysis of RNA isolated from the prostate tumour-derived P1 cell line (lane 4).

Identification of a minimal EGF promoter

To determine the minimal element of the EGF gene capable of driving reporter gene expression in the cell lines, the cells were transfected with a series of EGF-promoter/CAT gene constructs. Figure 2 shows a typical analysis of CAT activity from single transfection experiments of the P1 and SL1 cell lines. Reduction

Figure 3 EMSA of proteins binding to the -51 to -34 region of the mouse EGF gene promoter

Extracts of P1 or SL1 cells were incubated with radiolabelled double-stranded oligonucleotide A (see the Experimental section) either in the absence of unlabelled competitor or in the presence of 100-fold molar excess of double-stranded oligonucleotide competitor A (GCGAAGCTTGAGGTGGAGCCAGAAGAG), B (GAGGTGGAGCCAGAAGAG) or C (AAAGAAGAGGAGGTGGAGCCAGAAGAG). EMSA was performed as described previously [16]. The leftmost lane contains radiolabelled probe alone. The major complexes formed (X, Y and Z) are indicated.

of the promoter fragment size from 896 to 51 bases caused a small increase in the resulting CAT activity. However, across several experiments ($n = 5$ for P1 cell line; $n = 4$ for SL1 cell line) this increase was not statistically significant. Further reduction in the size of promoter fragment to 35 bases, that is to just upstream of the putative TATA box-like element (TTTAAAA) of the gene (see Figure 2B), reduced the promoter activity in both cell lines approximately 6–8-fold, to a level equivalent to or only slightly greater than that of the pCAT Basic vector (see, for example, Figure 4). Analysis of variance of several experiments indicated that this was a highly significant effect (for SL1 cell line $n = 4$ and $P = 0.006$; for P1 $n = 5$ and $P = 0.002$).

EMSA of the EGF promoter region required for basal activity

To address the issue of whether regulatory proteins bind to the region of the EGF promoter between 51 and 35 bases upstream of the transcriptional start site, EMSAs were performed using double-stranded oligonucleotide sequences. Initially the ability of proteins extracted from SL1 or P1 cells to bind to oligonucleotide A was investigated. This oligonucleotide pair spanned nucleotides 51 to 34 bases upstream of the EGF gene transcriptional start site, with an additional 5' extension unrelated to the sequence of the EGF promoter. As shown in Figure 3, when total extracts prepared from either the P1 or SL1 cell lines were incubated with radiolabelled oligonucleotide A, three major retarded species were observed, two of which migrated as a doublet (labelled X and Y in Figure 3) and one of which had a higher mobility (labelled Z). Overexposure of the gels revealed further minor bands (result not shown, but see Figure 6), although it is unclear whether these represent additional binding proteins or are degradation products of the major species. The major retarded species were efficiently competed for by incubation with a 100-fold excess of the unlabelled oligonucleotide pair A, suggesting that the retarded products observed represented high-affinity specific complexes. However, when the competition was

performed using a 100-fold excess of unlabelled double-stranded oligonucleotide spanning nucleotides -51 to -34 , oligonucleotide B (which is identical with oligonucleotide A but without the 5' extension), no competition was observed. This suggested that either proteins were interacting specifically with the DNA sequences within the 5' extension of oligonucleotide pair A or this extension was required to stabilize the end of the double-stranded oligonucleotide to facilitate protein binding. To distinguish between these two possibilities, a competition experiment was performed using a 100-fold molar excess of a double-stranded oligonucleotide spanning nucleotides -59 to -34 of the EGF promoter (oligonucleotide C). This double-stranded DNA competed effectively with the radiolabelled oligonucleotide pair A for binding proteins (Figure 3). This indicated that it was not the precise terminal DNA sequence of oligonucleotide pair A that was important for protein binding but suggested that the extension might be required to stabilize the 5' end of the double-stranded DNA to allow protein binding.

