Neuronal expression of the rat M_1 muscarinic acetylcholine receptor gene is regulated by elements in the first exon

Ian C. WOOD^{1,2}, Mireia GARRIGA CANUT¹, Claire L. PALMER, Stefania PEPITONI and Noel J. BUCKLEY

Wellcome Laboratory for Molecular Pharmacology, Department of Pharmacology, University College London, London WC1E 6BT, U.K.

Muscarinic acetylcholine receptor genes are members of the G-protein coupled receptor superfamily. Each member of this family studied to date appears to have a distinct expression profile, however the mechanisms determining these expression patterns remain largely unknown. We have previously isolated a genomic clone containing the M_1 muscarinic receptor gene and determined its gene structure [Pepitoni, Wood and Buckley (1997) *J. Biol. Chem.* **272**, 17112–17117]. We have now identified DNA elements responsible for driving cell specific expression in transient transfection assays of immortalized cell lines. A region of the gene spanning 974 nucleotides and containing 602 nucleotides of the first exon is sufficient to drive specific expression in cell lines. Like the M_4 and M_2 gene promoters, the M_1 promoter contains an Sp1 motif which can recruit transcription factor Sp1 and at least one other protein, although this site does not appear

to be functionally important for M_1 expression in our assay. We have identified a region within the first exon of the M_1 gene that regulates expression in cell lines, contains several positive and negative acting elements and is able to drive expression of a heterologous promoter. A polypyrimidine/polypurine tract and a sequence conserved between M_1 genes of various species act in concert to enhance M_1 transcription and are able to activate a heterologous promoter. We show that DNA binding proteins interact *in vitro* with single-stranded DNA derived from these regions and suggest that topology of the DNA is important for regulation of M_1 expression.

Key words: gene expression, single-stranded DNA-binding proteins, transcription.

INTRODUCTION

G-protein coupled receptors (GPCRs) are responsible for mediating a great deal of intercellular communication, particularly within the nervous system. The GPCR gene superfamily is estimated to contain over 1000 members, making it one of the largest in the mammalian genome. Within the nervous system, many of these genes are expressed in specific neuronal subpopulations. Despite the size and importance of this gene family, little is known about the transcriptional mechanisms that direct their cell specific expression. The five muscarinic receptor genes were the first complete GPCR subfamily to be cloned [1,2] and mapped within the nervous system [1,3]. Muscarinic receptors mediate the metabotropic actions of acetylcholine in the nervous system and effector tissues [4], where each muscarinic receptor gene (M_1 – M_5) shows a unique expression profile [1,3]. As such, they present an ideal model for examining the transcriptional mechanisms responsible for driving gene expression in subpopulations of differentiated neurones. The M_1 receptor is the most abundant subtype in the central nervous system and activation of the M_1 receptor leads to stimulation of phospholipases C [5] and A_2 [6], inhibition of cAMP production [1], activation of K^+ and Cl^- channels [7], and inhibition of the opening of K^+_M channels [8,9]. Relatively little is known about the functional effects of M_1 activation but they have been implicated in learning and memory [10,11]. Recently, transgenic mice lacking the M_1 receptor gene have been produced [12]. These mice are resistant to epileptic seizures induced by muscarinic agonists and show no muscarinic modulation of the M-type potassium channel in sympathetic ganglion neurones, suggesting that the M_1 receptor plays a role in the control of

neuronal excitability. As a first step to identifying the genomic elements responsible for cell-specific expression of the M_1 gene we have previously reported the isolation of the rat M_1 gene and shown that it consists of a 657 bp non-coding exon separated by a 13.5 kb intron from a single 2.54 kb coding exon [13]. In common with several other GPCRs, the M_1 gene lacks TATA and CAAT boxes and has an Sp1 binding site in the vicinity of the transcription start site. All of these features are shared by the rat M_4 gene [14,15] and the chicken M_2 gene [16]. Expression patterns of the M_1 and M_4 genes are broadly similar: both genes are expressed by autonomic ganglia and, in the central nervous system, expression is largely restricted to telencephalic regions. In the latter case, many neurones of the cortex, hippocampus and striatum co-express both M_1 and M_4 genes, but there are also subpopulations of neurones that express only one or neither gene [17,18]. Because these two genes are family members, share a common gene structure and have similar expression patterns, we were interested to see whether their expression involved common regulatory mechanisms. In previous studies, we have shown that the M_4 gene is under the control of a constitutive core promoter which is repressed in non-expressing cells, at least in part, by recruitment of a zinc finger silencer, RE1-silencing transcription factor/neuron-restrictive silencer factor (REST/NRSF) [19,20], to a repressor element 1/neuron-restrictive silencer element (RE1/NRSE) type silencing element [21,22]. Whilst expression driven by the promoter region of the M_2 gene has not been analysed in cells that do not endogenously express M_2 , it has been proposed that it may also contain a silencer element [16]. Analysis of the M_1 promoter region that drives cell specific expression *in vitro* has not identified any RE1/NRSE silencing element [13]. The M_2 gene does appear to be positively regulated

Abbreviation used: GPCR, G-protein coupled receptor.

¹ These authors contributed equally to the work.

² To whom correspondence should be addressed (e-mail i.wood@ucl.ac.uk).

by the GATA family of transcription factors which act via binding sites within the proximal promoter region and within the first exon [22a]. It would appear that despite the similar expression profiles of the M_1 and M_4 genes, transcriptional regulation of these members of the muscarinic receptor gene family is directed by distinct transcriptional mechanisms. In this study we have identified several regions within the first exon that are conserved between M_1 genes in a variety of species and act as regulators of expression. One of these regions is a polypyrimidine/polypurine tract, the others contain no recognizable motif. Two regions, including the polypyrimidine/polypurine tract and a region that is conserved between species, recruit single-stranded DNA-binding proteins. We also show that the Sp1 site present within the M_1 gene is capable of recruiting transcription factor Sp1 and at least one other protein *in vitro*, although this does not appear to be required for driving reporter gene expression in transient transfection assays.

