Control of Neurone-Specific Gene Expression:
Transcriptional Regulation of the
M₁ Muscarinic Acetylcholine Receptor Gene

By

Mireia Garriga-Canut B.Sc.

A thesis presented for the degree of
Doctor of Philosophy in the
University of London

Wellcome Laboratory for Molecular Pharmacology
Department of Pharmacology
University College London
ABSTRACT

The establishment of differentiated neuronal phenotype remains an outstanding problem of molecular neurobiology. One of the clearest manifestations of this molecular diversity is provided by the G-protein coupled receptor (GPCR) family. Each of the more than 1,000 members of this gene family has a unique expression profile and thereby offer an ideal model to examine the transcriptional mechanisms that underwrite this molecular diversity.

Muscarinic acetylcholine receptors (M₁ - M₅) make up one of the subfamilies of GPCR genes. The M₁ gene is the most abundant of the muscarinic receptor genes and is mainly expressed in telencephalic regions and autonomic ganglia. I have now investigated the regions of this gene that are capable of driving expression of a reporter gene in an M₁ specific manner. One of these regions is a polypyrimidine/polypurine (PPY/PPU) sequence capable of forming single stranded DNA, and the other (found downstream of the PPY/PPU tract) is conserved across species, has no recognisable motifs and is not able to form single stranded DNA by itself although it shows sensitivity to specific single stranded nucleases when next to the PPY/PPU tract. Both the PPY/PPU tract and the conserved region are bound by nucleolin, a multifunctional phosphoprotein, and act as cis-enhancing elements.

A second region important for expression of the M₁ gene has been identified to be bound by SHARP-1, a basic helix loop helix protein of unknown function expressed in the adult nervous system. Gal4 fusion experiments have shown that SHARP-1 functions as a repressor of both basal and activated transcription driven either by a TATA-containing or a TATA-less promoter in a position independent manner. Furthermore, SHARP-1 contains two independent repression domains, one at the C-terminus, which acts by a mechanism sensitive to the histone deacetylase inhibitor Trichostatin A (TSA), and the other at the bHLH domain, which works through a TSA insensitive mechanism. Co-transfection assays showed that SHARP-1 downregulates expression of the M₁ gene in M₁ expressing cells.

Data presented here shows that the transcriptional mechanisms that control expression of the M₁ gene are different to those that control expression of the other members of the same family. These results provide insight into the molecules and mechanisms employed in the establishment of aspects of differentiated neuronal phenotype.
ACKNOWLEDGEMENTS

I would like to thank the Wellcome Trust and my supervisor Professor Noel J. Buckley for letting me carry out this project in this lab. I would especially like to thank Drs Ian C. Wood and Avtar Roopra for lending guidance through the more difficult parts of the project. I hope they find somebody else to laugh at when I leave. It was fun for me, too. I would also like to thank the whole group both in London and Leeds for being so patient with me and making my time in the lab so enjoyable.

I would like to dedicate this work to my family, especially to my mum for letting me go away and my sister for pushing me to go away. A special thanks to Miriam and all those friends in Barcelona, like Cristina, Gemma, Montse, Oskar, Manoli and Fredy, who although being very far have kept in touch. Finally, I would also like to thank all those new friends that I have met in England, like Ana, Arantza, Elena, Unai, Koji, Louise, Hasanthi, Rav, David, Veronique, Damien and Tim for being so supportive during all these years and making my time here very enjoyable.
LIST OF FIGURES

CHAPTER 3

1 - Organisation of the rat M$_i$ muscarinic receptor gene. 54
2 – Expression of the M$_i$ gene by different cell lines. 57
3 – Deletional analysis of the rat M$_i$ gene. 60
4 - Binding of nuclear proteins to the M$_i$ GC box. 62
5 - Sp1 crosslinking. 65

CHAPTER 4

6 – Deletional analysis within the non-coding exon of the M$_i$ gene. 73
7 - S1 nuclease sensitivity. 77
8 – Single stranded mapping protocol. 80
9 - Single stranded mapping. 82
10 - Gel shift assay with single stranded DNA. 85
11 - Southwestern blot with single stranded DNA. 87
12 - λ Zap cDNA library screening. 90
13 – Nucleolin interacts with the PPY tract and conserved region in the λ Zap screening. 93
14 – Endogenous nucleolin binds to the PPY tract. 96

CHAPTER 5

15 – More deletional analysis within the non-coding exon of the M$_i$ gene. 109
16 – SHARP-1 binds to the M$_i$ promoter in a yeast one-hybrid screen. 113
17 – SHARP-1 is closely related to SHARP-2/STRA13/DEC1. 116
18 – Expression of SHARP-1 in different cell lines. 119
19 – SHARP-1 represses M$_i$ expression. 121
20 – SHARP-1 homodimerises. 124
21 - SHARP-1 acts as transcriptional repressor. 127

