Identification of a functional promoter element in the 5'-flanking region of the rat cMG1/TIS11b gene

Anthony N. CORPS, John C. PASCALL, Kathryn M. HADFIELD and Kenneth D. BROWN*

Department of Cellular Physiology, The Babraham Institute, Babraham Hall, Babraham, Cambridge CB2 4AT, U.K.

The cMG1 gene was originally identified as a mitogen-stimulated primary response gene. However, in contrast to genes such as *c-fos* and TIS11, cMG1 is also expressed at significant levels before and after the transient elevation induced by agonists. We have sequenced a 1.3 kb rat genomic cMG1 clone, which includes 931 bp upstream of the transcription start site identified by primer-extension analysis. A 1033 bp fragment, including this 5'-flanking sequence, directed the expression of the reporter gene chloramphenicol acetyltransferase (CAT) in transfected NIH-

3T3 cells. Progressive 5'-to-3' deletion indicated that an element located between -138 and -114 was responsible for most of this basal CAT expression. DNA mobility-shift assays showed that the sequence between -143 and -105 contained binding sites for cellular proteins, the principal complexes involving nucleotides between -119 and -105. We conclude that these complexes may represent the transcription factor-DNA element interactions that determine basal cMG1 expression.

INTRODUCTION

The primary response gene cMG1 (mouse TIS11b, human ERF1 [1–3]) is one of three members of a recently identified family of genes. Each gene codes for a protein containing two copies of a novel C₂H motif (CX₂CX₂CX₂H) that resembles a zinc finger [1,2], and one of the other members of the family, termed TIS11, TTP or Nup475 [4–6], has indeed been shown to bind zinc [6]. A number of proteins containing other zinc-finger motifs are established as transcription factors, and TIS11/Nup475 has been shown to be nuclear-localized [6], which is consistent with such a role. Although the function(s) of this mammalian gene family have not been defined, other proteins containing the CX₂CX₂CX₂H motif have been identified in *Drosophila*, where they are developmentally regulated [7,8], and in *Saccharomyces cerevisiae* [9]. Interestingly, over-expression of TIS11 or TIS11b in *S. cerevisiae* markedly inhibits growth [9].

TIS11, like other primary response genes such as *c-fos*, shows very low basal expression and substantial transient induction by a number of agonists (hormones and growth factors) acting on a range of cell types [4–6]. cMG1 and the third member of the family, TIS11d, show a different pattern of expression: their mRNAs are readily detected in quiescent cells before the addition of agonist and return to this basal level after an agonist-induced transient elevation [2,3,10,11]. There may therefore be separate control sequences in the cMG1 gene that mediate the basal and stimulated expression of cMG1. Although examples of control elements occurring in other positions have been reported (see the Discussion section), gene expression has most commonly been found to be regulated by the interaction of transcription factors with enhancer sequences located in the 5'-flanking region of the gene, typically within a few hundred bp of the transcription start site. This applies both to primary response genes showing transient activation [12], the most extensively studied of which are *c-fos* and *c-jun* [12–14], and to genes which display longer-term expression, for example in differentiated cellular phenotypes or under hormonal regulation [15]. In this paper, we show that a short sequence within the 5'-flanking region of the cMG1 gene

is both required for the constitutive expression of a CAT reporter gene in NIH-3T3 cells and binds specifically to proteins in cellular extracts.

EXPERIMENTAL

Materials

All oligonucleotides were synthesized at the Microchemical Facility of the Babraham Institute. [γ -³²P]ATP and [¹⁴C]chloramphenicol were obtained from ICN-Flow or Amersham International. Restriction enzymes and modifying enzymes were obtained from Gibco-BRL and Stratagene, and QIAGEN columns from Hybaid Ltd. Dulbecco's modified Eagle's medium (DMEM) and serum were from ICN-Flow and Imperial Laboratories. Antisera against c/EBP- α , - β and - δ were generously donated by Dr. S. L. McKnight (Tularik Inc., San Francisco, CA, U.S.A.). The characterization and use of these antisera has been described previously [16–18].

Isolation and sequencing of a rat genomic DNA fragment spanning the 5'-end of the cMG1 gene

A rat genomic library, consisting of a partial *Eco*RI digest of rat genomic DNA inserted into bacteriophage λ Charon 4A (a gift from Dr. H. Stanley, Institute of Zoology, Regents Park, London, U.K.), was probed using the 238 bp cDNA insert from cMG1-238 [1]. A single genomic recombinant was identified for further study. Restriction digests of this recombinant were hybridized to a fragment spanning nucleotides 24–1053 of the cMG1 cDNA sequence [1] and an approx. 8 kb *Xba*I-*Xba*I fragment was identified and subcloned. Restriction digests of this recombinant were then hybridized to a PCR fragment spanning bases 1–121 of the cMG1 cDNA sequence. A 1.3 kb *Eco*RI-*Eco*RI fragment, which spanned the 5'-end of the cMG1 gene, was identified and subcloned into plasmid pUC18 for sequencing and subsequent manipulation. Sequencing of both strands was performed on

Abbreviations used: CAT, chloramphenicol acetyltransferase; CRE, cyclic AMP response element; DMEM, Dulbecco's modified Eagle's medium; SRE, serum response element.

