Structure of the m1 Muscarinic Acetylcholine Receptor Gene and Its Promoter*

(Received for publication, March 11, 1997)

Stefania Pepitoni, Ian C. Wood, and Noel J. Buckley‡
From the Wellcome Laboratory for Molecular Pharmacology, Department of Pharmacology, University College London, London WC1E 6BT, United Kingdom

The m1 receptor is one of five muscarinic receptors that mediate the metabotropic actions of acetylcholine in the nervous system where it is expressed predominantly in the telencephalon and autonomic ganglia. RNase protection, primer extension, and 5'-rapid amplification of cDNA ends analysis of a rat cosmid clone containing the entire m1 gene demonstrated that the rat m1 gene consists of a single 657-base pairs (bp) non-coding exon separated by a 13.5-kilobase (kb) intron from a 2.54-kb coding exon that contains the entire open reading frame. The splice acceptor for the coding exon starting at −71 bp relative to the adenine of the initiation methionine. This genomic structure is similar to that of the m4 gene (Wood, I. C., Roopra, A., Harrington, C. A., and Buckley, N. J. (1995) J. Biol. Chem. 270, 30933–30940 and Wood, I. C., Roopra, A., and Buckley, N. J. (1996) J. Biol. Chem. 271, 14221–14225). Like the m4 gene, the m1 promoter lacks TATA and CAAT consensus motifs, and the first exon and 5'-flanking region are not germinally related. The 5'-flanking region also contains the consensus regulatory elements Sp-1, NZF-1, AP-1, AP-2, E-box, NFκB, and Oct-1. Unlike the m4 promoter, there is no evidence of a RE1/NRSE silencer element in the m1 promoter. Deletional analysis and transient transfection assays demonstrate that reporter constructs containing 0.9 kb of 5'-flanking sequence and the first exon are sufficient to drive cell-specific expression of reporter gene in IMR32 neuroblastoma cells while remaining silent in 3T3 fibroblasts.

G-protein coupled receptors are responsible for mediating a vast amount of intercellular communication throughout the body, especially in the nervous system. Current estimates put the size of this gene superfamily well in excess of 1000, making it one of the largest in the mammalian genome. In situ hybridization studies indicate that each individual member probably has a unique expression profile within the nervous system, yet the factors that determine and direct the expression patterns of the members of this gene family are largely unknown. Since the response of any neuron to a neurotransmitter is determined by the repertoire of receptors expressed at the cell surface, then it is essential to understand the mechanisms that determine the types of receptor gene expressed by individual neurons. The five muscarinic receptors are encoded by a subfamily of this gene superfamily (1, 2), and their gene products are responsible for mediating the metabotropic actions of acetylcholine in the nervous system and its effector tissues (3). Each of the five muscarinic receptor genes is differentially expressed throughout the central and autonomic nervous systems both in adulthood (4, 5) and during embryonic development, and each of the receptors exhibits a unique pharmacological profile (6, 7). The m1 receptor is the most abundant subtype in both the central and autonomic nervous systems and is found predominantly in the telencephalon (4), autonomic ganglion cells (4, 8, 9), and exocrine tissue (10). Activation of the m1 receptor leads to numerous responses, including stimulation of phospholipases C (7) and A2 (11), inhibition of cAMP production (5), activation of K+ and Cl- channels (12), and inhibition of opening of K+ channels in neural cell lines, and sympathetic and hippocampal neurons (13, 14). We have previously described the structure of the rat m4 gene and its promoter (5, 15–17) and have shown that expression is silenced in non-neuronal cells (16) via a RE1/NRSE (18, 19) type silencing element. There are many locations, such as the cerebral cortex, hippocampus, striatum, and autonomic ganglia, that co-express m1 and m4 receptor genes, and equally important in view of the silencing of the m4 promoter, there are many locations that express neither the m1 nor m4 genes, such as most regions of the mesencephalon and rhombencephalon and most non-neuronal tissue. Yet even within areas that co-express m1 and m4, not all individual cells express both m1 and m4 genes. There is, therefore, an intimate matrix of overlapping and nonoverlapping expression profiles between these homologous family members. We were thus interested to identify the control regions of the m1 promoter to ascertain if the m1 and m4 genes shared a common gene structure and regulatory elements. As a first step to addressing this issue, in the present study, we present the first description and analysis of the structure of the m1 muscarinic receptor gene and its promoter.