22 - SHARP-1 represses transcription through two independent domains. 130

23 - Repression by SHARP-1 is through two independent mechanisms. 133

24 - SHARP-1 preferentially binds to a B box. 136
ABBREVIATIONS

ACh- Acetylcholine
AdMLP- Adenovirus major late promoter
AGP- Acid glycoprotein
3AT- 3-amino-1,2,4-triazole
bHLH- basic helix loop helix
BMP- Bone morphogenetic protein
CBP- CREB-binding protein
CMV- Cytomegalovirus
CNF- Ciliary neurotrophic factor
CNS- Central nervous system
COL4A2- Collagen type IV A2
COX- Cytochrome c Oxidase
CREB- cAMP-response element binding protein
CTD- C-terminal repeats
CTF/NF1- CCAAT-binding transcription factor/nuclear factor-1
DAG- Diacylglycerol
DBD- DNA binding domain
DBH- Dopamine β-hydroxylase
DEC1- Differentiated embryo chondrocyte 1
DEP- Downstream promoter element
DSP1- Dorsal switch protein
DTT- Ditiothreitol
EBNA- Epstein-Barr virus nuclear antigen
EDTA- Ethylenediaminetetraacetic acid
EPII- E47 interaction partner 1
EMC- extramacrochaetae
E(spl)- Enhancer of Split
ETS- External transcribed spacer
FBP- FUSE binding protein
FGF- Fibroblast growth factor
FUSE- Far upstream sequence element
GAL4 DBD- GAL4 DNA binding domain
Gcn5- General controller of nutrition 5
GDNF- Glial cell-derived neurotrophic factor
GluR2- Glutamate receptor 2
GPCR- G-protein coupled receptor
GST- Glutathione S transferase
HAT- Histone acetyltransferase
HDAC- Histone deacetylase
HES-1- Hairy and Enhancer of Split homologue 1
HIV- Human Immunodeficiency Virus
hnRNP- Heterogeneous nuclear ribonucleoprotein
HO- Homothallic switching endonuclease
hprt- hypoxanthine-guanine phosphoribosyl transferase
Id- inhibitor of DNA binding
Inr- Initiator
IP3- Inositol phosphate
kb- kilobase
kDa- Kilo Dalton
Kr- Kruppel
KRAB- Kruppel-associated box
L1-CAM- L1 cell adhesion molecule
LIF- Leukaemia inhibitory factor
mAChR- Muscarinic acetylcholine receptor
Mash1- Mouse achaete-scute homologue
MBP- Myelin basic protein
MYC-CF1- MYC-common factor 1
MYC-PRF- MYC-plasmacytoma repressor factor
NCAM- Neurone cell adhesion molecule
NMDAR1- N-methyl-D-aspartate receptor 1
NUDR- nuclear DEAF-1-related
DEAF-1-Deformed epidermal autoregulatory factor-1
PAGE- polyacrylamide gel electrophoresis
PCAF- p300/CBP associated factor
PCR- Polymerase chain reaction
PDX1- Pancreas/duodenum homeobox 1
PIC- Pre-initiation complex
PIP<sub>2</sub>- Phosphatidylinositol-4,5-bisphosphate
PLC- Phospholipase C
PPY/PPU tract- Polypyrimidine/polypurine tract
PPY tract- Polypyrimidine tract
Rb- retinoblastoma
RBD- RNA binding domain
RE-1/NRSE- Repressor element 1/Neurone-restrictive silencer element
REST/NRSF- RE-1 silencing transcription factor/Neurone-restrictive silencer factor
rRNA- ribosomal RNA
SAGA- Spt-Ada-Gcn5-Acetyltransferase containing complex
SAP- Shrimp alkaline phosphatase
SCG10- Superior cervical ganglion-10
SDS- Sodium dodecyl sulphate
SHARP-1- Enhancer of Split and Hairy related protein 1
SHH- Sonic hedgehog
SIN3- Switch independent
STRA13- Stimulated with retinoic acid 13
SUC2- Sucrose 2
SWI/SNF- Switch/sucrose non-fermenter
TAF- TBP-associated factor
TBP- TATA binding protein
TE- Tris-EDTA
EDTA- Ethylenediaminetetraacetic acid
TF- Transcription factor
TGF- Transforming growth factor
tk- thymidine kinase
TRE- Tetracycline responsive element
TSA- Trichostatin A
VSM- Vascular smooth muscle
UAS- Upstream activator sequence
5'UTR- 5' untranslated region
UV- Ultraviolet
YY1- Ying Yang 1
CHAPTER 1

INTRODUCTION
INTRODUCTION

The primary goal of studies in developmental neurobiology has been to understand the cellular and molecular mechanisms that give rise to the establishment of neural circuits assembled during development. This includes specification of neuronal and glial cell types, and specification of connections between appropriate synaptic partners. There are thought to be many hundreds of different neuronal types. The phenotypic differences that characterise this multitude of cells are largely due to differential gene expression. Considering that the nervous system has been estimated to express more than 10,000 genes, with most of these genes having a unique expression profile, precise mechanisms that control the correct transcriptional pattern must exist.

The work presented here is an attempt to understand the transcriptional mechanisms responsible for the establishment of different neuronal phenotypes. As a model, the mechanisms that control transcription of \( M_1 \) muscarinic acetylcholine receptor (mAChR) gene, a gene that presents a unique expression profile within the nervous system, are analysed.

CUES THAT CONTROL DEVELOPMENT OF THE NERVOUS SYSTEM

Inducing factors

The nervous system arises from a population of cells located along the dorsal midline of a sheet of ectodermal cells (Reviewed in (Albright et al., 2000; Cowan et al., 1997; Kandel et al., 2000)). Experiments performed by Spemann and Mangold in the early 1920s (Spemann and Mangold, 1924) showed that ectodermal cells could be directed to generate neural cells in
response to signals secreted by a group of mesodermal cells termed the
organiser region. These experiments constituted the first evidence of the
existence of neural inducing factors and prompted a search for the identity
of these factors. In 1989, during a screening for candidate factors that
induce mesoderm differentiation, Smith and colleagues (Smith et al., 1989)
purified members of the fibroblast growth factor and transforming growth
factor β (TGF-β) families, and identified them as the mesoderm-inducing
factors. After the isolation of TGF factors, several inducing factors were
identified; this class of determinants consists of secreted and
transmembrane factors provided by one group of cells that are able to direct
the fate of cells exposed to them. Bone morphogenetic proteins (BMP), a
subclass of TGF proteins, are an example of inducing factors that exhibit a
broad range of cellular effects depending on developmental stage and
location (Mehler et al., 1997). Early in development, BMP4 and 7 direct
ectodermal cells to differentiate into epidermis. Other factors, including
follistatin, noggin and chordin, secreted by the organiser region, inhibit
the activity of BMP, causing cells of the ectoderm to differentiate to form
the neural plate. After the neural tube has closed, BMP factors are
responsible for generating several classes of neurones in the dorsal spinal
cord. Sonic Hedgehog (SHH), a secreted morphogen, can induce different
cell fates along the dorsal ventral axis of the neural tube at different
concentration thresholds (Briscoe and Ericson, 1999). As such, BMPs and
SHH induce dorsal and ventral fates in the neural tube and thereby
establish dorso-ventral polarity along the neural axis. Intermediate cell
types exhibit a distinct phenotype that is thought to result from co-
operative interactions of these two classes of inductive signals.

Intrinsic factors

Different cell types may have different responses to the same
inducing factor, consequently giving rise to the generation of different
neuronal cell types. The reason for this is that inducing factors exert their effect by interacting with membrane receptors in the target cell, but distinct cells may possess different cell receptors. Furthermore, stimulation of the target receptor initiates cascades of signalling, ending with changes in gene expression, and again, distinct cells may contain different effector systems. Cells gradually acquire independence from extrinsic signals and become progressively more reliant on intrinsic programs, triggered by factors that include cell surface receptors and transcription factors.

It is now known that TGF-β family members elicit their response by activating TGF-β receptors, which triggers the phosphorylation of SMAD family members (Itoh et al., 2000). SMADs are a family of transcription factors present in the cytoplasm, which after receptor-mediated phosphorylation, dimerise with other members of the same family and translocate to the nucleus, where they regulate, in co-operation with other transcription factors, the expression of target genes. Accordingly, both TGF-β receptors and the SMAD family of transcription factors act as intrinsic cues involved in the dorso-ventral patterning of the neural tube.