The DNA sequence data reported in this paper appear in the EMBL Nucleotide Sequences Database under the accession number X86571.

* To whom correspondence should be addressed.

double-stranded plasmid DNA using the Sequenase kit (United States Biochemical), as described by the manufacturer.

Primer extension analysis

A 20 bp oligonucleotide, complementary to bases 140–121 of the previously published cMG1 cDNA sequence [1], was end-labelled using [γ - 32 P]ATP and T4 polynucleotide kinase. It was then hybridized at 45 °C to poly(A)⁺ RNA isolated from stimulated RIE-1 cells and extended using reverse transcriptase in cDNA synthesis buffer (Amersham cDNA Synthesis System Plus kit) for 60 min at 42 °C. After extraction and precipitation by standard methods, the samples were denatured and subjected to electrophoresis on a 6% (w/v) acrylamide sequencing gel.

Insertion and nested deletion of cMG1 5'-flanking sequence in pCATbasic

A 1033 bp *Eco*R1–*Sau*3A genomic fragment, corresponding to 931 bp of 5'-flanking sequence and 102 bp of untranslated mRNA, was ligated into *Eco*R1–*Bam*H1-cut Bluescript. The insert was then excised using *Hind*III and *Xba*I and ligated into pCATbasic (Promega) to give the plasmid termed cMG5'[931]CAT (CAT is chloramphenicol acetyltransferase).

Progressive 5'-to-3' deletion of the inserted sequence was performed using the double-stranded Nested Deletion Kit (Pharmacia), according to the manufacturer's instructions. In brief, the cMG5'[931]CAT plasmid was cut using *Hind*III and the site was filled using α -thionucleotides, which prevented subsequent digestion by exonuclease III through this site. The plasmid was then cut at the *Eco*RV site, located between those for *Hind*III and *Eco*R1 in the multi-cloning site derived from Bluescript. Timed incubation with exonuclease III was then used to produce progressive deletions of the cMG1 5'-flanking sequence, and single-stranded ends were removed using S1 nuclease. The plasmids were re-ligated and used to transform TG1 bacteria. Plasmids from selected colonies were purified on QIAGEN columns; the position of the deletion in each plasmid was determined by sequencing, using the M13 reverse primer site 21 bp upstream of the *Hind*III site into which the cMG1 5'-flanking sequence had been inserted. The deleted plasmids are termed cMG5'[*n*]CAT, where the cMG1 5'-flanking sequence remaining after deletion begins *n* bp upstream from the transcription start site.

Cell culture, transfection and analysis of CAT expression

Stock cultures of Swiss 3T3 cells and RIE-1 cells were maintained and passaged in DMEM containing 10% (v/v) and 5% (v/v) newborn calf serum respectively, as described previously [19,20]. NIH-3T3 cells were treated in the same manner as Swiss 3T3 cells, and PC12 cells were grown in DMEM containing 10% (v/v) fetal calf serum.

For transfection, NIH-3T3 cells were seeded into 35-mm dishes at 7.5×10^4 cells/dish. After 24 h they were transfected with the cMG5'[*n*]CAT plasmids (0.5–2 μ g/dish), using the DOTAP transfection reagent (Boehringer) prepared according to the manufacturer's instructions. The cells were incubated with DOTAP/plasmid overnight in DMEM containing 10% (v/v) calf serum, and then for a further 24 or 48 h in fresh medium. The dishes were then placed on ice, rinsed three times with ice-cold PBS, scraped into 0.5 ml of PBS and transferred into Microfuge tubes. After a further rinse with PBS, the cells were resuspended in 40 μ l of 250 mM Tris, pH 7.8, and subjected to three cycles of freezing (solid CO₂/ethanol mixture) and thawing (37 °C for 2 min), with vigorous vortexing between cycles. The

disrupted cells were centrifuged (4 min, 10000 *g*) and the supernatants were incubated at 60 °C for 10 min and then re-centrifuged. A fixed proportion of the supernatant was incubated for 3 h at 37 °C in the presence of acetyl-CoA (2 mg/ml) and [14 C]chloramphenicol (50 nCi) in a final volume of 70 μ l of 200 mM Tris, pH 7.8. The reactions were extracted using ethyl acetate and the acetylated chloramphenicol was separated from the unreacted material by TLC on silica-gel plates (Sigma, T-6770) using chloroform:methanol (19:1). The radioactive spots visualized by autoradiography were cut out and counted in a scintillation counter.

PC12 cells were not reliably transfected by this method and were therefore transfected by electroporation. Between $(1-2) \times 10^7$ cells in DMEM were incubated for 5 min at room temperature with 10 μ g of test plasmid and 10 μ g of the control plasmid RSV-lac, placed in an electroporation cuvette and treated at 300 V, 960 μ F (time constants between 20 and 21 ms) in a Gene Pulser (Bio-Rad). The treated cells were incubated on ice for 8 min, dispersed in 10 ml of DMEM/serum, divided between two 60-mm dishes and incubated for 8 h at 37 °C. After this incubation, the cells were rinsed twice with PBS, treated with 400 μ l of reporter lysis buffer (Promega) for 15 min at room temperature and subjected to one cycle of freeze/thawing as described above. Aliquots of the supernatant were assayed for CAT activity as described above and for β -galactosidase activity as described previously [21].