MATERIALS AND METHODS

Cell culture

Cell lines were cultured in 5% CO₂ at 37 °C and all media were supplemented with 6 g/l penicillin, 10 g/l streptomycin and 2 mM L-glutamine. IMR32 and 3T3 cells were grown in Dulbecco's modified Eagle's medium (Life Technologies Inc.) supplemented with 10% (v/v) fetal-calf serum. Neuro 2a cells were grown in minimal essential medium (Life Technologies Inc.) supplemented with 10% fetal-calf serum. NB4 1A3 cells were grown in Ham's F10 (Life Technologies Inc.) supplemented with 15% horse serum and 2.5% fetal-calf serum.

Reverse transcription PCR

RNA was extracted from cell lines and brain tissue using RNazol B (Biogenesis Ltd.) and reverse transcribed using oligo(dT) and MMuLV reverse transcriptase (Promega). The oligonucleotides used to amplify the M_1 gene were: Rm1 +299s (5'-GGGCT-CCCTTTCCTCATGTC), Rm1 -212s (5'-CCCCACAAACC-AGACACC), Rm1 -140s (5'-GGCATGTGCTCAGTCCC), Rm1 +507s (5'-GGGACATGCCAATCACTG), Rm1 31a (5'-CTGGTGCCAAGACAGTGATGTTGGG) and Rm1 111a (5'-GCTGGCAGCCGAGAGGCACA). Cycling conditions were 95 °C for 5 min, and then 30 cycles of 95 °C for 30 s, 55 °C for 1 min and 72 °C for 1 min followed by a final step of 72 °C for 10 min. Aliquots of the reaction mixture were visualized on a 2% (w/v) agarose gel. PCR products were cloned using the pGEM-T Easy vector (Promega) and recombinant plasmids sequenced using *Taq* polymerase, fluoresceinated dye terminators and an Applied Biosystems 377 automated DNA sequencer. Sequences of the human mouse and rat M_1 genes were submitted to EMBL: accession numbers AJ006520, AJ006521 and AJ006522 respectively.

Reporter plasmid construction

Reporter plasmids were constructed using the luciferase reporter vector pGL3 Basic (Promega). An oligonucleotide corresponding to sequences -372 to -352 and containing a *Bgl*II linker was used in conjunction with an oligonucleotide to the luciferase gene, GL2, to amplify the region between -372 and +602 by PCR from the plasmid pGL3 K/P1.6 [13]. The PCR product was digested with *Bgl*II and *Hind*III and cloned into *Bgl*II/*Hind*III-cut pGL3 Basic to generate pGL3 -372/+602. pGL3 -372/+602ΔSp1, which has a deletion of nucleotides +17 to +23 of the M_1 gene, was generated by fusion PCR as follows: PCR

products obtained from the amplification of pGL3 -372/+602 with the oligonucleotides RV3 and M_1 ΔSp1a (5'-CGCCCCAC-CCAACCCACCCAGCCACACAC), and GL2 and M_1 ΔSp1s (5'-CTGGGTGGGTGGGTGGGCGTGTGGGCA), were combined and used in a second round of amplification with the oligonucleotides RV3 and GL2 (vector primers for use with pGL3; Promega). The resulting PCR product was digested with *Bgl*II and *Hind*III and cloned into *Bgl*II/*Hind*III-digested pGL3 Basic. pGL3 -372/+602 was partially digested with *Sma*I and religated to generate pGL3 -372/+166. Oligonucleotides corresponding to sequences +259 to +241, +299 to +316, +348 to +366, +399 to +417 and +504 to +486 (with *Bgl*II linkers), were used in conjunction with an oligonucleotide corresponding to nucleotides -372 to -352 (also with a *Bgl*II linker) to construct pGL3 -372/+241, pGL3 -372/+316, pGL3 -372/+366, pGL3 -372/+417 and pGL3 -372/+504 respectively by PCR. pGL3 -372/+602 ΔPPY was constructed by fusion PCR with the two sets of oligonucleotides, fusion M_1 Δ+412/+485s (5'-CCCAGATGCTGCCTGAGGAAGGC-TGGCG) and an oligonucleotide corresponding to sequences -372 to -352 and fusion M_1 Δ+412/+485a (5'-CCCTGGC-ACGCCAGCCTTCCTCAGGCA) and an oligonucleotide corresponding to sequences +602 to +583. A PCR product obtained with M_1 +396s and GL2 was digested with *Xma*I and cloned into *Xma*I-digested pGL3 Basic to make pGL3 +396/+569 M_1 . The plasmid pGL3 -1390/+602 M_1 was partially digested with *Sma*I and religated to give pGL3 +166/+602 M_1 . Both of these constructs were digested with *Hind*III and annealed oligonucleotides containing the initiator sequence from the adenovirus late promoter [23] with *Hind*III compatible ends was inserted to produce pGL3 +396/+569 M_1 Inr and pGL3 +166/+602 M_1 Inr respectively. The same oligonucleotides were inserted into *Hind*III-digested pGL3 Basic to produce pGL3 Inr. DNA sequencing was performed to check the identity of the reporter plasmids constructed.