MATERIALS AND METHODS

Cosmid Library Screening—A rat cosmid library in pWE15 (Stratagene) was screened using a mixture of three [α-32P]dATP tailed coding region oligodeoxynucleotides (RmLA, GGCACCCGTCCTCAGGGGACCCTT- TCCGGTTGCCAACAGACGTAGTGTG; RmLB, ACTCAGGGTCCG- AGCTGCTTTCCTCCTCCTGACAGTGAGTCTT, and Rm1C, GG- GGCCTTGGGAGCTTCTGGCCACCGCTCTTTGTCCAGGCGAG- AGC). Filters were hybridized overnight at 37 °C in buffer containing 4 × saline/sodium/phosphate/EDTA (150 mM NaCl, 10 mM NaH2PO4, 1 mM EDTA), 5 × Denhardt's reagent (0.1% Ficol, 0.1% bovine serum albumin, 0.1% polyvinylpyrrolidone), 250 mg/ml yeast RNA, 500 mg/ml denatured salmon sperm DNA, and 0.1% SDS (1 × 106 dpm/ml) before washing at 55 °C in 1 × saline/sodium/phosphate/EDTA, 0.1% SDS, four times for 15 min each. Filters were then exposed to x-ray film at −70 °C for a week. Positive colonies were mapped using probes based upon RACE5 sequence of the upstream exon.

* This work was supported by the Wellcome Trust. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Wellcome Laboratory for Molecular Pharmacology, Dept. of Pharmacology, University College London, London WC1E 6BT, United Kingdom. Tel.: 44 0 171 380 7244; Fax: 44 0 171 380 7245; E-mail: n.buckley@ucl.ac.uk.
Reverse Transcription-PCR—Reverse transcription-PCR analyses were carried out as described previously (16). RNA was treated with RNasefree DNase I (Promega) prior to reverse transcription to ensure that subsequent amplification products were derived from RNA and not contaminating genomic DNA. PCR cycling conditions were (95 °C/30 s, 55 °C/30 s, and 72 °C/30 s) for 30 cycles, and the reaction was allowed to proceed at 42 °C for 5 min followed by a period of 72 °C/10 min. Resulting amplified products were analyzed on a 2% agarose gel.

RESULTS

Organization of the Rat m1 Gene—Screening of 10^9 recombinants in a rat cosmid library with coding region primers yielded a single colony, S1–139. The same colony hybridized to oligonucleotides derived from 5′-RACE cDNA sequence.

Since repeated screening of cDNA libraries failed to generate any significant upstream cDNA sequence information, 5′-RACE was used to generate clones corresponding to the 5′-untranslated region of the m1 cDNA. Use of AMV-RT generated clones that routinely terminated less than 60 bp upstream of the splice acceptor site. Supplementation of the PCR with dimethyl sulfoxide, formamide, or glyceral failed to yield any longer clones, and neither did the use of MeHgOH to denature the RNA prior to reverse transcription. However, use of the thermostable polymerase Tth (Promega) for the reverse transcription yielded several longer clones, which were used to generate the underlined sequence seen in Fig. 1. A long poly-proximate tract stretches for about 88 bp between positions +415 and +503 (see Fig. 1), and this may serve as a premature transcriptional stop during reverse transcription. Interestingly, around this point, the sequence of the rat 5′-untranslated region diverges sharply from its porcine homologue (22).

Homology downstream from this point is very high (88% over 294 bases to the initiation codon although in the porcine sequence, the consensus splice site lies further upstream than 361 with respect to the initiation codon (22). Both the position of the splice site and homology of the coding exon are highly conserved between rat and mouse with the exception of a single base change - CCTTCTTTTCTAGGAG (23).