DNA mobility-shift assay

For the preparation of extracts for band-shift assays, cells were grown on 9-cm dishes for 6–8 days (3T3 and PC12) or 10–14 days (RIE-1). Whole-cell and nuclear extracts for mobility-shift assays were prepared using methods based on those described previously [22,23]. For whole-cell extracts, cells were digested with trypsin in a minimum volume to detach them from the dish, and washed twice by centrifugation and resuspension in 10 ml of ice-cold PBS. They were then centrifuged again and resuspended in 70–200 ml (depending on the number of cells/dish) of extraction buffer (450 mM NaCl/0.4 mM EDTA/25% (v/v) glycerol/0.5 mM dithiothreitol/0.5 mM PMSF/20 mM Hepes, pH 7.8) and subjected to three cycles of freeze/thawing as described above. Finally, the disrupted cells were centrifuged at 10000 *g* at 4 °C for 2 min and aliquots of the supernatant were frozen and stored at –70 °C until required. For the preparation of nuclear extracts, dishes of cells were placed on ice and the cells were rinsed three times with ice-cold PBS and once with Buffer A (1.5 mM MgCl₂/10 mM KCl/10 mM Hepes, pH 7.9). They were then lysed in the dish in 1.5 ml of Buffer A containing 0.2% (v/v) Nonidet P.40, scraped into centrifuge tubes and centrifuged for 3 min at 300 *g* in an MSE Centaur bench-top centrifuge. The nuclei were resuspended in 70–150 μ l of extraction buffer and were subjected to three cycles of freeze/thawing as described above.

Pairs of oligonucleotides were hybridized by heating equimolar mixtures to 94 °C for 10 min and then allowing them to cool to room temperature. The oligonucleotide pairs (1–2 pmol) were end-labelled using [γ - 32 P]ATP and T4 polynucleotide kinase and the products were separated from unincorporated [γ - 32 P]ATP using NucTrap columns (Stratagene), equilibrated and eluted with STE (100 mM NaCl/1 mM EDTA/10 mM Tris, pH 8.0) in a final volume of approx. 100 μ l.

DNA mobility-shift assays were performed essentially as described [22]. The binding reactions consisted of 1–2 μ l of cell or nuclear extract (or extraction buffer in controls), 1 μ l (10–20 fmol) of 32 P-labelled oligonucleotide pairs and 1 μ l of

unlabelled competitor oligonucleotide pairs (as required), in a total volume of 20 μ l of reaction buffer [1 mM $MgCl_2$ /4% (w/v) Ficoll/0.5 mM dithiothreitol/0.01% (w/v) poly(dI-dC)/20 mM Hepes, pH 7.9]. The reactions were incubated on ice for 45 min and were then loaded onto a non-denaturing 4% (w/v) acrylamide gel, which had been pre-run in TBE (89 mM Tris base/89 mM boric acid/2.5 mM EDTA) at 150 V for 2.5 h. The gel was run for a further 2–2.5 h, dried and exposed to X-ray film for 20–40 h.

RESULTS

Sequence of the cMG1 5'-flanking region

In order to determine the 5'-flanking sequence of the cMG1 gene, we isolated and sequenced a 1.3 kb rat genomic clone which hybridized to probes derived from the cMG1 cDNA sequence [1], as described in the Experimental section. Comparison with the cDNA sequence indicated that this genomic clone contained 931 bp of 5'-flanking sequence, 134 bp of 5'-untranslated mRNA, the first 57 bp of the cMG1 coding region and 189 bp of an intron. Figure 1(a) shows the sequence of the first 1132 bp of this

clone, including the first 12 bp of the intron. The site of the exon/intron boundary was identified by the discrepancy between the genomic and cDNA sequences at this point, and is marked by a classic splice site (AG:GT), which is also found in the human ERF-1 sequence at this position [24].

The transcription start site (+1 in Figure 1a) was identified by primer-extension analysis using an oligonucleotide primer complementary to bases 140–121 of the cMG1 cDNA sequence [1]. The predominant extension product ran just below the 162-base marker on a sequencing gel (Figure 1b), indicating that the start site lay 21 bases upstream of the cDNA sequence; this is the position designated +1 (GAG...) in Figure 1. However, it should be noted that a start site two bases earlier (giving an mRNA beginning GAGAG...) might account for the lower-intensity band observed at about 163 bases in primer-extension analysis (Figure 1b). Figure 1 also shows that there is a TATAAA sequence at –33 to –30. When we compared the cMG1 5'-flanking sequence against a database of transcription factor recognition sites (using the MAP program from the University of Wisconsin Genetics Computer Group package) a number of consensus sites were identified. Some of these were for factors

Figure 1 DNA sequence of the 5'-flanking region of the cMG1 gene

(a) The methods used for the derivation and sequencing of a rat genomic clone are described in the text. The TATAAA sequence (–33 to –28) is shown in bold type, and the transcription start site identified by primer-extension analysis (b) is marked +1. The cDNA sequence previously reported [1] begins at base 22, and the translation initiation codon (bases 135–137) is shown in italics. Consensus binding sequences for certain transcription factors are underlined. (b) Primer extension analysis using an oligonucleotide primer complementary to bases 140–121 of the cMG1 cDNA sequence [1] was performed as described in the text and the products were run against ^{32}P -labelled pBR321-derived markers. The principal product is indicated by an arrow.