DNA transfections

Qiagen column purified DNA was transfected into cells using Tfx[™] 50 (Promega) according to the manufacturer's instructions. Transfections were carried out in 10 mm wells using 500 ng reporter plasmid, 2 ng pRL-CMV and 1.5 μl Tfx[™] 50 in a total volume of 200 μl and applied for 1 h. Cells were harvested 1–2 days later into 100 μl Passive Lysis Buffer (Promega) of which 30 μl was used in Promega's DLR[™] system. Luminescence was measured using a Turner TD-20E luminometer; firefly luciferase was normalized to *Renilla* luciferase and the results expressed relative to normalized luminescence driven from the promoterless pGL3 Basic.

Gel mobility shift assays

Nuclear protein was extracted from tissue culture cell lines as described by Dent and Latchman [24] and the concentration determined by a D_c protein assay kit (Bio-Rad). For single-stranded DNA binding, oligonucleotides were labelled using T4 polynucleotide kinase and [γ -³²P]ATP. Oligonucleotides spanning the Sp1 binding site in the M_1 promoter were synthesized: M_1 Sp1s (5'-GTACGGTGGGTGGGCTGGGT) and M_1 Sp1a (5'-GTACCAGCCCCACCCACC). When annealed these oligonucleotides form a 5' overhang at each end which was labelled by Klenow fill-in using [α -³²P]dATP. Nuclear protein (1–5 μg) was incubated on ice, with or without competitor DNA, for 20 min in 19 μl of a solution containing 20 mM Hepes, pH 7.9, 100 mM KCl, 5 mM MgCl₂, 8% (v/v) glycerol and 1 μg of calf thymus DNA. Approximately 10000–20000 c.p.m. of radioactive probe

was added to each reaction and the incubation continued at room temperature for 20 min. For supershift experiments 1 μ l of anti-Sp1 antibody (1 mg/ml, Santa Cruz) was added to the reactions and the incubation continued for a further 20 min. Reactions were electrophoresed on 22.25 mM Tris/borate, pH 8.3, 0.5 mM EDTA, 4% (w/v) polyacrylamide gels, and the gel was dried and exposed to X-ray film.

Southwestern analysis

Proteins (15 μ g) were separated by SDS/PAGE [6% (w/v) gel] and transferred on to a Hybond C+ Extra nylon membrane (Amersham). The proteins were then renatured and probed with radioactive probes at 1×10^6 d.p.m./ml in the same buffer used for the gel shift assays as described by Vinson et al. [25].

UV crosslinking

Gel mobility shift reactions were prepared as described and subjected to UV exposure from an inverted transilluminator for 60 min as described by Cooney et al. [26]. Samples were separated by SDS/PAGE [10% (w/v) gel] and the gel was dried and exposed to X-ray film.

Yeast one-hybrid screening

A 913 bp region of the M₁ promoter, corresponding to nucleotides -727 to +186, was cloned into the yeast vector pBM2389 [27] and transformed into the yeast strain SFY526 [28]. This yeast strain was then transformed with DNA from an E17 mouse cDNA yeast expression library (Clontech), using the protocol of Schiestl and Gietz [29], and transformants were grown on complete supplement mixture -His/-Leu/-Trp (Bio 101, Vista, CA, U.S.A.), containing 5 mM 3-amino-1,2,4-triazole (Sigma) to select for interactions. Library candidates were tested for their ability to specifically activate the M₁-containing reporter plasmid by re-transformation back into SFY526. Library plasmids producing interacting proteins were sequenced for identification.

RESULTS

The M₁ 5' untranslated region contains highly conserved sequences across species

Previously we have isolated the rat gene encoding the M₁ muscarinic receptor [13]. A sequence alignment of the rat [13] and porcine [30] M₁ genes showed that the non-coding sequence was highly conserved between the two species. The first 492 nucleotides from the longest rat cDNA clone showed little homology to the 5' end of the porcine cDNA. Further comparisons between these two sequences revealed homology between the 5' end of the porcine cDNA and a region between -212 to -126 of the rat gene (Figure 1a). As the 3' end of this sequence showed sequence similarity with a splice donor, and the region immediately upstream of +489 showed sequence similarity with a splice acceptor, we used a PCR assay to determine whether this region was transcribed from the rat gene. Several sense primers in the rat 5' region were used in conjunction with an antisense primer, Rm1 31a, in the coding exon of cDNA templates derived from a human cell line (IMR32), mouse cerebral cortex or rat cerebral cortex. Two of these primers, Rm1 +299s and Rm1 +499s, showed specific amplified products of the expected size from each of the species of cDNA tested (Figure 1b). Neither of the two primers to the rat sequence showing homology to the 5' end of the porcine cDNA, Rm1 -212s and Rm1 -140s, gave amplified products from cDNA in this PCR

Figure 1 A sequence in rat similar to the 5' end of the pig M₁ cDNA is not transcribed

(a) Sequence alignment of the 5' end of the pig cDNA [30] and nucleotides -212 to -118 of the rat gene. The positions of the primers used in PCR analysis are denoted by arrows. (b) PCR was performed using an antisense primer in the coding exon, Rm1 31a, and the indicated sense primers on cDNA from human IMR32 cells, mouse cerebral cortex and rat cerebral cortex. PCR products were electrophoresed through a 2% (w/v) MetaPhor agarose gel (FMC BioProducts, Rockland, ME, U.S.A.). DNA marker shown on the outer lanes is 1 kb ladder (Life Technologies Inc.). (c) PCR was performed on genomic DNA using an antisense primer, Rm1 111a, and the sense primers indicated.