Mutational analysis to define the precise region of the EGF promoter required for basal activity

To test the suggestion that the proteins interacting with the region of the EGF promoter between 51 and 35 bases upstream of the transcriptional start site were required for basal activity of the EGF promoter, an additional EGF promoter/CAT construct containing 51 bases of EGF 5'-flanking region was prepared exactly as for EGF5'[51]CAT (see the Experimental section) except that dG residues at bases 51, 49, 48 and 46 within the

Figure 4 Effect of mutation of bases between -51 and -46 of the EGF promoter on functional activity (A) and protein binding (B)

A mutated EGF promoter construct, EGF5'[51M]CAT, was prepared in which the upstream sequence was altered from the wild-type sequence by the replacement of dG residues with dA residues at 51, 49, 48 and 46 bases upstream of the EGF transcriptional start site (see the Experimental section). (A) compares the promoter activity of the wild-type and mutated constructs when transfected into P1 cells. Cells were transfected with DNA constructs as indicated and 8 h later the cells were harvested and processed for assay of CAT activity. The mobilities of unmodified chloramphenicol (Chl) and acetylated chloramphenicol (AcChl) are indicated. (B) compares the ability of mutated and wild-type oligonucleotides to compete for protein binding with the wild-type sequence using EMSA. Extracts of P1 or SL1 cells were incubated with radiolabelled oligonucleotide A either in the absence of unlabelled competitor or in the presence of a 100-fold molar excess of either oligonucleotide A or mutated oligonucleotide D. EMSA was performed as described [16]. The sequence of the mutated oligonucleotide is 5'-GCGAAGCTTAAATAGAGCCAGAAGAG-3' (upper strand only shown with the mutated bases italicized). The leftmost lane contains radiolabelled probe only. The major complexes formed (X, Y and Z) are indicated.

Figure 5 An Sp1 consensus oligonucleotide competes for protein binding to the -51 to -34 region of the EGF gene promoter

Extracts of P1 or SL1 cells were incubated with radiolabelled oligonucleotide A either in the absence of unlabelled competitor or in the presence of a 100-fold molar excess of either oligonucleotide A or a double-stranded Sp1 consensus sequence (oligonucleotide E). The leftmost lane contains radioactive probe only. The major complexes formed (X, Y and Z) are indicated.

upstream PCR primer were altered to dA residues (Figure 4). When this plasmid, termed EGF5'[51M]CAT, was transfected into P1 cells, the resultant CAT activity was much reduced with respect to plasmid EGF5'[51]CAT to a level similar to that of plasmid EGF5'[35]CAT (Figure 4A). Similar results were obtained using SL1 cells (not shown).

To test the hypothesis that disruption of the DNA sequence between 51 and 46 bases upstream of the transcriptional start site altered/prevented binding of proteins required for activity of the EGF promoter, the ability of a double-stranded oligonucleotide D containing the dG to dA mutated bases to compete for binding of proteins to the oligonucleotide pair A was determined by EMSA. As shown in Figure 4(B), whereas the unlabelled oligonucleotide pair A competes effectively for protein binding with the radiolabelled probe A, essentially no competition was observed with the mutated oligonucleotide pair D, when using protein extracts from either the P1 or SL1 cell lines. This indicated that the DNA-binding proteins that bound to the radiolabelled probe failed to bind to the mutated DNA sequence.

Identification of an Sp1-binding site in the minimal EGF promoter

As the region of the EGF gene between 35 and 51 bases upstream of the EGF transcriptional start site contains a sequence element similar to an extended Sp1 consensus sequence [17], we tested the possibility that one of the proteins binding to this region of the EGF gene was Sp1. Initially, the ability of a double-stranded consensus Sp1 oligonucleotide sequence (oligonucleotide E) to compete for binding of proteins with radiolabelled double-stranded oligonucleotide A was investigated. As shown in Figure 5, the Sp1 double-stranded oligonucleotide was capable of competing with the radiolabelled double-stranded oligonucleotide A for binding of the proteins forming all of complexes X and Y, and most of complex Z, from both the P1 and SL1 cell lines, indicating that these proteins shared a DNA-sequence-binding specificity with Sp1.