During development, the neural tube gives rise to the forebrain, midbrain, hindbrain and spinal cord. The hindbrain presents a number of swellings along the rostro-caudal axis, termed rhombomeres, each of them containing sensory and motor neurones that innervate individual branchial arches. How is the identity of an individual rhombomere established? Different members of the family of Hox genes, which share a 60-amino acid DNA binding domain termed the homeodomain, are expressed in different rhombomeres, and expression of specific combinations of Hox genes specifies the identity of different classes of neurones in the hindbrain. Therefore, HOX proteins act as intrinsic
factors involved in the rostro-caudal patterning of the nervous system. *Hox* gene expression in the hindbrain is regulated in part by mechanisms that are intrinsic to the neural tube but is also influenced by signals from surrounding mesodermal cells.

**Identification of transcription factors through genetic studies**

The emergence of the central role of transcription factors as determinants of neuronal identity became evident with genetic studies in *Drosophila*. In *Drosophila*, sensory organs develop in a precise pattern (reviewed in (Campuzano and Modolell, 1992; Fisher and Caudy, 1998)). Mutations that suppress formation of sensory organs or promote development of extra sensory organs in ectopic positions allowed the identification of a group of transcription factors of the basic helix loop helix (bHLH) family involved in sensory organ cell fate determination. For example, mutations in a gene termed *achaete* result in fewer sensory organs, whilst mutation in another gene termed *hairy* results in the production of ectopic sensory hairs, indicating a role for *achaete* in promoting the formation of neuronal precursors and a role for *hairy* in repressing formation. Cloning of *achaete* and *hairy* showed that they act by activating and repressing transcription, respectively.

Identification of *Drosophila* and *Caenorhabditis elegans* developmental factors led to the cloning of their vertebrate homologues. These studies revealed a remarkable degree of evolutionary conservation in developmental regulatory programs. Genetic studies in mice and zebrafish demonstrated that a high proportion of the identified genes have critical functions in establishing identity of neural cells. Disruption of *Mash1* (mouse *achaete-scute* homologue) leads to the loss of specific subset of neurones (Guillemot *et al.*, 1993). However, *Mash1* does not commit multipotent cells to a neural fate, like its *Drosophila achaete-scute*
counterparts, but rather promotes the differentiation of a committed neuronal precursor (Sommer et al., 1995). In Hes-1 (a homologue of hairy and enhancer of split) deficient mice, postmitotic neurones appeared prematurely, suggesting a role for Hes-1 in controlling the proper timing of neurogenesis and regulating neural tube morphogenesis (Ishibashi et al., 1995).

Identification of transcription factors through their target genes

Much work has been done on early development and many genes required for this process have been identified through genetic approaches. Many of these genes encode transcription factors and function by regulating gene expression. The target genes of these transcription factors however have proved difficult to identify and very few targets have been identified with any certainty. An alternative approach to sort out the problem is to identify the factors responsible for the transcriptional control of genes expressed in terminally differentiated neurones, including those genes that code for the transmitters that signal to other organs and the receptors that permit the cell to respond to incoming signals. Once the factors that control expression of these genes are known, their regulators can be similarly identified, ultimately resulting in the identification of a network of regulators responsible for neuronal development.

Using such an approach led to the identification of the homeobox-containing protein PHOX2a/ARIX. PHOX2a was originally identified in a screen for proteins that interact with an enhancer element responsible for cell specific expression of the neurone cell adhesion molecule (NCAM) gene (Valarche et al., 1993). Expression of Phox2a in all noradrenergic neurones suggested a role for PHOX2a as part of the regulatory cascade.
which determines the noradrenergic phenotype (Valarche et al., 1993). Subsequently, PHOX2a was isolated by virtue of its interaction with an enhancer element present in the dopamine β-Hydroxylase (DBH) gene (Zellmer et al., 1995), which codes for an enzyme in the synthesis pathway of the neurotransmitter noradrenaline. *In vivo* studies showed that mutant mice for Phox2a completely lack the locus coeruleus, the main noradrenergic centre of the brain and that affected sensory ganglia fail to express DBH transiently as they normally do (Morin et al., 1997). Furthermore, these mice also failed to express RET, a receptor subunit of glial cell-derived neurotrophic factor (GDNF), and undergo a massive increase in apoptosis. The authors suggest that is through the regulation of Ret expression that PHOX2a exerts its trophic control. Therefore, these results show that PHOX2a, a factor originally identified during promoter studies of the NCAM gene, has a role in controlling neuronal cell-type differentiation as well as neuronal survival.

Another example of a factor isolated through its target gene is represented by the zinc finger protein repressor element silencing transcription factor/neurone-restrictive silencer factor (REST/NRSF). REST/NRSF was originally identified as a DNA binding protein that interacts with a conserved regulatory element found in the promoters of many neurone-restricted genes, such as the sodium type II channel (Chong et al., 1995; Maue et al., 1990), SCG10 (Mori et al., 1990; Schoenherr and Anderson, 1995) and the M4 muscarinic acetylcholine receptor (Mieda et al., 1997; Wood et al., 1996) genes, and acts as a repressor of neuronal gene expression in non-neuronal cells. A database search using consensus REST/NRSF binding sites, (repressor element 1/neurone-restrictive silencer element; RE-1/NRSE), identified over 100 genes, including two transcription factors that may act as positive regulators of neuronal differentiation and at least 8 non-neuronal genes (Buckley,
unpublished observations)(Schoenherr et al., 1996). In this case, the large size of the RE-1/NRSE, 21 bp, permitted the identification of potential targets in a database search. Although database searches are not possible with small binding sites, promoter studies like those that led to the identification of REST/NRSF are a viable approach to identifying transcription factors, and subsequently, their cognate target genes. This approach can thus help to elucidate the mechanisms by which transcription controls the establishment of cell type.

**MUSCARINIC RECEPTORS**

**Muscarinic Acetylcholine receptors: a model to study neuronal differentiation**

GTP-binding protein coupled receptors (GPCRs) represent the largest gene family of the mammalian genome, containing over 1,000 members. This family includes genes coding for receptors responsive to light, neurotransmitters and odorants. Five mAChRs genes have been identified and constitute the first complete GPCR subfamily to be cloned and mapped within the nervous system (Bonner et al., 1987; Bonner et al., 1988; Buckley et al., 1988). Expression of the mAChR genes begins at the stage of terminal neurogenesis/early differentiation (Buckley, unpublished observations). In situ hybridisation and immunocytochemistry experiments (reviewed in (Caulfield, 1993; Hulme et al., 1990)) have shown that within the brain, each of the muscarinic receptor genes (M₁-M₅) has a unique expression profile. However, this expression profile overlaps considerably, with some cells expressing one or another receptor subtype, but other cells expressing a mixture of two or more subtypes, therefore, this family represents a useful model with which to study the processes that drive differential gene expression.
within the nervous system. Understanding the molecular mechanism by which cell-specific expression of mAChR subtypes is attained should provide important insights into the development of differentiated neuronal phenotype.