assay, demonstrating that this region of the rat gene was not transcribed, at least in IMR32 cells, mouse and rat cerebral cortex. As a control these primers were shown to produce amplified products from a genomic DNA template in conjunction with a second antisense primer (Figure 1c). Interestingly, a single PCR band was seen for human and rat cDNA, whereas mouse cDNA showed two bands, suggesting the use of an alternatively spliced exon. Each of the amplified products were cloned into the pGEM-T Easy cloning vector and sequenced. An alignment of these sequences is shown in Figure 2, as is the corresponding sequence of the pig M₁ gene reported by Kubo et al. [30]. A sequence comparison between the 5' untranslated region of human, mouse, rat and pig showed that there was extensive homology between the four species. The larger and smaller PCR products from the mouse cDNA were found to correspond to products with or without the alternative exon (lower case sequence in Figure 2) respectively. We could find no evidence for the use of an alternative exon in the human or rat M₁ genes, although the positions of the splice donor site of exon 1 and the splice acceptor site of the coding exon appeared to be conserved. The alternatively spliced exon found in the mouse gene was 84% (78 bp out of 93 bp) identical to the equivalent region in the pig cDNA. It is not known whether this exon is alternatively or constitutively spliced in the pig M₁ gene. PCR of genomic DNA has shown that both the human and rat genomes contain the sequence for this alternatively spliced exon which is located approx. 500 bp 3' of exon 1. Sequence analysis of these PCR products showed that although the sequence of this exon was very similar between these species, a deletion of 11 pyrimidines in the splice acceptor site had occurred in the human and rat M₁ genes (results not shown), and this may explain why the splice acceptor site does not appear to be used in these species. There

Figure 2 Sequence comparison of the M_1 5' untranslated region from human, mouse, rat and pig

Sequences for the human, mouse and rat genes were determined experimentally using the cloned PCR products shown in Figure 1, whilst the pig sequence has been reported previously [30], accession number J04192. Regions of sequence homology between the species are boxed, the vertical dashed lines represent the positions of splice junctions, and the sequence of the alternative exon is shown in lower case. The coding sequence is shown in bold, and the positions of the primers used in PCR analysis are denoted by arrows.

have been no reported functional differences between the mouse M_1 receptor and the human or rat M_1 receptor and the consequences, if any, of the alternatively spliced product in mouse remains unclear. We were unable to amplify human and mouse exon 1 sequences using rat primers further 5' to +299, suggesting that the remainder of this exon may not be as highly conserved.

A 974 bp region of the rat M_1 gene was capable of driving reporter gene expression

To identify the genomic *cis* elements that were responsible for driving cell specific M_1 expression we cloned a region of the rat gene between -372 and +602, upstream of a luciferase gene, and assayed for its ability to drive reporter gene expression. This plasmid, pGL3 -372/+602, drives luciferase expression 4.9-fold relative to the empty vector in IMR32 cells (Figure 3), but does not drive expression in the non- M_1 -expressing neuronal lines, Neuro 2a and NB4 1A3 (0.3- and 0.4-fold respectively). Some expression is seen in 3T3 cells (2-fold) although not to the extent seen in IMR32 cells. Like the two previously characterized muscarinic receptor genes, the rat M_4 [14,15] and chicken M_2 [16], the rat M_1 gene contains a consensus binding site for the transcription factor Sp1, near the transcription initiation site (GC box between positions +17 and +23 of Pepitoni et al. [13]). To determine whether this Sp1 site is important for M_1 expression it was deleted to produce the construct pGL3 -372/+602 Δ Sp1 which lacks nucleotides +17 to +23 of the M_1 gene. The levels of luciferase activity driven by this construct were not significantly different from the levels driven by pGL3 -372/+602, suggesting that this GC box is not involved in regulating expression within

these cells. Although deletion of the GC box shows no functional significance in the cell lines used we have isolated two cDNA clones encoding Sp1 in a yeast one-hybrid screen to identify transcription factors that bind to the M_1 promoter. Double stranded oligonucleotides corresponding to nucleotides +10 to +28 of the M_1 promoter were radioactively labelled and used in a gel mobility shift assay with protein extracts from IMR32 and 3T3 cells. Protein extracts from IMR32 cells showed a protein-DNA complex (labelled A in Figure 4) that was competed by an excess of unlabelled probe [Figure 4 (upper panel), lane 2] but not by an unrelated sequence [Figure 4 (upper panel), lane 3], and the complex was unaffected by an antibody to Sp1 [Figure 4 (upper panel), lane 4]. 3T3 protein extracts produced a DNA-protein complex that migrated more slowly than that formed with IMR32 extracts [Figure 4 (upper panel), lane 5]. This complex was specific as it was competed by an excess of unlabelled probe [Figure 4 (upper panel), lane 6], but not by an unrelated sequence [Figure 4 (upper panel), lane 7]. Inclusion of an antibody to Sp1 resulted in a supershifted band and a reduction in the intensity of the signal from the DNA-protein complex [Figure 4 (upper panel), lane 8]. We are not sure why the Sp1 antibody reduced the signal of the DNA-protein complex; however, it is a polyclonal antibody and thus may have formed large complexes which were unable to migrate through the gel. We have observed this phenomenon before using the same antibody [21]. As the antibody was specific for Sp1 and had an effect on the DNA-protein complex formed with 3T3 protein extracts but not with that formed by IMR32, we concluded that the DNA-protein complex formed with 3T3 extracts contained Sp1. To identify the molecular mass of proteins that interacted with this region a UV crosslinking experiment was performed. The results of this

Figure 3 Expression of luciferase reporter gene constructs in transient transfections of cell lines