Before proceeding to a finer analysis of the promoter of the m1 gene, we first wished to establish whether the S1–239 codon contained sufficient information to direct expression of the m1 gene by transfecting the entire cosmids into the human m1 expressing neuroblastoma, NBOK1. Reverse transcription-PCR, using primers derived from noncoding and coding exons specific for the rat m1 sequence, revealed the presence of rat m1 transcripts in the transfecants (data not shown), thus verifying that the cosmid contained all information necessary to drive expression, at least in transient transfection assays. No rat m1 transcripts were detected in untransfected NBOK1 cells.

Identification of the Transcriptional Start Site—Two independent strategies were used to identify the position of the transcription start site of the m1 gene. Initially, primer extensions were performed on rat brain cortex poly(A)+ RNA using the two primers, 132a and 235a (see Fig. 1). Despite using several protocols and different reverse transcriptases, no prod-
**FIG. 1.** Organization of the rat m1 muscarinic receptor gene. Panel A shows the relative sizes and positions of the coding (Ex2) and non-coding (Ex1) exons on a fragment of the S1–239 cosmid clone while panel B shows an enlargement of the 5' flanking sequence. B, BamHI; E, EcoRI; H, HindIII; K, KpnI; N, NcoI; P, PstI; S, SalI; Sa, SacI. Panel C shows the sequence of 1.6 kb of 5' flanking sequence and the 657-bp noncoding exon. Positions of HindIII, KpnI, EcoRI, PstI, and SacI sites are indicated. The position of the ΔH/P1.8 deletion is also indicated. Positions of consensus regulatory elements Sp-1, NZF-1, AP-1, AP-2, E-box, NFκB, and Oct-1 are boxed. Dashed underlines indicate primers used in PCR analysis. Transcription start sites identified by RNase protection are indicated by asterisk; the two guanine residues representing the m1 Muscarinic Promoter 17114
tion initiation sites of and 186 nucleotides were observed, corresponding to transcrip-
...double underline
...extended product corresponding to position...conjunction with AMV reverse transcriptase, the length of this...obtain a 171-nucleteotide product using the 235a primer in...We did, however,...for the slower migration of RNA compared with DNA.

FIG. 2. Identification of transcription start site. Map of a region of the m1 gene between −678 and +656 showing the positions of the 3 antisense RNA probes used in RNase protection experiments is shown at top. Bottom panel shows RNase protection analysis of the m1 gene. RNA probes were internally labeled with [α-32P]UTP according to standard procedures. Each of the labeled RNA probes were hybridized to 10 μg of rat brain cortex RNA (lanes 1, 5, and 9), 100 μg of rat brain cortex RNA (lanes 2, 6, and 10), 10 μg of tRNA (lanes 3, 7, and 11), or 100 μg of tRNA. The positions of major protected bands and their size in nucleotides are indicated by arrows on the left. Sizes have been corrected for the slower migration of RNA compared with DNA.

FIG. 3. PCR analysis of genomic DNA and cDNA. PCR analysis of rat brain cortex cDNA (a), m1 genomic clone, H/P-1.4/+0.6pGL3 (b), or no template (c) using a selection of m1 primers as indicated. PCR products were electrophoresed on a 2% metaphor gel, 1-kb DNA markers. Templates are rat brain cortex cDNA (a), H/P-1/4/+0.6pGL3 plasmid DNA (b), and water (c). Primer pairs used for amplification are indicated.