In the present study, analysis of the mechanisms that direct cell specific expression of the M₁ gene is presented. Studies of M₁ transcriptional regulation in particular, and any neurone-specific gene in general contribute to the understanding of the processes that lead to the establishment of cellular phenotype and how these sum to produce a neuronal phenotype.

**Muscarnic receptors**

mAChR mediate the slow responses of acetylcholine (ACh) (Hulme et al., 1990) via their effect on GTP-binding proteins (G-proteins) (reviewed in (Dohlman et al., 1991) and (Strader et al., 1994)). Acetylcholine binding leads to conformational changes in the receptor protein which enable the receptor to bind to specific G-proteins (consisting of α, β and γ subunits). This interaction promotes the exchange of GTP for GDP on the G protein α subunit. GTP binding leads to the dissociation of the G-protein from the receptor and the dissociation of the α subunit (α-GTP) from the βγ complex. α-GTP and βγ are then able to activate or inhibit specific effector enzymes and/or ion channels, leading to a physiological response.

**Structure**

As with other GPCRs, mAChRs consist of 7 transmembrane (TM) α-helices with the N-terminal region facing the extracellular space and the C-terminal domain in the cytosol (Bonner et al., 1987; Bonner et al., 1988). It has been predicted that the seven TM domains are arranged in a
ring-like fashion. A similar structural arrangement has been described for bacteriorhodopsin whose three-dimensional structure has been studied in great detail by X-ray crystallography (Henderson and Unwin, 1975). Site-directed mutagenesis and affinity labelling studies have shown that specific residues, some conserved among all mAChR subtypes are critically involved in the recognition of ACh and other muscarinic agonists, and also in G-protein coupling selectivity (reviewed in (Wess, 1995)).

Receptor effector coupling

Each of the muscarinic receptor subtypes has the potential to stimulate different cellular responses by coupling to specific G-proteins, which then interact directly either with an effector such as an ion channel or a second messenger. Stimulation of M₂ and M₄ receptors has mainly been linked to inhibition of adenylyl cyclase and inhibition of Ca²⁺ channels through selective coupling to G-proteins of the Gᵢₒ family. Stimulation of M₁, M₃ and M₅ is mainly linked to activation of phospholipases and activation/inhibition of K⁺ channels through coupling to G-proteins of the Gᵢₒ family (Peralta et al., 1988).

Phospholipase Cβ₁ (PLCβ₁), coupled through Gαq or Gα₁₁, serves as the primary effector for the M₁ receptor (Berstein et al., 1992). Activation of PLC mediates hydrolysis of the membrane phospholipid phosphatidylinositol-4,5-bisphosphate (PIP₂) to generate diacylglycerol (DAG) and inositol triphosphate (IP₃). DAG binds to and stimulates protein kinase C, while IP₃ liberates calcium stored in the endoplasmic reticulum by binding to the IP₃ receptor, an IP₃-sensitive calcium channel. M₁, M₃ and M₅ receptors can also stimulate the production of IP₃ by a mechanism that involves the tyrosine kinase-dependent phosphorylation of phospholipase Cγ (Gusovsky et al., 1993).
$M_1$, $M_3$ and $M_5$ receptors, but not $M_2$ nor $M_4$, are linked to phospholipase A2 activation (Conklin et al., 1988). Activation of phospholipase A2 catalyses the hydrolysis of membrane phospholipids to generate free arachidonic acid and the corresponding lysophospholipid. Arachidonic acid serves as a substrate for the formation of a number of bioactive eicosanoids, including prostaglandines, thromboxanes and leukotrienes.

$M_1$ and $M_3$ receptors are also responsible for inhibition of $K_M$-current (or also called M-current) (Fukuda et al., 1988; Marrion et al., 1991). Following an action potential, $K_M$ channels, which are a subtype of voltage-gated $K^+$ channels, open allowing $K^+$ outside of the cell and causing hyperpolarisation of the cell and therefore inhibiting excitability. Muscarinic agonists mediate the closure of the $K_M$ channels, resulting in a less negative membrane potential and an increase in excitability.

Finally, muscarinic receptors mediate inhibition of a subclass of $Ca^{2+}$-activated $K^+$ channels known as AHP channels (Lancaster and Adams, 1986; Madison et al., 1987). AHP channels generate a long lasting after-hyperpolarisation following the influx of $Ca^{2+}$ during a preceding action potential.

In summary, different subtypes of muscarinic receptors recognise the same neurotransmitter, ACh, but may elicit different responses within the cell when they are activated. The repertoire of receptors and effector systems expressed by a neurone will therefore determine how the response is manifested.
**Functional studies**

Activation of mAChR leads to a variety of responses including modulation of ganglionic transmission in autonomic ganglia, contraction of smooth muscle, relaxation of cardiac muscle, stimulation of secretion in glandular tissue and both stimulation and inhibition of central nervous system (CNS) neuronal excitability (Hulme et al., 1990). Relatively little is known about the functional role of mAChR in CNS, but they have been implicated in learning, memory, attention and cognition as well as many sensory, motor and autonomic processes (Di Chiara et al., 1994; Dragunow, 1996).

Transgenic mice lacking the M₁ receptor gene are resistant to epileptic seizures induced by the muscarinic agonist pilocarpine. In addition, they show no muscarinic modulation of the M-type potassium channel in sympathetic ganglion neurones, thereby supporting the notion that M₁ receptors play a role in the control of neuronal excitability (Hamilton et al., 1997).

Muscarinic involvement in memory can be implied by the observations that administration of muscarinic agonists to trained rats produces an improved memory response, whilst muscarinic antagonists impair memory function (Messer et al., 1990). In another study, it was shown that treatment of apolipoprotein E-deficient mice, which are widely used as a model for Alzheimer's diseases, with the M₁ agonist AF150(S) abolishes their memory deficits (Fisher and Caudy, 1998).