To determine whether any of the proteins binding to the -51 to -34 region of the EGF promoter was Sp1, supershift analysis was performed using a polyclonal antiserum to Sp1. The

Figure 6 Demonstration of the binding of Sp1 to the -51 to -34 region of the EGF gene promoter by supershift analysis

Extracts of P1 or SL1 cells were incubated with radiolabelled oligonucleotide A either alone or in the presence of a polyclonal antiserum to Sp1 (I) or the corresponding preimmune serum (P). The righthand panel shows a longer exposure of the same gel as that shown in the lefthand panel. The leftmost lane in each panel contains radioactive probe only. The major complexes formed (X, Y and Z) are indicated. S indicates the supershifted complex formed after incubation with the Sp1 antiserum.

radiolabelled double-stranded oligonucleotide A was incubated with cell extracts from the P1 or SL1 cell lines in the presence or absence of the Sp1 antiserum or the corresponding preimmune serum. The results of this experiment are shown in Figure 6, which includes a short and a longer autographic exposure of the same gel. Addition of the Sp1 antiserum to the DNA-protein binding reaction reduced the formation of complex X when compared with the complexes formed either in the absence of antiserum or in the presence of the preimmune serum. This indicated that the Sp1 antiserum was disrupting the binding of Sp1 (or a cross-reacting species) to the double-stranded oligonucleotide. Overexposure of the gel shows the presence of a super-shifted species (arrowed S in Figure 6) in samples containing the Sp1 antiserum but with none of the other treatments. These results indicate that Sp1 is involved in the formation of complex X. However, they do not dismiss the possibility that additional proteins are involved in the formation of this complex. Additional experiments will be required to address this issue.

DISCUSSION

We have previously established a transgenic mouse line (termed EGF/Tag) in which a 5.5 kb fragment of the mouse EGF gene promoter directs expression of a SV40 T-antigen reporter construct. Expression of the reporter gene was directed to the salivary glands, the stomach and prostate gland in transgenic mice. Use of the oncogenic SV40 T-antigen as a reporter resulted in hyperproliferative (salivary glands and stomach) or tumorigenic (prostate gland) changes in the tissues. We have now derived cell lines from the SMG and prostate glands of these animals. The approach of directed expression of T-antigen in transgenic mice has been used previously to facilitate the establishment of immortalized cell lines from several tissues including kidney [18], hypothalamus [11], pituitary [19] and pancreas [10]. In contrast with these cell lines which retained differentiated characteristics, the cell lines we have derived from

EGF/Tag mice fail to express the endogenous EGF gene to significant levels, although reverse transcriptase PCR analysis indicates that low levels of EGF mRNA are present in the cultures (J. C. Pascall, unpublished work). In addition, we have failed to demonstrate the presence of EGF in either of the cell lines by immunocytochemistry (results not shown), using an antiserum to EGF which we have previously used to immunolocalize EGF in SMG from EGF/Tag mice [9]. This suggests that only low levels of EGF gene expression are occurring in the cell lines. We have isolated two further cell lines from the SMG of the EGF/Tag mice which also fail to synthesize significant EGF mRNA (J. C. Pascall, unpublished work), indicating that the SL1 cell line is not atypical in this respect.

The lack of EGF expression by the cell lines is surprising as we have previously shown the co-expression of EGF and Tag in the GCT cells of the SMG of EGF/Tag mice [9]. It is possible that, during establishment of the SL1 cell line, a small subpopulation of cells may have been selected that had already stopped expressing EGF *in vivo*. Alternatively, expression of the endogenous EGF gene may require cell-cell or cell-environment interactions that are disrupted when the cells are placed in culture. A further possibility is that, in establishing the cultures, cells expressing higher levels of SV40 T-antigen are selected and this results in dedifferentiation of the cell line with respect to the *in vivo* situation.

Despite the unexpected absence of EGF expression in the cell lines, the EGF promoter drives expression of T-antigen in these cells. We reasoned therefore that the cells could be used for investigating regulatory elements of the promoter. However, it is important to consider the influence of SV40 T-antigen on this system. SV40 T-antigen is not in itself a transcription factor, although it can activate expression of genes containing TATA boxes in their promoters. However, for this to occur, not only is a TATA box element required but also at least one upstream transcription-factor-binding site [20]. Therefore, although T-antigen may contribute to the activation of the EGF promoter in these cells, site-specific transcription-factor binding to the gene promoter is presumably still required for this transcription activation.