To further verify that transcription of the m1 gene initiates at +1 and the 1st exon does not contain any introns, a comparative PCR between cDNA from brain cortex and genomic DNA was performed. cDNA was generated from rat brain cortex RNA that previously had been treated with DNase to remove genomic DNA. Two different sense primers were used, 696s, which contains sequence upstream of the proposed transcription start site, and 539s, which contains sequence downstream of the proposed transcription start site. Each of the sense primers was used in PCR in conjunction with the three antisense primers 461a, 235a, and 132a (the positions of each of the primers is shown in Fig. 1). The PCR products obtained from this analysis are shown in Fig. 3. Primer pairs 539s/46a, 539s/235a, and 539s/132a generated 71-, 303-, and 412-bp amplified products, respectively, from both H/P-1.4/+0.6pGL3 and cDNA, indicating that the intervening sequence is exonic and contains no introns. However, primer pairs 696s/46a, 696s/235a, and 696s/132a generated 231-, 461-, and 568-bp amplified products only when H/P-1.4/+0.6pGL3 was used as template. No signal was seen using cDNA as template, thereby indicating primer 696s must lie upstream of the transcription initiation site. These data are consistent with the guanine residues at positions +1 and +3 being the dominant transcription initiation sites. Furthermore, the sizes of the PCR products obtained using 539s with each of the antisense primers is the same in cDNA and genomic DNA, indicating that there are no introns between 539s and 132a.

Sequence Analysis of the 5'-Flanking Region of the m1 Gene—Inspection of the sequence of the upstream exon and 5'-flanking sequence reveals consensus binding elements for AP-1, NZF-1 (24), AP-2, Oct-1, and NFκB. No TATA or CAAT consensus elements are present. The sequence flanking the transcription start site show no homology with any known initiator sequence. No significant homology with the 5'-flanking region or the promoter of the m4 gene is found, nor is it...
found with any other sequence in the database.

Expression of Promoter-Reporter Constructs in Cell Lines—
Reverse transcription-PCR analysis demonstrated the presence of m1 transcripts in IMR32 cells and their absence in 3T3 fibroblasts (Fig. 4). The four reporter constructs (K/P-1.6/+0.6pGL3, H/P-1.4/+0.6pGL3, ΔH/P-1.3/+0.6pGL3, and S/P-0.88/+0.6pGL3) used to assay the promoter activity in these cells all terminated at the PstI site 65 bp upstream of the splice site and started 1.60, 1.39, 1.26, and 0.88 kb upstream of the transcriptional start site, respectively. All four constructs expressed 4–5-fold above background in m1-expressing IMR32 cells, and only the larger K/P-1.6/+0.6pGL3 construct drove expression significantly above background in the nonexpressing 3T3 fibroblasts.

DISCUSSION

G-protein coupled receptors are a diverse and widely expressed family of receptors that mediate signaling throughout the body both in development and adulthood. Within the nervous system, they represent one of the most significant sources of phenotypic diversity, yet little is known of the factors and mechanisms that regulate this cell-specific expression. As such, understanding the mechanisms governing the transcriptional regulation of members of this gene family can offer insight into the establishment and maintenance of specific patterns of gene expression within the nervous system. In the present study, we have shown that, in common with several other members of the G-protein-coupled receptor gene family, including the m4 (15, 16, 17, 24), V1a vasopressin (25), D1a dopamine (26, 27), and C5a (28) receptor genes, the m1 muscarinic receptor gene consists of a single coding exon and a single noncoding exon. Another feature shared with most, but by no means all, other G-protein-coupled receptor genes whose promoters have been examined for genes is the absence of TATA, CAAT, or initiator consensus elements. Examples include the promoters of the 5HT1a (29), 5HT2a (30), 5HT1c (31), serotonin receptors, V1a vasopressin receptor (25), D2 and D1a receptors (26, 27, 32), SSTR1 somatostatin receptor (33), and NPY-1 receptor (34).

Inspection of 1.6 kb of 5′-flanking sequence revealed several consensus regulatory elements including one AP-1 site, two AP-2 sites, two NFκB sites, an E-box and an NZF-1 element. The latter is an element recognized by a zinc finger protein that is expressed in the developing nervous system (24). Since it has been shown that two neuronal proteins, one NFκB-like and one distinct from NFκB (BETA) (35) can interact with the NFκB recognition sequence and activate transcription from the proenkephalin and HIV promoters (36), it will be interesting to ablate these sites and monitor the effect on expression of the m1 gene. As with the m4 promoter, no CRE elements are found in the proximal promoter. The existence of a 88-bp polypyrimidine/polypurine tract in the noncoding exon between +415 and +503 is intriguing in light of studies on other promoters such as the malic enzyme (37), EGF receptor (38), and the mouse c-Ki-ras (39), which have shown deletion of such tracts to decrease promoter activity. The role of the polypyrimidine/polypurine tract in transcription of the m1 muscarinic receptor gene remains to be examined.