In Alzheimer's disease patients, cholinergic signalling has been found to be heavily impaired as a result of a degeneration of the basal forebrain nucleus (Whitehouse et al., 1982). Activation of M₁ and M₃ muscarinic receptors prevents amyloid-β deposition into senile plaques, suggesting that decreased cholinergic input contributes to the formation
of senile plaques observed in Alzheimer's disease patients (reviewed in (Hellstrom-Lindahl, 2000)). It has been shown that amyloid-β deposition disrupts muscarinic receptor coupling to G-proteins (Ladner and Lee, 1998; Pavia et al., 1998). Impaired muscarinic receptor-G-protein coupling leads to an increased formation of amyloid-β, which in turn may exert additional modulation of the cholinergic system.

Another major hallmark of Alzheimer's disease is neurofibrillary lesions, which are due to the dissociation of hyperphosphorylated tau protein from microtubules, leading to disruption of neuronal cytoskeleton and interference with intracellular transport mechanisms (reviewed in (Hellstrom-Lindahl, 2000)). It has been shown that stimulation of M1 receptor can cause a decrease in the levels of phosphorylated tau (Sadot et al., 1996), therefore, decreased cholinergic activity associated with Alzheimer's disease may also contribute to an increased phosphorylation of tau that ultimately leads to cell death.

Accordingly, drug development for treatment of Alzheimer's disease has targeted enhancement of the cholinergic system via cholinergic agonists or cholinesterase inhibitors which directly prolong and/or increase the synaptic concentration of acetylcholine (Fisher, 2000; Newhouse et al., 1997; Nordberg and Svensson, 1998).

The involvement of muscarinic receptors in motor control can be seen by the fact that M4 receptor-deficient mice show an increase in basal locomotor activity and are hypersensitive to the stimulatory locomotor effects of D1 dopaminergic receptor activation (Gomeza et al., 1999). Dopaminergic and cholinergic receptors reciprocally modulate the release of each other's ligand, probably at the level of the striatum, where the two receptors are co-expressed, i.e.; dopamine inhibits the release of ACh, and ACh inhibits dopamine release. Consistent with this, both
dopamine receptor agonists and muscarinic receptor antagonists are clinically useful in the treatment of Parkinson's disease, a movement disorder caused by the loss of dopaminergic nerve terminals in the striatum (Barbeau, 1962). On the other hand, schizophrenia is thought to be due to an excessive activity of dopaminergic synapses, and in turn, the role of muscarinic receptors is being considered. High doses of some muscarinic antagonists produce psychotic-like symptoms such as delusions, hallucinations and memory loss (Perry and Perry, 1995). The use of muscarinic agonists in the treatment of schizophrenia is currently being investigated.

Localisation

Each of the muscarinic receptors has been shown by in situ hybridisation studies to have a unique pattern of mRNA tissue distribution within the CNS (Buckley et al., 1988; Vilaro et al., 1990). The M1 is the most abundant subtype, being particularly abundant in the hippocampus, cerebral cortex, striatum and dentate gyrus. M2 transcripts are relatively rare within the CNS and only found in significant quantities in the medial septum and pons, with lower levels present in the thalamus. M3 is localised predominantly in the forebrain with expression also in some thalamic nuclei. The striatum contains the highest levels of M4 transcripts, whilst lower levels are present in the cortex and hippocampus. M5 transcripts are present at very low levels in the hippocampus.

Immunocytochemistry studies using specific antibodies against each receptor subtype have complemented the in situ hybridisation studies, providing more information on the abundance and precise location of the proteins themselves (Dorje et al., 1991; Levey et al., 1991). The results of these studies show that the M1 subtype is present in the cortex and striatum, being localised in the cell bodies and neurites.
Expression of $M_2$ receptor protein is found in the hindbrain, basal forebrain striatal neurons, mesopontine tegmentum and cranial motor nuclei. This distribution is coincident with that of cholinergic neurones. The $M_2$ receptor is also found in non-cholinergic neurones in cortical and sub-cortical structures. Immunoprecipitation studies (Wall et al., 1991) have identified $M_3$ receptor protein in the cortex, hippocampus, striatum and olfactory tubercule, and lower densities in the hindbrain region. The $M_4$ receptor is localised in the neostriatum, olfactory tubercule and islands of Calleja. $M_5$ receptor proteins could not be detected with this technique, probably because of low receptor protein levels.

**Regulation of expression**

As previously described, acetylcholine can stimulate a wide range of second messenger systems by acting through each of the five muscarinic receptor subtypes. Multiple mechanisms exist for the regulation of the expression and activity of each of the receptors, providing a mechanism by which the responses to acetylcholine are processed and determine what effects these have on the cell. Regulation can occur at three basic levels. Firstly, the cellular response to agonists acting on cell surface receptors diminishes rapidly with continued exposure of the cells to this stimulus. This phenomenon has been termed desensitisation, and is the result of uncoupling of the receptor and effector mechanism. Secondly, continued exposure of cells to agonist results in a loss of cell surface receptors, a process known as down-regulation, which includes degradation of receptor protein and decrease of mRNA levels. Recovery of mAChR from down-regulation requires expression activation to give rise to newly synthesised protein. Finally, regulation at the level of transcription defines the distribution pattern described above for each of the muscarinic receptor subtypes and is probably the most important mechanism of regulation during development.
CONTROL OF TRANSCRIPTION

In eukaryotes, three RNA polymerases are responsible for transcription. RNA polymerase I transcribes rRNA, RNA polymerase II transcribes mRNA and RNA polymerase III transcribes tRNA and other small RNAs. RNA polymerase II is therefore responsible for transcription of the muscarinic receptors.

Transcription starts when RNA polymerase binds to the promoter of a gene. The promoter is the region of the gene involved in binding of RNA polymerase to initiate transcription (Lewin, 1997). This includes the DNA sequence surrounding the transcription start site, a downstream promoter element (DPE) and a sequence present in many but not all genes termed the TATA box (Breathnach and Chambon, 1981). In higher eukaryotes, the TATA box is found about 30 base pairs upstream of the transcription start site and plays an essential role in accurately positioning the transcription start site. The TATA box is recognised by the TATA-binding protein (TBP) (Cavallini et al., 1989; Hahn et al., 1989; Hoey et al., 1990; Hoffman et al., 1990; Horikoshi et al., 1989; Schmidt et al., 1989), which forms part of a large complex termed TFIID. Besides TBP, the TFIID complex contains between 8 and 14 TBP-associated factors (TAFs), which help to orient and stabilise the pre-initiation complex (PIC) (Albright and Tjian, 2000). In TATA-less promoters, TFIID is still required for transcription (Pugh and Tjian, 1991). The basic transcription machinery, which contains RNA polymerase II, TFIID and several general transcription factors and associated proteins (including TFIIA, B, E, F and H) (Matsui et al., 1980; Samuels et al., 1982), is sufficient for efficient and accurate initiation from a core promoter in vitro. In prokaryotes, comparable rates of transcription are achieved in vivo and in vitro.
However, in eukaryotes, DNA is packaged into chromatin, and even a strong core promoter is inactive (reviewed in (Struhl, 1999)). In prokaryotes, a family of DNA binding proteins interact with specific DNA sequences found upstream of the transcription start site and modulate the levels of basal transcription. Repressors are required to keep gene activity at low levels in a subset of promoters and act by occlusion or generation of repressosome structures. Activators act on a set of promoters, that are inherently weak or specifically repressed, stimulating transcription by directly interacting with RNA polymerase (Hochschild and Dove, 1998). In eukaryotes, activators and repressors bind to cis-regulatory elements (Maniatis et al., 1987) and are essential to control the levels of expression in response to developmental programs (i.e. defining spatio-temporal expression), metabolic requirements, and external stimuli.