Figure 2 CAT expression from cMG5'[n]CAT plasmids in transfected NIH-3T3 cells

NIH-3T3 cells were transfected with the cMG5'[931]CAT and deletion plasmids; mock-transfected control cells were subjected to the transfection protocol without added DNA. Extracts were prepared from the cells and CAT activity was determined, as described in the text. The results shown are from a single experiment and similar results were obtained in at least two other experiments for each deleted plasmid.

Table 1 CAT expression from cMG5'[n]CAT plasmids in transfected NIH-3T3 cells

NIH-3T3 cells were transfected with the cMG5'[n]CAT plasmids as described in the text; mock-transfected control cells were subjected to the transfection protocol without added DNA. Extracts were prepared from the cells and CAT activity was determined, as described in the text. The results shown are the means of duplicate dishes in a single experiment; similar results were obtained for each plasmid in two other experiments.

Plasmid	[¹⁴ C]Chloramphenicol conversion (c.p.m.)
cMG5'[200]CAT	3528
cMG5'[177]CAT	4710
cMG5'[163]CAT	4103
cMG5'[138]CAT	3868
cMG5'[114]CAT	1212
cMG5'[96]CAT	1555
cMG5'[90]CAT	1328
cMG5'[66]CAT	928
cMG5'[51]CAT	129
(Mock transfection)	76

showing tissue- or gene-specific activity (HNF-5, MCBF and H4TF1) [25], which are unlikely to be generally involved in the expression of cMG1. However, the search also revealed recognition sites for certain transcription factors (AP2, PEA3, NF-IL6, *c-myc* and E2F) that have been implicated in signalling responses associated with proliferation, differentiation or development; these sites are shown in Figure 1.

Direction of CAT expression in NIH-3T3 cells by the 5'-flanking sequence of the cMG1 gene

A 1033 bp restriction fragment, containing 931 bp of the 5'-flanking sequence and 102 bp of untranslated mRNA, was ligated into the pCATbasic plasmid as described in the Experimental section. This construct (cMG5'[931]CAT) directed a substantial expression of CAT activity when transfected into NIH-3T3 cells (Figure 2). Under our standard assay conditions, extracts from mock-transfected cells, or from cells transfected with pCATbasic,

Figure 3 DNA mobility-shift analysis of complexes between oligonucleotide pairs from -143 to -105 of the cMG1 5'-flanking region and proteins in NIH-3T3 cell extracts

Whole-cell extracts from NIH-3T3 cells were incubated with ³²P-labelled oligonucleotide pairs and 100-fold excess unlabelled oligonucleotide pairs, as indicated, and subjected to PAGE under non-denaturing conditions. The first track in each group contains ³²P-labelled oligonucleotide pairs without added extract. The results shown in (a) and (b) are from two separate experiments. In addition to the incubations shown, the oligonucleotide pairs C and F were labelled with ³²P and used in parallel incubations in experiments (a) and (b) respectively, but no complexes were observed. Similar results were obtained in at least one other experiment with each combination of labelled and unlabelled oligonucleotide pairs. (c) Shows the positions of the oligonucleotide pairs within the cMG1 5'-flanking sequence. Pair F comprises bp -119 to -105 from the deletion plasmid cMG5'[114]CAT, in which bp -119 to -115 are not of the cMG1 5'-flanking sequence, and pair G comprises the c/EBP consensus sequence described in the text and used in Figure 4.

catalysed less than 0.2% conversion of [¹⁴C]chloramphenicol, while extracts from cells transfected with the positive control plasmid pRSV-CAT catalysed between 50 and 60% conversion; extracts from cells transfected with cMG5'[931]CAT typically catalysed 5-10% conversion of [¹⁴C]chloramphenicol, and in some experiments up to 25% conversion was achieved.