(a) A representation of the M₁ promoter region used in transfection analysis. The position of the transcription start site is indicated by the arrow and the positions of the Sp1 site, the polypyrimidine/polypurine tract (PPY) and the region highly conserved across species (Con) are shown. The vertical lines delimit reporter constructs used in the deletion series, and the positions of the terminal nucleotides are given. (b) All data are normalized to *Renilla* luciferase expression driven by co-transfected pRL-CMV (Promega) and promoter activity is expressed as, fold over luciferase activity obtained with the empty reporter vector, pGL3 Basic. Data represent the mean of at least three independent experiments each performed in triplicate and the error bars represent the S.E.M. The reporter constructs are named with respect to the 5'-most transcription initiation site from Pepitoni et al. [13] being defined as +1.

experiment showed that the M₁ promoter region interacted with a 25 kDa protein from IMR32 cells and a 100 kDa protein from 3T3 cells [Figure 4 (lower panel), lanes 1 and 3]. Both these interactions were specific as they could be competed by an excess of unlabelled probe [Figure 4 (lower panel), lanes 2 and 4]. The size of the interacting protein identified from 3T3 cells was con-

sistent with it being Sp1. Deletion of a region between +166 and +602 within the upstream exon to give pGL3 -372/+166 resulted in the loss of luciferase expression in each cell line tested (Figure 3). Analysis of the DNA sequence within this 436bp region did not identify any consensus elements for DNA-binding proteins although there was a stretch of pyrimidines between

Figure 4 Gel mobility shift assay and SDS/PAGE analysis of proteins binding to the region of the M_1 gene between nucleotides +10 and +30

Upper panel: labelled DNA was incubated with nuclear protein isolated from cell lines in the presence of no competitor, unlabelled excess probe (m1 Sp1 competitor), oligonucleotides containing an unrelated sequence (Non specific competitor), or an antibody to the Sp1 protein (Sp1 antibody). Arrows labelled A and Sp1 indicate positions of specific protein-DNA complexes. Lanes are labelled (1-8) starting from the left-hand side. Lower panel: labelled DNA was incubated with nuclear protein from cell lines in the presence of either no competitor or excess unlabelled probe. After UV crosslinking, the protein-DNA complexes were subjected to SDS/PAGE (6% gel). The positions of molecular mass standards are shown on the left. Arrows labelled A and Sp1 show the positions of specifically bound proteins. Lanes are labelled (1-4) starting from the left-hand side.

nucleotides +412 and +485. Stretches of pyrimidines have been shown to be important in transcriptional regulation of several genes including chicken malic enzyme [31], vascular smooth muscle α -actin [32,33], hEGFR [34], *c-ets-2* [35], and c-Ki-Ras [36]. To functionally test the regions within +166 to +602, a series of constructs were produced containing discrete deletions within this region, and their ability to drive luciferase expression was analysed. The results are shown in Figure 3. Deletion of the region between +241 and +602 to give pGL3 -372/+241 resulted in the total loss of luciferase expression in all cell lines.

Deletion of the polypyrimidine/polypurine tract between nucleotides +415 and +485 to give pGL3 -372/+602 Δ PPY resulted in a substantial loss of expression in IMR32 cells. Deletion of the region between +504 and +602 to give pGL3 -372/+504 resulted in a similar attenuation of luciferase activity as the deletion of the polypyrimidine/polypurine tract. Simultaneous deletion of both these regions to give pGL3 -372/+417 did not give a further loss of luciferase activity suggesting that the two regions +415 to +485 and +504 to +602 acted in concert to enhance expression of the M_1 gene. Although pGL3 -372/+417 showed reduced expression compared to pGL3 -372/+602 it was not as low as that seen for either pGL3 -372/+166 or pGL3 -372/+241. A further deletion was made leading to construct pGL3 -372/+366 which resulted in an increase of luciferase expression in both IMR32 and 3T3 cells. The expression in IMR32 cells was similar to that driven by pGL3 -372/+602; however, the expression in 3T3 cells was greater, suggesting that the region between +366 and +417 contained a repressor element. Deletion of a further 50 nucleotides to give pGL3 -372/+316 resulted in the loss of this non-specific activity, although this construct still drove higher levels of luciferase expression than seen with pGL3 -372/+241, suggesting that the region between +241 and +316 could act as an enhancer. None of the constructs showed appreciable activity in Neuro 2a or NB4 1A3 cells. In summary, deletion analysis of the M_1 promoter has identified two regions, the polypyrimidine/polypurine tract between +412 and +485 and a region between +504 and +602, conserved across species, which acted in concert to enhance expression in IMR32 cells. The region between +311 and +366 could act as a potent enhancer in both IMR32 and 3T3 cells but this activity was repressed by the presence of the region +366 to +417. Expression of the reporter genes was greater in the non-neuronal 3T3 cells than for either of the neuronal cells, Neuro 2a or NB4 1A3, suggesting that different mechanisms may be responsible for regulating M_1 expression in these different cell types.

Regions of the non-coding exon can activate a heterologous promoter

As all of the deletions shown in Figure 3 that affect expression of the reporter construct occur downstream of the transcription start site, we investigated whether these deletions were affecting gene transcription or mRNA stability. In order to test this we juxtaposed two regions of the M_1 non-coding exon to a heterologous promoter. For the heterologous promoter we used the adenovirus major late promoter which like the M_1 gene lacks a TATA box [23]. Figure 5 shows the levels of luciferase expression driven by these constructs. A region of the DNA corresponding to +166/+602 was cloned upstream of the initiator from the adenovirus major late promoter, to generate pGL3 +166/+602 Inr. This corresponds to the region that was deleted in pGL3 -372/+166 and caused loss of reporter gene expression in all the cell lines tested (see Figure 3). As can be seen in Figure 5, pGL3 +166/+602 Inr activated transcription 8.8-fold in IMR32 cells but did not drive expression in Neuro 2a or NB4 1A3 cells. In fact Neuro 2a and NB4 1A3 showed a 2-fold repression. As in the case of pGL3 -372/+602, transcription was also activated in 3T3 cells but to a much lesser extent (2.2-fold). A further construct was made consisting of the polypyrimidine/polypurine tract and the conserved region between +396 and +569. In this case transcription from the adenovirus major late promoter was strongly activated in IMR32 cells (24.5-fold) and more weakly in 3T3, Neuro 2a and NB4 1A3 cells (6.9-, 2.9- and 2.9-fold