Although a remarkable number of common basic principles mediate the regulation of both eukaryotic and prokaryotic gene transcription, eukaryotic transcription appears to be enormously more complex. The salient difference is the packaging of eukaryotic DNA into chromatin, making promoters inaccessible to the basic transcription machinery. In eukaryotes, transcriptional regulators may act by interacting directly with components of the basic transcription machinery, such as in the case of prokaryotic regulators, or by altering chromatin structure.

**Activation through interaction with the basic transcription machinery**

TBP has been shown to be able to substitute for the TFIID complex as the nucleating factor for the assembly of the PIC in vitro (Peterson et al., 1990). However, while TFIID was found to be able to direct activation by
multiple transcription factors, TBP failed to support activator-dependent transcription (Kambadur et al., 1990; Pugh and Tjian, 1990; Tanese et al., 1991). Pugh and Tjian suggested that TAFs serve as potential targets for some transcriptional regulators that bind to the cis-regulatory elements (Pugh and Tjian, 1990). To date, several factors have been identified that interact with TAFs or other members of the PIC.

Spl (the first mammalian activator to be cloned (Dynan and Tjian, 1983a)) and CREB are two factors that contain an activation domain characterised by a high percentage of glutamine residues. Both factors have been found to interact directly with hTAF1130 (Tanese et al., 1996) (or its Drosophila homologue, dTAF1110 (Gill et al., 1994)) and mutations in Spl that reduce binding to hTAF1130, also reduce activation, suggesting that Spl mediates activation by interacting with a member of the basic transcription machinery.

CTF/NF1 factors are characterised by a proline-rich activation domain (Mermod et al., 1989). CTF/NF1 has been shown to interact with TBP through a heptapeptide found within the proline-rich domain highly similar to the C-terminal repeats (CTD) of RNA polymerase II (Xiao et al., 1994). However, in a reconstituted system free of TAFs, TBP was unable to support activation mediated by the CTF/NF1 proline-rich activation domain fused to the Gal4 DNA binding domain (Tanese et al., 1991). When TBP associated factors were added to this system, activated levels of transcription were restored, suggesting that one or more TAFs are required to mediate activation by CTF/NF1.

Activation domains from the yeast Gal4p (Gill and Ptashne, 1987) and glucocorticoid receptor (Hollenberg and Evans, 1988) transcription factors contain a large proportion of acidic amino acids. Both factors are able to synergistically activate transcription in a mammalian system.
(Kakidani and Ptashne, 1988), suggesting that they target the same molecule which is part of the basic transcription machinery and common to yeast and mammalian cells.

Although some regulators have similar types of activation domains, such as glutamine-rich, proline-rich or acidic domains, the number of distinct activation domains identified in different transcription factors largely outnumbers the targets in the basic transcription machinery. A series of specialised adapters, or co-factors, have been identified and are required to relay information imparted by some DNA-bound factor to the basic transcription machinery.

CREB-binding protein (CBP), and the closely related factor p300, are adapter proteins believed to participate in the activities of hundreds of different transcription factors (Vo and Goodman, 2001). Interaction of p300/CBP with components of the basic transcriptionary machinery, such as TFIID and TFIIB, have been suggested to contribute to p300/CBP function as co-activators (Abraham et al., 1993; Lee et al., 1996).

Transcription factors that activate by interaction with the basic transcription machinery are thought to do so by 1) stimulating the recruitment of the general transcription factors to the promoter or by 2) catalysing the conversion of the PIC from a closed to an open state (reviewed in (Lemon and Tjian, 2000)). Based on the formation of active transcription complexes in vitro, it was initially proposed that activator-mediated recruitment of a TFIID-TFIIA-TFIIB complex was followed by an ordered assembly of the PIC (Buratowski, 1994). However, an alternative model in which recruitment of TFIID is followed by the recruitment of a pre-assembled RNA polymerase II holoenzyme has also been suggested. It now seems that although some components of the core machinery almost certainly pre-exist as tightly associated complexes, compartmentalisation
of the nucleus plays an important role, increasing the concentration of selected factors at the right locus where transcription of specific genes must occur (reviewed in (Lemon and Tjian, 2000)).

Mechanisms of repression

In addition to activation, gene transcription may be actively repressed (reviewed in (Ogbourne and Antalis, 1998)). Repressors bind to regulatory cis-elements and as in the case of activators can repress transcription by interacting with the basic transcription machinery. *Drosophila* Dorsal Switch Protein (DSP1) represses transcription by directly interacting with TBP, which results in the displacement of TFIIA, and therefore in interference of the PIC assembly (Kirov et al., 1996). Kruppel (Kr) is another *Drosophila* factor essential in organogenesis during late embryonic development. When acting from a site close to a basal promoter, monomeric Kr interacts with TFIIB to activate transcription, whereas at high concentrations, Kr dimerises and interacts with TFIIE beta, a subunit of TFIIE, resulting in transcriptional repression (Sauer et al., 1995).

Other repressors act by affecting the activity of activators. This can happen at many stages. In some cases, repressors inhibit the binding of activators to their DNA-binding site whilst in other cases repressors have been found to interact with DNA-bound activators. The *Drosophila* repressor Cactus binds to the activator dorsal, interfering with the function of the nuclear localisation signal and therefore preventing its translocation to the nucleus (Kidd, 1992). Emc (or its mammalian homologue ID) are HLH proteins that lack the basic DNA-binding domain present in bHLH proteins, therefore are able to dimerise, but can not bind DNA. EMC forms non-DNA-binding heterodimers with bHLH activators of the Achaete-Scute and Daughterless family members, preventing these
activators from binding to DNA. A different mechanism to inhibit the binding of an activator appears to be by displacement or competition for the same site by the repressor. Sp3 is a factor that binds to the same site as Sp1, but unlike Sp1, cannot activate transcription (Hagen et al., 1994).