In order to define the region of cMG5'[931]CAT that was responsible for directing CAT expression, progressive 5'-to-3' deletion of the inserted sequence was used to produce a series of plasmids with decreasing amounts of proximal 5'-flanking sequence upstream of the CAT reporter. Very little change in activity was observed when the cMG1 5'-flanking sequence was reduced from 931 bp to 200 bp upstream of the transcription

Figure 4 DNA mobility-shift analysis comparing complexes formed by oligonucleotide pairs from the cMG1 5'-flanking region and a consensus c/EBP-binding sequence

Whole-cell extracts from NIH-3T3 cells were incubated with ^{32}P -labelled oligonucleotide pairs and 100-fold excess unlabelled oligonucleotide pairs as indicated [D, cMG5'-flank -124 to -105; G, c/EBP consensus (see legend to Figure 3)], and subjected to PAGE under non-denaturing conditions. The tracks marked α , β and δ show incubations to which anti-c/EBP- α , - β or - δ antibodies (0.4 μl) were added for the second half of the binding reaction. Similar results were obtained in a second experiment.

start site, but a substantial loss of activity occurred on deletion to 42 bp (Figure 2). We then tested additional plasmids with deletions within this region. CAT expression was driven by a plasmid retaining 138 bp of the cMG1 5'-flanking sequence, but the majority (75–80%) of CAT expression was lost when the cMG1 5'-flanking sequence was reduced to 114 bp (Table 1). Loss of the remaining CAT expression (i.e. to a level similar to that in cells transfected with pCATbasic) occurred when the upstream sequence was reduced from 66 to 51 bp.

DNA mobility-shift assay of the region directing CAT expression

Given the results from the cMG5'[n]CAT deletion analysis, we investigated the ability of proteins in cellular extracts to bind to DNA fragments spanning the region 138–114 bp upstream of the transcription start site. When a pair of ^{32}P -labelled oligonucleotides corresponding to bp -143 to -105 were incubated with extracts from NIH-3T3 cells and electrophoresed on non-denaturing polyacrylamide gels, several retarded complexes were observed (Figure 3). The formation of most of these complexes, including the major doublet, was blocked by the inclusion of a 100-fold excess of the unlabelled pair -143 to -105 (Figure 3, competitor A). However, complex formation was not affected by a 100-fold excess of the unlabelled pair -138 to -110 (Figure 3, competitor B), and ^{32}P -labelled -138 to -110 was itself unable to form complexes with proteins in the extracts (results not shown). These results suggested that the complexes were being formed by sequences at one or both ends of the -143 to -105 sequence, i.e. including or overlapping with the end(s) of the -138 to -114 region defined by deletion analysis.

In order to localize the binding region more precisely, oligonucleotide pairs corresponding to the 5' and 3' halves of the -143 to -105 sequence were therefore prepared, and these were

Figure 5 DNA mobility-shift analysis of complexes between oligonucleotide pairs from -143 to -105 of the cMG1 5'-flanking region and proteins in extracts from control and stimulated RIE-1 cells

Confluent, quiescent cultures of RIE-1 cells were incubated for 30 min at 37 °C with vehicle (CON), insulin (INS; 1 $\mu\text{g}/\text{ml}$), angiotensin II (All; 100 nM) or epidermal growth factor (EGF; 20 ng/ml). Whole-cell extracts were prepared as described in the text, incubated with ^{32}P -labelled -143 to -105 and 100-fold excess unlabelled -143 to -105, as indicated, and subjected to PAGE under non-denaturing conditions. Similar results were obtained in a second experiment.

used in both labelled and unlabelled form for binding and competition. Formation of the major doublet using ^{32}P -labelled -143 to -105 was largely unaffected by oligonucleotide pairs corresponding to bp -143 to -124 (Figure 3a, competitor C), but was blocked by the inclusion of -124 to -105 (competitor D). Consistent with these results, ^{32}P -labelled -143 to -124 did not produce significant complex formation, while ^{32}P -labelled -124 to -105 formed complexes equivalent to the major doublet (Figure 3a). Further reduction of the sequence to bp -119 to -105 still allowed both competition and complex formation (Figure 3b, competitor E). However, we then prepared an oligonucleotide pair corresponding to bp -119 to -105 of the plasmid cMG5'[114]CAT (i.e. in which bp -119 to -115 were not of the cMG1 5'-flanking sequence) which had not directed the expression of CAT activity in the transfection assay (Table 1). This oligonucleotide pair was unable to form complexes, and did not inhibit the formation of complexes by ^{32}P -labelled -143 to -105 or -119 to -105 (Figure 3b, competitor F). Taken together, the results indicate that the main protein complex is formed across the 3' end of the -138 to -114 region defined by the cMG5'[n]CAT transfection experiments.

Comparison with the c/EBP family

The sequence of bp -119 to -105 (GGTGTCAATCAACTT) contains the CAATC motif that is found in the palindromic consensus sequence recognized by the c/EBP family of transcription factors [16]. We therefore compared the activity of the

Figure 6 Analysis of cMG1 mRNA expression and cMG1 5'-flanking region activity in PC12 cells

(a) RNA was prepared from control and epidermal growth factor (EGF)-stimulated RIE-1 and PC12 cells, and was subjected to Northern blotting followed by hybridization to a ^{32}P -labelled cMG1 cDNA probe; all methods have been described previously [11]. Analysis of 28 S and 18 S rRNA, and reprobing for GAPDH mRNA, indicated that there was no significant difference in RNA loading between tracks. CON, Control. (b) Whole-cell extracts from PC12 cells or NIH-3T3 cells were incubated with ^{32}P -labelled oligonucleotide pairs and 100-fold excess of unlabelled oligonucleotide pairs, as indicated, and subjected to PAGE under non-denaturing conditions, as described in the text. (c) PC12 cells were transfected with pCATbasic, cMG5'[931]CAT or cMG5'[138]CAT, as described in the text. The autoradiograph presented shows CAT conversion by extracts from duplicate dishes of cells from each electroporation.