Deletion analysis of EGF promoter function in the SL1 and P1 cell lines indicates that only 51 bases of DNA upstream of the transcriptional start site was required for promoter activity. However, deletion of an additional 16 bases, that is to a point 3 bases upstream of the putative TATA box in the gene, eliminated essentially all of the promoter activity. Gel-retardation analysis demonstrated that at least three proteins bound to this region of the gene. The identity of two of these proteins remains to be determined, but the third is likely to be Sp1 (or a related protein recognized by the Sp1 antiserum). Although Sp1 binds to G/C-rich sequences, typically GGGCGGG [21], the EGF promoter contains a sequence between -51 and -41 (GAGGTGAGC) which differs at only one base (underlined) from an extended Sp1 consensus sequence G/TG/AGGCGG/TG/AG/AC/T [17]. A similar Sp1-binding site (TGGGAGGAGC) has been identified previously in the promoter of the human growth hormone gene [17].

Sp1 is known to be ubiquitous, and therefore would not be an obvious candidate for a transcription factor involved in the tissue-specific regulation of EGF expression. However, levels of Sp1 expression have been shown to vary at least 100-fold between tissues [22]. In addition, the Sp1-binding sites of different genes have at least a 10–20-fold range of binding affinities [21], resulting in genes that differ in the levels of Sp1 required for their activation. These two observations suggest that, although Sp1 itself is unlikely to be solely responsible for the tissue-specific

expression of individual genes, it could interact with other transcription factors to produce such an effect. For example, muscle-specific expression of the cardiac α -actin gene requires the interaction of three transcription factors, MyoD1, CArG box factor and Sp1 [23]. It is also interesting to consider the regulation of the uteroglobin gene. It, like EGF, is expressed in a variety of epithelial tissues. Its expression has recently also been shown to be regulated by Sp1 (and the closely related factor Sp3) via a binding site between 60 and 70 bases upstream of the mRNA transcriptional start site, a similar region to the position of the Sp1 site in the EGF promoter [24].

A recent observation suggests another possible role for the binding of Sp1 to the EGF promoter. It has been demonstrated previously that Sp1 is subjected to post-translational modification by phosphorylation [25]. It has now been shown that, in dedifferentiated tissues such as regenerating rat liver, Sp1 is less phosphorylated than in terminally differentiated hepatocytes [26]. The effect of this dephosphorylation is to increase the affinity of Sp1 for its target promoter DNA sequences and hence the activity of these promoters [26]. Since the EGF promoter binds Sp1, it raises the possibility that RNA expression from the EGF promoter might be increased in regenerating proliferating tissues. Interestingly, the promoter of the transforming growth factor α gene, which codes for another EGF receptor ligand, has also been shown to be regulated by Sp1 [27]. These workers have also demonstrated that proteins bind to the Sp1-like sites in the transforming growth factor α promoter *in vivo* [28]. Therefore it is possible that expression of the transforming growth factor α and EGF genes could be regulated in a similar manner.

The identity of the other proteins that interact with the EGF promoter in the -51 to -34 region remains to be determined. As the majority of the specific protein-DNA interactions are efficiently competed for by an Sp1 consensus oligonucleotide, but the binding of only one is disrupted by the anti-Sp1 serum, it is possible that the other proteins may be one of the other closely related members of the family (Sp2, Sp3 and/or Sp4). In particular, Sp3 and Sp4 have been shown to bind to the same consensus DNA target sequence as Sp1 [29]. Alternatively the proteins could simply be ones having an overlapping DNA-binding specificity with that of the Sp1 family, e.g. Krox24 [30].

In conclusion we have derived two cell lines which we have used to identify a minimal promoter fragment of the mouse EGF gene. Future work will involve identification of the proteins, in addition to Sp1, that bind to this region of the promoter and their roles in the regulation of expression of the EGF gene.

We thank Professor S. Jackson (Wellcome/CRC Institute, Cambridge, U.K.) for kindly providing an antiserum to transcription factor Sp1 and Dr. David Brown (Babraham Institute) for assistance with statistical analysis. This work was supported by the Biotechnology and Biological Science Research Council. J.C.P. is supported by a Babraham Research Fellowship.