Binding to a DNA element by an activator is a pre-requisite for activation, but it is not sufficient. Contact must be made between the activator and the PIC or other components of the activation pathway. Some repressors are found to directly interact with activators, blocking this step of transcription initiation. A very well characterised example is represented by the interaction between the yeast activator Gal4p and the negatively acting factor Gal80p. In the absence of galactose, Gal80p binds to Gal4p, masking its activation domain and preventing activation of the divergent promoters of GAL1/10. In the presence of galactose, Gal80p undergoes a conformational change, exposing the activation domain of Gal4p, allowing transcriptional activation (Johnston et al., 1987; Ma and Ptashne, 1987b; Mylin et al., 1989; Mylin et al., 1990). In other cases the repressor needs to bind DNA to exert its effect. The transcription repressor MYC-PRF binds to a DNA sequence adjacent to the site for the transcription activator MYC-CF1, masking its activation domain, and therefore, inhibiting transcription from the c-myc gene (Kakkis et al., 1989).

Chromatin remodelling

In prokaryotes, genomic DNA is associated with histone-like proteins, however they are unlikely to have a transcriptional inhibitory role, since promoters are easily accessible and fully functional in vivo. In eukaryotes, DNA is packaged into chromatin, a beaded structure where the basic unit is the nucleosome. Each nucleosome is formed by a “core particle”, consisting of a stretch of 146 bp of DNA wrapped 1.75 times
around a histone octamer, and a "linker" DNA fragment of variable length connecting one nucleosome to the next (reviewed in (Kornberg and Lorch, 1999; Wolffe and Guschin, 2000)). The core octamer, which is formed by two of each the histone proteins H2A, H2B, H3 and H4 (Kornberg, 1974), makes contacts with the phosphodiester backbone of the DNA every 10 base pairs (Luger et al., 1997). Between 15 and 30 residues at the amino termini of all the histones are unstructured and commonly referred as "tails" (Luger et al., 1997), and appear to promote fibber formation perhaps by contacting adjacent nucleosomes or by influencing the configuration of linker DNA (Tse et al., 1998). H1 is an additional histone that interacts with the linker DNA, promoting coiling or folding of chromatin in the fibber (Carruthers et al., 1998). Opposite to what was initially thought, chromatin is not a static structure, but rather it is functionally specialised and structurally heterogeneous, playing a key role in transcriptional control. Both histone-DNA interactions in the core particle and histone tails interactions in the chromatin fibber are thought to contribute to repression of transcription by blocking the access of transcription factors and the basic transcription machinery to DNA (reviewed in (Kornberg and Lorch, 1999; Wolffe and Guschin, 2000)). Many transcription factors recruit multiprotein complexes with ATP-dependent chromatin remodelling activity, which modulates repression due to histone-DNA interactions in the core particle, whilst others recruit complexes with histone acetyltransferase (HAT) or histone deacetylase (HDAC) activity, which targets the histone tails and relieves repression due to condensation in the chromatin fibber or re-establishes it, respectively.

In general, acetylation of the core histone tails corresponds with a shift to a more open and accessible chromatin structure, whilst hypoacetylated states are linked to a repressed structure (reviewed in
A connection between acetylation and transcription became evident when the yeast transcriptional activator Gcn5p was shown to have HAT activity (Brownell and Allis, 1996), and that this activity was required for the activation of target genes in vivo (Kuo et al., 1998). Subsequently, many mammalian factors involved in transcriptional activation have been shown to have HAT activity. GCN5 (Candau et al., 1997) and PCAF (Yang et al., 1996b) are the mammalian homologues of yeast Gcn5p. Both factors are part of a larger complex, the SAGA complex, where additional factors are required to allow substrate specificity. p300 and its close homologue CBP are also HATs recruited by numerous transcriptional activators, such as CREB and MyoD, to stimulate transcription of specific genes (Bannister and Kouzarides, 1996; Ogryzko et al., 1996). The general transcription factor TAFII250, which is required for PIC assembly and to promote activated transcription has been shown to also have HAT activity (Mizzen et al., 1996), but surprisingly, its HAT activity is required for transcription at only a subset of promoters (O'Brien and Tjian, 2000).

Acetylation of histone tails can be reversed by histone deacetylases. Numerous proteins that were originally identified as co-repressors have now been shown to possess deacetylase activity, such as in the case of the yeast factor Rpd3p (Rundlett et al., 1996). In some cases, HDACs interact directly with DNA-bound factors. For example, YY1 directly interacts with HDAC1, and this interaction is essential for its transcriptional repression function (Yang et al., 1996a). However, in most cases, interaction of a HDAC factor with the DNA-bound transcriptional repressor is mediated by an adapter. The mSIN3 factors are members of a large complex including HDAC1 and HDAC2 and are responsible for mediating repression by a large number of transcription factors, such as MAD (Ayer et al., 1995) and
REST/NRSF (Grimes et al., 2000; Huang et al., 1999; Naruse et al., 1999; Roopra et al., 2000).

Finally, chromatin remodelling factors affect chromatin in a distinct manner to HATs and HDACs. These factors cause a net change in the position of the histone octamer relative to DNA, modifying the nuclease digestion pattern of nucleosomal DNA (Cote et al., 1994). For this, chromatin remodelling factors use the energy derived from ATP hydrolysis to break and re-establish histone-DNA contacts within the core particle of the nucleosome (Tyler and Kadonaga, 1999). The SWI/SNF chromatin remodelling complex was originally identified as an activator of gene expression of the SUC2 and HO genes in yeast (Winston and Carlson, 1992). This complex has recently been shown to be responsible for the transcriptional control of about 6% of S. cerevisiae genes, both in a positive and a negative manner (Holstege et al., 1998), suggesting a role for chromatin remodelling complexes not only in transcriptional activation, but also repression.

If nucleosome remodelling factors are targeted to specific regions of DNA by transcription factors, how do transcription factors penetrate chromatin to access the DNA binding sites? Some experiments are helping to answer this question. Some activators, such as the glucocorticoid receptor, can bind DNA packaged into chromatin, whereas others, such as CTF/NF-1 cannot. Glucocorticoid receptors have been shown to recruit SWI/SNF complex, which in turn may facilitate the binding of other factors (Neely et al., 1999). Similarly, studies of the HO endonuclease gene in yeast have shown a sequential recruitment of regulators (Cosma et al., 1999; Krebs et al., 1999). Binding of a first set of transcriptional activators to upstream sequences preceded events that required the SWI/SNF complex, followed by the SAGA complex, which promoted localised histone acetylation, enabling the binding of a second set of regulators. Binding of
this second set of regulators was required to direct the binding of the RNA polymerase holocomplex. Therefore the answer is that some transcription regulators facilitate the access of chromatin remodelling factors to DNA, and these chromatin remodelling factors facilitate the access of additional transcription factors.