Figure 5 Both the +166/+602 and +396/+569 regions of the M₁ non-coding exon can activate expression driven by a heterologous promoter

All data are normalized to *Renilla* luciferase expression driven by co-transfected pRL-CMV, and promoter activity is expressed as fold over luciferase activity obtained with the reporter vector, pGL3 Inr. Data represent the mean of at least three independent experiments each performed in triplicate.

Figure 6 Gel mobility shift assays and Southwestern analysis of the M₁ enhancer regions

Single-stranded labelled DNA corresponding to sequences (a) +415 to +485 (pPy), (b) +486 to +569 (489s) and (c) +569 to +486 (572a) was incubated with nuclear protein isolated from cell lines in the presence of no competitor (no comp.), unlabelled excess probe (sp. comp.) or oligonucleotides containing an unrelated sequence (non-sp. comp.). Arrows indicate the position of specific DNA-protein complexes. (d-f) Nuclear proteins were subjected to SDS/PAGE (6% gel) and electroblotted on to nylon membranes. After renaturation of the proteins the blot was incubated with the same probes used for the gel mobility shift assays as described in the Materials and methods section. Numbers on the left-hand side correspond to the position of migration of molecular mass standards and are shown in kDa.

respectively), consistent with the deletion analysis of this region within the M₁ promoter.

The conserved region and the polypyrimidine tract both bind single-stranded DNA-binding proteins

Gel mobility shift assays were performed with regions of the M₁

promoter that we had found to be functionally important for M₁ expression. A single-stranded probe consisting of the polypyrimidine tract was found to bind proteins from both IMR32 and 3T3 cells (Figure 6a). This binding was specific for the single-stranded probe and a probe synthesized from the antisense sequence did not show any DNA-protein interactions (results not shown). Sequences from both strands of the conserved region

between nucleotides +486 and +569 were found to form specific single-stranded-DNA-protein interactions. Gel mobility shift assays with each of the probes showed DNA-protein complexes with both IMR32 and 3T3 cell extracts even though they only appear to enhance expression in IMR32 cells (Figures 6b and 6c). A Southwestern analysis was performed to identify the sizes of the proteins interacting with the DNA probes from the different cell lines. The Southwestern analysis for each of the probes is shown in Figures 6(d), 6(e) and 6(f). For each of these probes, two hybridizing bands of approximately 105 and 95 kDa were identified with IMR32 protein extracts, and only a single band of approximately 105 kDa with protein extracts from 3T3 cells. Surprisingly each of the probes showed binding to the same molecular mass proteins in both of the cell lines, raising the possibility that they may be interacting with the same DNA-binding proteins. As freshly produced blots were used for each of the probes the observed signals could not be the result of any carryover or incomplete removal of a previously used probe. As yet we do not know whether the two identified bands obtained from IMR32 protein extracts were due to different proteins or modifications of the same protein. A UV crosslinking experiment, using the polypyrimidine tract probe, produced the same results and also showed that this binding can be competed by an excess of unlabelled oligonucleotide of the same sequence but not by an oligonucleotide containing an unrelated sequence (results not shown).

DISCUSSION

To date the gene structures for three of the family of muscarinic receptor genes have been identified: the rat M_4 [14,15], the chicken M_2 [16] and the rat M_1 genes [13]. All of these genes appear to have a similar gene structure consisting of a single coding exon and a 5' exon containing the 5' untranslated region. The rat M_4 gene has been reported to contain an alternatively spliced exon in the 5' untranslated region [15] and we have identified an alternatively spliced exon in a similar position of the mouse M_1 gene. The significance of this exon is unclear as we could not detect its use in either the human or the rat M_1 gene in the tissues examined. Interestingly, the sequence for this exon is present in both the human and rat genomes but the splice acceptor site lacks eleven of the pyrimidines that are present in the mouse, presumably resulting in this splice acceptor site being less amenable to splicing.

In the central and autonomic nervous systems, the expression patterns of the M_1 and the M_4 genes are very similar but not identical. In the autonomic ganglia, for example, both genes appear to be co-expressed in the majority of neurones [37,38] whereas in the striatum, over 80% of the neurones express M_1 and 40–50% express M_4 [17]. In addition, Bernard et al. [18] showed that 85% of the large striatal interneurons have M_1 and most co-express M_4 , whilst 90% of medium-sized neurones co-express M_1 and M_4 , resulting in populations of neurones that express one of the genes, both of the genes or neither. We were therefore interested to know whether transcription of these two genes was governed by similar or disparate mechanisms. Previously we, and others, identified some of the DNA elements responsible for cell specific expression of the M_4 gene [14,15,21,22]. We have now analysed the regions of the M_1 gene that are responsible for cell specific expression and compared these with the regions that were found to be important for M_4 expression. The results show that the M_1 gene is positively regulated by several elements within the 5' exon, since deletion of a region of the gene containing these elements (+241 to +602)

abolishes expression in the M_1 -expressing cell line, IMR32. None of the reporter constructs tested showed any appreciable luciferase activity in the non-expressing cell lines used; however, luciferase activity in 3T3 cells appeared greater than activity in Neuro 2a and NB4 1A3 cells. Indeed activity in the two neuronal cell lines, Neuro 2a and NB4 1A3, was found to be lower than that of the empty reporter plasmid, suggesting that these constructs may be actively silenced in these cell lines. This is similar to the M_4 gene which contains a constitutively active core promoter which is silenced, at least in non-neuronal cells, by the transcriptional repressor REST/NRSF [21,22].