REFERENCES

- 1 Carpenter, G. and Cohen, S. (1979) *Annu. Rev. Biochem.* **48**, 193–216
- 2 Carpenter, G. and Cohen, S. (1990) *J. Biol. Chem.* **265**, 7709–7712
- 3 Cohen, S. (1962) *J. Biol. Chem.* **237**, 1555–1562
- 4 Rall, L. B., Scott, J., Bell, G. I., Crawford, R. J., Penschow, J. D., Niall, H. D. and Coghlan, J. P. (1985) *Nature (London)* **313**, 228–231
- 5 Gubits, R. M., Shaw, P. A., Gresik, E. W., Onetti-Muda, A. and Barka, T. (1986) *Endocrinology* **119**, 1382–1387
- 6 Byyny, R. L., Orth, D. N. and Cohen, S. (1972) *Endocrinology* **90**, 1261–1266
- 7 Pascall, J. C., Saunders, J., Blakeley, D. M., Laurie, M. S. and Brown, K. D. (1989) *J. Endocrinol.* **121**, 501–506
- 8 Kurachi, H. and Oka, T. (1986) *J. Endocrinol.* **109**, 221–225
- 9 Pascall, J. C., Surani, M. A., Barton, S. C., Vaughan, T. J. and Brown, K. D. (1994) *J. Mol. Endocrinol.* **12**, 313–325

- 10 Efrat, S., Linde, S., Kofod, H., Spector, D., Delannoy, M., Grant, S., Hanahan, D. and Baekkeskov, S. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9037–9041
- 11 Mellon, P. L., Windle, J. J., Goldsmith, P. C., Padula, C. A., Roberts, J. L. and Weiner, R. J. (1990) *Neuron* **5**, 1–10
- 12 Cartier, N., Lacave, R., Vallet, V., Hagege, J., Hellio, R., Robine, S., Pringault, E., Cluzeaud, F., Briand, P., Kahn, A. and Vandewalle, A. (1993) *J. Cell Sci.* **104**, 695–704
- 13 Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W. J. and Bell, G. I. (1983) *Science* **221**, 236–240
- 14 Feinberg, A. P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
- 15 Pascall, J. C. and Brown, K. D. (1988) *J. Mol. Endocrinol.* **1**, 5–11
- 16 Corps, A. N., Pascall, J. C., Hadfield, K. N. and Brown, K. D. (1995) *Biochem. J.* **311**, 251–258
- 17 Lemaigre, F. P., Courtois, S. J., Lafontaine, D. A. and Rousseau, G. G. (1989) *Eur. J. Biochem.* **181**, 555–561
- 18 Sigmund, C. D., Okuyama, K., Ingelfinger, J., Jones, C. A., Mullins, J. J., Kane, C., Kim, U., Wu, C. Z., Kenny, L., Rustum, Y., Dzau, V. J. and Gross, K. W. (1990) *J. Biol. Chem.* **265**, 19916–19922
- 19 Windle, J. J., Weiner, R. I. and Mellon, P. L. (1990) *Mol. Endocrinol.* **4**, 597–603
- 20 Gruda, M. C., Zabolotny, J. M., Xiao, J. H., Davidson, I. and Alwine, J. C. (1993) *Mol. Cell. Biol.* **13**, 961–969
- 21 Kadonaga, J. T., Jones, K. A. and Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23
- 22 Saffer, J. D., Jackson, S. P. and Annarella, M. B. (1991) *Mol. Cell. Biol.* **11**, 2189–2199
- 23 Sartorelli, V., Webster, K. A. and Kedes, L. (1990) *Genes Dev.* **4**, 1811–1822
- 24 Dennig, J., Hagen, G., Beato, M. and Suske, G. (1995) *J. Biol. Chem.* **270**, 12737–12744
- 25 Jackson, S. P., MacDonald, J. J., Lees-Miller, S. P. and Tjian, R. (1990) *Cell* **63**, 155–165
- 26 Leggett, R. W., Armstrong, S. A., Barry, D. and Mueller, C. R. (1995) *J. Biol. Chem.* **270**, 25879–25884
- 27 Chen, X. R., Azizkhan, J. C. and Lee, D. C. (1992) *Oncogene* **7**, 1805–1815
- 28 Chen, X. R., Wright, K. L., Berkowitz, E. A., Azizkhan, J. C., Ting, J. P.-Y. and Lee, D. C. (1994) *Oncogene* **9**, 3179–3187
- 29 Hagen, G., Muller, S., Beato, M. and Suske, G. (1992) *Nucleic Acids Res.* **20**, 5519–5525
- 30 Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R. and Charnay, P. (1990) *Mol. Cell. Biol.* **10**, 3456–3467