c/EBP consensus recognition sequence (TGCAGATTGCGCA-ATCTGCA; [16]) in the mobility-shift assay with that of the cMG1 5'-flanking region oligonucleotide pair -124 to -105 having the same length. Extracts from each cell-line tested (NIH-3T3, RIE-1 and PC12) contained proteins which bound to the c/EBP consensus sequence (Figure 4 and results not shown). However, these complexes were slightly larger than those formed by cMG1 -124 to -105, and each unlabelled oligonucleotide pair showed only limited ability to inhibit complex formation by the other labelled oligonucleotide pair (Figure 4). Furthermore, the complex formed by NIH-3T3 extract with the c/EBP consensus sequence was 'super-shifted' by antiserum against c/EBP β , but the complexes formed with cMG1 -124 to -105 were not affected by antisera against c/EBP- α , - β or - δ (Figure 4).

Distribution of the proteins forming complexes with cMG5'[n]CAT

The pattern of binding complexes obtained using extracts from NIH-3T3 cells was also observed in extracts from Swiss 3T3 fibroblasts and RIE-1 epithelial cells (Figure 5 and results not shown), both of which express cMG1 mRNA [1,2,11]. Incubation of RIE-1 cells for 30 min with the mitogens epidermal growth factor, angiotensin II or insulin, each of which increases the level

of cMG1 mRNA [11], did not have any substantial effect on the binding complexes observed (Figure 5). All of the bands observed were also obtained when nuclear extracts of NIH-3T3 cells or RIE-1 cells were used instead of whole-cell extracts (results not shown).

In contrast to the 3T3 and RIE-1 cell lines, neither control nor stimulated PC12 cells express cMG1 mRNA at levels detectable by Northern blotting, even after long exposures of the blots (Figure 6a and [2]). It was therefore of some interest to find that extracts from PC12 cells gave a similar pattern and amount of binding complexes in the mobility-shift assay to those obtained from the other cell-lines (Figure 6b). Furthermore, consistent with the hypothesis that these complexes are involved in gene expression, PC12 cells expressed substantial levels of CAT when transfected with the cMG5'[931]CAT or cMG5'[138]CAT plasmids (Figure 6c). These results suggest that PC12 cells contain transcription factors that can act at the cMG1 promoter, and that the lack of endogenous cMG1 expression in these cells is therefore due to additional factors (see below).

DISCUSSION

The mRNAs of both cMG1/TIS11b and its human equivalent ERF-1 are detected readily in various unstimulated cell-lines [1,2,24], while the related gene TIS11/Nup475 shows little or no basal mRNA expression [2,4-6]. A number of results have focused our attention on the -119 to -105 region of the cMG1 5'-flanking sequence, as being required for basal cMG1 mRNA expression. First, comparison of this region with the 5'-flanking sequences of ERF-1 and TIS11/Nup475 shows that bp -118 to -105 are present at a similar position in ERF-1 [24], but that there is no similar sequence in the first 997 bp upstream of TIS11/Nup475 [6]. Secondly, deletion analysis of the 5'-flanking sequence of the cMG1 gene in CAT plasmids indicated that the majority of basal CAT expression depends on elements located in the region between bp -138 and -114 (Table 1). Thirdly, DNA mobility-shift analysis using oligonucleotides taken from -143 to -105 revealed a number of protein complexes, of which the major doublet could be formed by the sequence -119 to -105, which overlaps the 3'-end of the -138 to -114 region (Figure 3). Critically, formation of this complex did not occur when the cMG1 5'-flanking sequence between bp -119 to -115 was replaced with the equivalent base pairs from the deleted plasmid cMG5'[114]CAT (Figure 3), which did not direct CAT expression in the transfection assay (Table 1). The finding that bp -138 to -110 did not support complex formation suggests that bp -109 to -105 may also be critical for protein binding. However, when these bases in -143 to -105 or -124 to -105 were substituted by a *Xba*I restriction site, complex formation still occurred (A. N. Corps, unpublished work); thus, base pairs beyond -110 may only be required to maintain a double-stranded DNA context for the binding site. Taken together, our results are consistent with the hypothesis that bp -119 to -105 contain a control element which binds to regulatory proteins (as observed in the mobility-shift assay) to mediate the basal expression of cMG1.