In our earlier studies, we have shown that the core promoter of the m4 muscarinic receptor gene is constitutively active and cell-specific expression is achieved by silencing expression in non-neuronal (15, 16, 40) via a RE1/NRSE-type silencing element (18, 19). Interestingly, inspection of 2.5 kb of flanking sequence of the m1 promoter reveals no RE1/NRSE element. This observation is corroborated by the failure of a radiolabeled single-stranded RE1/NRSE oligodeoxynucleotide to hybridize to a digest of the m1 cosm id under conditions that generate a strong hybridization signal to digests of the R3–6 m4 cosm id (data not shown). Hence, unlike its m4 counterpart, it is unlikely that the m1 gene is under the control of the zinc finger silencer REST/NRSF (41, 42).

A recent report describes the promoter of
Deletional analysis revealed that constructs containing between 0.6 and 1.4 kb of 5’-flanking sequence and the entire noncoding exon were sufficient to drive reporter gene expression in IMR32 cells, a neuroblastoma that expresses an endogenous m1 gene. This cell line expresses more m1 mRNA than any other cell line that we have thus far screened, but even so, its level of expression compared with rat cerebral cortex is low (see Fig. 4), at least as judged by reverse transcription-PCR. All reporter constructs drove 4–5-fold luciferase expression relative to the promoterless vector, pGL3 Basic (Fig. 5). This modest stimulation of reporter gene activity is presumably a reflection of the relatively low levels of endogenous m1 expression in IMR32 cells. Only the Kp I-1.6/+0.6pGL3 reporter construct drove reporter gene expression in 3T3 fibroblasts, which express no endogenous m1 receptor, showing that constructs containing as little as 0.88 kb of 5’-flank and the non-coding exon are sufficient to drive cell-specific expression, at least in transient transfection assays. The low level of expression driven by the Kp I-1.6/+0.6pGL3 construct may indicate the presence of a weak non-neuronal activator between the Kp I site (−1.6 kb) and the HindIII site (−1.4 kb). Since these deletions ablate the E-box, AP-1, and the distal NF1 and NFkB sites, then it is clear that they are not necessary for cell-specific expression. The role, if any, of the NF1-1, NFkB, and AP-2 sites between the Sac I site and the transcriptional initiation site await determination by a finer analysis of the proximal promoter. However, interpretation of all transient transfection assays is limited by the fact that reporter gene expression is driven by multiple episomal copies of the reporter vector. Consequently, there are numerous examples of reporter constructs that are capable of driving apparent cell-specific expression in transient transfections that nevertheless fail to recapitulate appropriate cell and/or stage-specific expression in transgenic mice, as in the case of the dopamine β-hydroxylase gene where reporter constructs containing 0.6 kb of 5’-flanking sequence can drive cell-specific expression in transient transfection assays (44, 45) but give no expression in transgenic mice (46). Reporter gene expression of other G-protein-coupled receptor promoters has revealed that less than 1 kb of 5’-flanking sequence is sufficient to drive cell-specific expression of the V1a vasopressin (25), D1 dopamine (27), type-I angiotensin II (47), and m4 muscarinic (15, 16, 17) receptor promoter constructs in transient transfection assays. However, whether such discrete constructs are capable of driving tissue- and stage-specific expression in transgenic mice has not been reported for any members of the G-protein-coupled receptor gene family.

The characterization of two muscarinic receptor promoters now enables us to examine the differential transcriptional regulation of these members of the G-protein-coupled receptor gene family. Future studies are aimed at dissecting the m1 proximal promoter to determine whether the core promoter is constitutively active, as in the case of the m4 promoter, or whether progressive deletions lead to ablation of expression in m1-expressing cells.

Acknowledgment—We thank Dr. Magali Waelbroeck (Université Libre de Bruxelles) for the NBOK1 cells.