Of the regions we have identified to be functionally important in regulating M_1 expression only one contains a recognizable motif, the polypyrimidine/polypurine tract, deletion of which results in the loss of approximately 60% of expression in M_1 -expressing cell lines. Polypyrimidine/polypurine tracts have been shown to be important in driving expression of several genes such as the chicken malic enzyme [31], vascular smooth muscle α -actin [32,33], hEGFR [34], *c-ets-2* [35], and c-Ki-Ras [36]. In the malic enzyme gene, the polypyrimidine/polypurine tract assumes different isoforms containing single-stranded and triplex structures and also interacts with both single-stranded and double-stranded DNA-binding proteins. The polypyrimidine/polypurine tract of the vascular smooth muscle α -actin gene binds in a single-stranded manner to both negative and positive acting factors and also interacts with double-stranded DNA-binding proteins [32,33]. The polypyrimidine tract within the M_1 gene binds in a single-stranded manner to proteins from both IMR32 and 3T3 cells (Figures 6a and 6d). Neither double-stranded probes nor the polypurine tract from the antisense strand show any interactions with proteins (results not shown), even though regions rich in purines have been shown to interact with DNA-binding proteins [32,33]. Furthermore the polypyrimidine/polypurine tract is able to drive expression of a heterologous promoter in IMR32 cells, and to a much lesser extent in 3T3 cells, and is capable of interacting with protein from these lines (Figures 6a and 6d). Southwestern analysis shows (Figure 6d) that whilst this region interacts with only a single protein in 3T3 cells, it interacts with two proteins (or alternative forms of the same protein) in IMR32 cells. It will be interesting to determine whether the unique protein (or isoform) present in IMR32 cells is responsible for enhanced activity of this DNA element. Interestingly, the polypyrimidine/polypurine tract within the M_1 gene is not present in the M_1 cDNA reported from pig [30], although this motif may still be present within the pig M_1 gene. The region adjacent to the polypyrimidine/polypurine tract contains a stretch of 100 nucleotides that is almost completely conserved between the human, mouse, rat and pig M_1 genes (Figure 2). We have not identified any consensus DNA-binding motifs within this region, but this region does bind to proteins in a single-stranded manner (Figures 6b, 6c, 6e and 6f). Unlike the polypyrimidine/polypurine tract, both strands of this region are capable of interacting with single-stranded DNA-binding proteins. Southwestern analysis showed that both strands of this region bound to proteins of the same molecular mass as those bound by the polypyrimidine tract (Figures 6d–6f). Although the proteins bound by the three probes appeared to be the same molecular mass we cannot, as yet, say if they are identical. These two elements appear to work in concert since the deletion of either one is equivalent to the deletion of both. As both regions bind to proteins in a single-stranded manner and polypyrimidine/polypurine tracts have been shown to adopt a single-stranded nature it is possible that the polypyrimidine/polypurine tract induces or stabilizes a single-stranded conformation in the conserved region.

One of the transcription factor binding sites that is present in all three of the muscarinic genes cloned so far is that of an Sp1 site. Both the M₄ gene and the M₂ gene contain three Sp1 sites [14,16]. A tandem Sp1 site in the M₄ promoter binds to several DNA-binding proteins one of which is present only in those regions of the brain in which M₄ is expressed [21]. Interaction of DNA-binding proteins with the Sp1 sites in the M₂ promoter has not been tested but we identified the Sp1 protein in a yeast one-hybrid library screen using a region of the M₁ promoter as bait. We have shown that the Sp1 site within the M₁ gene was capable of binding to Sp1 and at least one other DNA-binding protein (Figure 4). *In vitro* DNA-binding assays and UV crosslinking experiments (Figure 4) showed that in IMR32, the M₁-expressing neuroblastoma, the Sp1 binding site interacted with a protein of approx. 20–25 kDa that was not recognized by an antibody to Sp1. Sp1 protein present in extracts from 3T3 cells does bind to this region but no other proteins were identified. Sp1 has been implicated in being important for transcription from TATAless promoters, however, deletion of the Sp1 site does not appear to alter the expression of the M₁ gene in the cell lines tested (Figure 3, pGL3 –372/+602 ΔSp1). Whether this site will prove to be important in expression of the M₁ gene in different neuronal subpopulations, within the mature or developing nervous system, remains to be determined.

In summary, we have identified several regulatory domains, contained within exon 1, that are partially responsible for conferring cell specific expression of the M₁ gene. Distinct domains appear to be responsible for regulating expression in non-expressing non-neuronal and non-expressing neuronal cells. Further studies are aimed at identifying accompanying transcription factors, and at assessing the role of these elements in regulating transcription *in vivo*.

This work is supported by the Wellcome Trust. M.G.C. is a holder of a Wellcome Trust Prize studentship. We would like to thank Avtar Roopra and Peter McIntyre for their helpful comments on the manuscript.