Although the results from the deletion and mobility-shift assays suggest that formation of complexes at the -119 to -105 element is necessary and sufficient for cMG1 expression, the results obtained using PC12 cells indicate that additional factors may influence the endogenous expression of cMG1. PC12 cells do not express significant levels of endogenous cMG1 mRNA; nevertheless, they contain proteins which form complexes with oligonucleotide pairs containing the cMG1 -119 to -105 sequence, and they express CAT from the cMG5'[931]CAT and

cMG5[138]CAT plasmids (Figure 6). There are several possible explanations for the lack of endogenous expression of cMG1 in PC12 cells. The proteins forming the complex may require activation (for example, by phosphorylation) in order to interact either directly or via other proteins with the transcriptional machinery [26]; this occurs, for example, at the cyclic AMP response element (CRE) [27]. Alternatively, dominant inhibitory factors, binding outside the region studied here, may be associated with the endogenous cMG1 gene in PC12 cells; this interpretation is consistent with the observation that the cMG5[931]CAT plasmid is active when transfected into PC12 cells. This result is also compatible with a third explanation, namely that the cMG1 gene is disrupted or deleted in this cell-line.

Comparison of the cMG1 -119 to -105 sequence against the EMBL database revealed a perfect 13-base homology with the 5'-flanking region of the *Xenopus* albumin gene, about 85 bp upstream of the TATA box but on the opposite strand relative to the coding sequence [28]. There is also a 13 out of 15 bp identity of cMG1 -119 to -105 with a region of the rat albumin gene on the coding strand 105 bp upstream of the TATA box [29], and the 5'-end of this region overlaps with the DEIII regulatory element (element F in the mouse) [29,30]. The latter is one of several elements in the albumin promoter and conveys 3- to 4-fold activation of transcription [29,30]. This effect is similar in magnitude to that observed in the present study, raising the possibility that similar mechanisms may be involved.

At present, we do not know the identity of the proteins which form the complexes with the cMG1 -119 to -105 sequence. The mouse albumin element F is bound with low affinity by c/EBP [30], and, as noted above, the cMG1 -119 to -105 sequence is centred on the motif CAATC, which is half of the palindromic consensus sequence recognized by the c/EBP family of transcription factors [16]. There is no dyad symmetry in the cMG1 5'-flanking sequence, but members of this family have been shown to interact with rather diverse sequences in addition to the palindromic consensus: for example, a recent study has shown that c/EBP α is involved in complex formation at the dissimilar elements mediating the effects of cyclic AMP and insulin respectively on the phosphoenolpyruvate carboxykinase gene [18]. We therefore tested for the possible involvement of members of this family in complex formation at the -119 to -105 site in the cMG1 gene. However, there was little mutual inhibition between the cMG1 -124 to -105 oligonucleotide pair and a c/EBP consensus oligonucleotide pair in complex formation, and the complexes formed by the cMG1 -124 to -105 oligonucleotide pair were not affected by antisera against c/EBP- α , - β or - δ (Figure 4). It therefore seems unlikely that any of these factors is responsible for the basal cMG1 transcription driven from this region of the 5'-flanking sequence. In addition to the CAATC motif, there is a 14-bp sequence in the cMG1 5'-flanking region (bases -186 to -173) which contains the 9-bp NF-IL6 (c/EBP β) consensus sequence [25,31] and is particularly closely related to the NF-IL6 site in the tumour necrosis factor- α gene [31]. Interestingly, this 14-bp sequence is also perfectly repeated in the intron and is conserved in ERF-1 [24]. However, at least the first of the two sites does not appear to have any effect on the observed basal CAT expression, which did not decrease between plasmids cMG5[200]CAT and cMG5[163]CAT (Table 1). Whether these sites are involved in the response of cells to agonist-induced stimulation remains to be determined.

During the experiments described here, we found no evidence for the stimulation of CAT expression from any of the constructs by a variety of growth factors active on NIH-3T3 cells (results not shown). The most likely reason for this lack of effect is the relatively high basal expression of cMG1 (compared with, for

example, that of TIS11 or *c-fos*), such that a transient stimulation of CAT expression from cMG5[931]CAT plasmids is not detectable against the basal accumulation of CAT enzyme. Examination of the cMG1 5'-flanking sequence indicates that no DNA-binding elements are present for some of the transcription factors most commonly associated with primary response gene activation [AP1, serum response element (SRE), CRE, *sis*-inducible element] [12-14]. However, some of the consensus sequences that are present (Figure 1) are conserved in the human gene [24], and these could mediate the effects of the various agonists which modulate the expression of cMG1 [1,2,11]. Finally, the possibility also exists that there are control sequences that lie outside the region covered in this study. For example, the *JunB* gene has functional SRE and CRE located about 2.1 kb downstream of the transcription start site, in the 3'-flanking region [32]. Furthermore, the *JE* gene contains a 7-nucleotide motif within its 3'-untranslated region, which mediates the induction of *JE* by platelet-derived growth factor, and is also found in a similar 3' position in a number of primary response genes [33]. Interestingly, this motif (TTTTGTA) is present in the long 3'-untranslated region of cMG1 mRNA (at bp 2473-2479 in the cDNA sequence [1], which is about 1350 bp beyond the coding region), is conserved in the mouse and human cMG1 homologues [2,24] and is also present in TIS11 [6]. Further work will be required to determine whether this or any of the other potential sites described can mediate the stimulated expression of cMG1.