REFERENCES

- Bonner, T. I., Buckley, N. J., Young, A. C. and Brann, M. R. (1987) *Science* **237**, 527–532
- Bonner, T. I., Young, A. C., Brann, M. R. and Buckley, N. J. (1988) *Neuron* **1**, 403–410
- Buckley, N. J., Bonner, T. I. and Brann, M. R. (1988) *J. Neuroscience* **8**, 4646–4652
- Hulme, E. C., Birdsall, N. J. M. and Buckley, N. J. (1990) *Ann. Rev. Pharmacol. Toxicol.* **30**, 633–673
- Peralta, E. G., Askenazi, A., Winslow, J. W., Ramachandran, J. and Capon, D. J. (1988) *Nature (London)* **334**, 434–437
- Conklin, B. R., Brann, M. R., Buckley, N. J., Ma, A. L., Bonner, T. I. and Axelrod, J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8698–8702
- Jones, S. V. P., Heilman, C. J. and Brann, M. (1991) *Mol. Pharmacol.* **40**, 242–247
- Fukuda, K., Higashida, H., Kubo, T., Maeda, A., Akiba, I., Bujo, H., Mishina, M. and Numa, S. (1988) *Nature (London)* **335**, 355–358
- Marion, N. V., Smart, T. G., Marsh, S. J. and Brown, D. A. (1989) *Br. J. Pharmacol.* **98**, 557–573
- Messer, W. S., Bohnett, M. and Stibbe, J. (1990) *Neuroscience Lett.* **116**, 184–189
- Messer, W. S., Stibbe, J. R. and Bohnett, M. (1991) *Brain Res.* **564**, 66–72
- Hamilton, S. E., Loose, M. D., Qi, M., Levey, A. I., Hille, B., McKnight, G. S., Idzerda, R. L. and Nathanson, N. M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13311–13316
- Pepitoni, S., Wood, I. C. and Buckley, N. J. (1997) *J. Biol. Chem.* **272**, 17112–17117
- Wood, I. C., Roopra, A., Harrington, C. and Buckley, N. J. (1995) *J. Biol. Chem.* **270**, 30933–30940
- Mieda, M., Haga, T. and Saffen, D. W. (1996) *J. Biol. Chem.* **271**, 5177–5182
- Rosoff, M. L., Wei, J. and Nathanson, N. M. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14889–14894
- Weiner, D. M., Levey, A. I. and Brann, M. R. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7050–7054
- Bernard, V., Normand, E. and Bloch, B. (1992) *J. Neuroscience* **12**, 3591–3600
- Schoenherr, C. J. and Anderson, D. J. (1995) *Science* **267**, 1360–1363
- Chong, J. A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J. J., Zheng, Y., Boutros, M. C., Altschuler, Y. M., Frohman, M. A., Kraner, S. D. and Mandel, G. (1995) *Cell* **80**, 949–957
- Wood, I. C., Roopra, A. and Buckley, N. J. (1996) *J. Biol. Chem.* **271**, 14221–14225
- Mieda, M., Haga, T. and Saffen, D. W. (1997) *J. Biol. Chem.* **272**, 5854–5860
- Rosoff, M. L. and Nathanson, N. M. (1998) *J. Biol. Chem.* **273**, 9124–9129
- Pengue, G. and Lania, L. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1015–1020
- Dent, C. L. and Latchman, D. S. (1993) in *Transcription Factors: A Practical Approach* (Latchman, D. S., ed.) pp. 1–26, IRL Press, Oxford
- Vinson, C. R., LaMarco, K. L., Johnson, P. F., Landschulz, W. H. and McKnight, S. L. (1988) *Genes Develop.* **2**, 801–806
- Cooney, A. J., Tsai, S. Y. and Tsai, M.-J. (1993) in *Transcription Factors: A Practical Approach* (Latchman, D. S., ed.) pp. 49–79, IRL Press, Oxford
- Liu, J., Wilson, T. E., Milbrandt, J. and Johnston, M. (1993) *METHODS: A Companion to Methods in Enzymology* **5**, 125–137
- Bartel, P. L., Chien, C.-T., Sternglanz, R. and Fields, S. (1993) *BioTechniques* **14**, 920–924
- Schiestl, R. H. and Gietz, R. D. (1989) *Curr. Genet.* **16**, 339–346
- Kubo, T., Fukuda, K., Mikami, A., Maeda, A., Takahashi, H., Mishina, M., Haga, T., Haga, K., Ichiyama, A., Kangawa, K., et al. (1986) *Nature (London)* **323**, 411–416
- Xu, G. and Goodridge, A. G. (1996) *J. Biol. Chem.* **271**, 16008–16019
- Sun, S., Stoffel, E. S., Cogan, J. G., Strauch, A. R. and Getz, M. J. (1995) *Mol. Cell Biol.* **15**, 2429–2436
- Cogan, J. G., Sun, S., Stoffel, E. S., Schmidt, L. J., Getz, M. J. and Strauch, A. R. (1995) *J. Biol. Chem.* **270**, 11310–11321
- Mavrothalassitis, G. J., Watson, D. K. and Papas, T. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1047–1051
- Johnson, A. C., Jinno, Y. and Merlino, G. (1988) *Mol. Cell Biol.* **8**, 4174–4184
- Hoffman, E. K., Trusko, S. P., Murphy, M. and George, D. L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2705–2709
- Hassal, C. J. S., Stanford, S. C., Burnstock, G. and Buckley, N. J. (1993) *Neuroscience* **56**, 1041–1048
- Brown, D. A., Buckley, N. J., Caulfield, M. C., Duffy, S. M., Jones, S., Lamas, J. A., Marsh, S. J., Robbins, J. and Selyanko, A. A. (1995) in *Molecular Mechanisms of Muscarinic Acetylcholine Function* (Wess, J., ed.) pp. 165–182, R. G. Landes Company, Georgetown

Received 10 December 1998/25 February 1999; accepted 19 March 1999