In summary, we have cloned the 5'-flanking region of the cMG1 gene and identified the start site for transcription and a TATAAA sequence. We have demonstrated that a sequence element, located between 119 and 105 bp upstream of the start site, appears to be responsible for basal expression of the gene. Future studies will be directed towards identifying the protein(s) which interact with this promoter element and defining the mechanisms involved in inducible cMG1 expression.

We thank Miranda Gomperts for performing the primer extension analysis and Dr. S. L. McKnight for anti-c/EBP antibodies. A. N. C. was supported by a grant from the British Diabetic Association.

REFERENCES

- Gomperts, M., Pascall, J. C. and Brown, K. D. (1990) *Oncogene* **5**, 1081-1083
- Varnum, B. C., Ma, Q., Chi, T., Fletcher, B. and Herschman, H. R. (1991) *Mol. Cell. Biol.* **11**, 1754-1758
- Barnard, R. C., Pascall, J. C., Brown, K. D., McKay, I. A., Williams, N. S. and Bustin, S. A. (1993) *Nucleic Acids Res.* **20**, 3580
- Varnum, B. C., Lim, R. W., Sukhatme, V. P. and Herschman, H. R. (1989) *Oncogene* **4**, 119-120
- Lai, W. S., Stumpo, D. J. and Blackshear, P. J. (1990) *J. Biol. Chem.* **265**, 16556-16563
- DuBois, R. N., McLane, M. W., Ryder, K., Lau, L. F. and Nathans, D. (1990) *J. Biol. Chem.* **265**, 19185-19191
- Mohler, J., Weiss, N., Murli, S. et al. (1992) *Genetics* **131**, 377-388
- Ma, Q., Wadleigh, D., Chi, T. and Herschman, H. (1994) *Oncogene* **9**, 3329-3334
- Ma, Q. and Herschman, H. R. (1995) *Oncogene* **10**, 487-494
- Gomperts, M., Pascall, J. C. and Brown, K. D. (1990) *Biochem. Soc. Trans.* **18**, 568-569
- Gomperts, M., Corps, A. N., Pascall, J. C. and Brown, K. D. (1992) *FEBS Lett.* **306**, 1-4
- McMahon, S. B. and Monroe, J. G. (1992) *FASEB J.* **6**, 2707-2715
- Treisman, R. (1994) *Curr. Opin. Genet. Dev.* **4**, 96-101
- Han, T.-H., Lamph, W. W. and Prywes, R. (1992) *Mol. Cell. Biol.* **12**, 4472-4477
- Lucas, P. C. and Granner, D. K. (1992) *Annu. Rev. Biochem.* **61**, 1131-1173
- Cao, Z., Umek, R. M. and McKnight, S. L. (1991) *Genes Dev.* **5**, 1538-1552
- Diehl, A. M., Michaelson, P. and Yang, S. Q. (1994) *Gastroenterology* **106**, 1624-1637
- O'Brien, R. M., Lucas, P. C., Yamasaki, T., Noisin, E. L. and Granner, D. K. (1994) *J. Biol. Chem.* **269**, 30419-30428
- Brown, K. D. and Blakeley, D. M. (1983) *Biochem. J.* **212**, 465-472

- 20 Blay, J. and Brown, K. D. (1985) *Biochem. J.* **225**, 85–94
- 21 Rosenthal, N. (1987) *Methods Enzymol.* **152**, 704–720
- 22 Dent, C. L. and Latchman, D. S. (1993) in *Transcription Factors: A Practical Approach* (Latchman, D. S., ed.), pp. 1–26, IRL Press, Oxford
- 23 Dignam, J. D., Lebovitz, R. M. and Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475–1489
- 24 Bustin, S. A., Nie, X.-F., Barnard, R. C. et al. (1994) *DNA Cell Biol.* **13**, 449–459
- 25 Faisst, S. and Meyer, S. (1992) *Nucleic Acids Res.* **20**, 3–26
- 26 Hunter, T. and Karin, M. (1992) *Cell* **70**, 375–387
- 27 Lalli, E. and Sassone-Corsi, P. (1994) *J. Biol. Chem.* **269**, 17359–17362
- 28 Schorpp, M., Dobbeling, U., Wagner, U. and Ryffel, G. U. (1988) *J. Mol. Biol.* **199**, 83–93
- 29 Herbomel, P., Rollier, A., Tronche, F., Ott, M.-O., Yaniv, M. and Weiss, M. C. (1989) *Mol. Cell. Biol.* **9**, 4750–4758
- 30 Maire, P., Wuarin, J. and Schibler, U. (1989) *Science* **244**, 343–346
- 31 Akira, S., Isshiki, H., Sugita, T. et al. (1990) *EMBO J.* **9**, 1897–1906
- 32 Perez-Albuerne, E. D., Schatteman, G., Sanders, L. K. and Nathans, D. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11960–11964
- 33 Freter, R. R., Irminger, J.-C., Porter, J. A., Jones, S. D. and Stiles, C. D. (1992) *Mol. Cell. Biol.* **12**, 5288–5300

Received 24 April 1995; accepted 19 May 1995