Other mechanisms of transcriptional control

Post-translational modification of chromatin not only by acetylation, but also by other means such as phosphorylation and methylation, are mechanisms that regulate the access of transcription factors to their DNA targets and also may act as specific molecular signals for DNA recognition (Strahl and Allis, 2000).

In addition to histone modifications, genomic DNA may also be modified. Methylation of CpG dinucleotides plays an important role in the inactivation of the X-chromosome, genomic imprinting, immobilisation of mammalian transposons, suppression of transcriptional noise and the control of tissue-specific gene expression (Bird and Wolffe, 1999). Methylation of DNA can interfere with transcription either by preventing the binding of basic transcription machinery and specific transcription factors that require contact with cytosine or by influencing nucleosome stability. MeCP2 was the first methylated-DNA binding protein to be identified (Lewis et al., 1992), and it has been shown that it is able to interact with a HDAC complex via mSIN3 (Jones et al., 1998), promoting a chromatin repressive environment. Subsequently, other methyl-CpG binding proteins have been identified and shown to recruit different chromatin remodelling and HDAC complexes (Hendrich and Bird, 1998).

The sum of all of the above mentioned mechanisms plus all the potentially yet uncharacterised processes act in concert throughout the
genomic to give rise to the phenotypic diversity between cells seen within whole organisms in general and the nervous system in particular.

**Transcriptional control of muscarinic genes**

To date, the mechanisms that control expression of a subset of neural specific genes have been studied, contributing to the understanding of the processes that give rise to neuronal differentiated phenotypes. Within this group, the M2 and M4 muscarinic receptors genes are included. Transcription of the M4 gene appears to be under the control of two promoters and cell specific expression is achieved, at least in part, through selective repression in M4 non-expressing cells mediated by the binding of the transcription factor REST/NRSF to an RE-1/NRSE present 550 bp upstream of the transcription start site (Wood et al., 1996) (Mieda et al., 1997; Roopra, 1997). In the case of the M2 subtype, expression levels in heart are regulated by the GATA family of transcription factors (Rosoff and Nathanson, 1998) and both the level of endogenous M2 AChR protein and gene transcription are increased by CNTF and LIF (Rosoff et al., 1996). Although the M2 gene is also expressed in the nervous system, the role of GATA factors in controlling expression in this tissue has not been studied.

**Transcriptional control of the M1 mAChR gene**

The expression patterns of the M1 and M4 genes are broadly similar, with both genes being expressed in autonomic ganglia and in the central nervous system, where expression is largely restricted to telencephalic regions. In the latter case, many neurones of the cortex, hippocampus and striatum co-express both M1 and M4 genes, but there are also subpopulations of neurones that express only one or neither gene (Bernard et al., 1992; Weiner et al., 1990). Since these two genes are family members, share a common gene structure and have similar expression patterns, it may be expected they share common regulatory
mechanisms. The aims of my project have been to identify the regulatory elements that control expression of the \( M_i \) gene, identify the transcription factors that bind to these regulatory elements and to determine the mechanisms by which these transcription factors work, in an attempt to understand the mechanisms that lead to the establishment of cellular phenotype in general and neuronal phenotype in particular.
CHAPTER 2

MATERIALS AND METHODS
MATERIALS AND METHODS

Cell Culture - Cell lines were cultured in 5% CO$_2$ at 37°C and all media was supplemented with 6 g/litre penicillin, 10 g/litre streptomycin and 2mM L-glutamine. IMR32, 3T3 and Neuro2a cells were grown in Dulbeco’s modified Eagle's medium (Life Technologies Inc.) supplemented with 10% foetal calf serum. NB4 1A3 cells were grown in HAMS F10 (Life Technologies Inc.) supplemented with 15% horse serum and 2.5% foetal 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 Moloney Murine Leukaemia Virus (MmuLV) reverse transcriptase (Promega). The oligonucleotides used to amplify the M$_i$ gene in the cultured cell lines were: RM$_i$ 1069s 5'-CTGGTCAAGGAGAAGAAGGCAG-CT and RM$_i$ 1518a 5'-GTCTCTCTGGGCTCCAGGAAGG. The oligonucleotides used to amplify the hypoxanthine-guanine phosphoribosyl transferase, (HPRT) gene were: hprt 231s 5'-CCTGCTGGATTACATTAAAGCACTG and hprt 567a 5'-CCTGAAGTACTCATTATAGTCAAGG. Oligonucleotides used to amplify the SHARP-1 gene were: SHARP-1.125s 5'-AGGATACCTACAAATTACCGC and SHARP-1.441a 5'-CGCGAGGTATTGCAAGAC. Numbers are relative to the translation start site of SHARP-1. Aliquots of the reaction mixture were subjected to electrophoresis on a 2% MetaPhor agarose gel.

Reporter plasmid constructs - Reporter plasmids were constructed using the luciferase reporter vector pGL3 Basic (Promega). A 2 kb HindIII fragment obtained from B/P4.0pGEM3Z (Pepitoni et al., 1997) was cloned into the HindIII site of pGL3 (Promega) to yield pGL3 -1390/+602 M$_i$. A 0.75 kb Acc65I/EcoRI fragment was excised from B/P4.0pGEM3Z and cloned.
into Acc65I/EcoRI cut pGL3 -1390/+602 M1 to yield pGL3 -1602/+602 M1. Exonuclease digestion (Erase-A-Base, Promega) of pGL3 -1390/+602 M1 was used to generate pGL3 -1260/+602 M1. To generate pGL3 -881/+602 M1, a 0.5 kbSacI fragment was excised from pGL3 -1390/+602 M1 and then religated. The construct pGL3 -1390/+602 M1 was used as template for PCR to generate the following constructs: To clone pGL3 -372/+602 M1, a PCR product obtained using the primers M1 -372s,BglII and GL2 (Promega) was digested with BglII and HindIII and cloned into pGL3 basic digested with BglII and HindIII. pGL3 -372/+602 M1 ΔSp1 was constructed by fusion PCR. Briefly, two sets of oligos; fusion M1ΔSp1s/M1+602a and M1 -372s/fusion M1ΔSp1a, which contain an overlapping sequence were used for PCR, and the products electrophoresed in a 2% MetaPhor agarose gel. A mix of equimolar amounts of each PCR product was used as template in a PCR reaction using the primers M1 -372s and M1 +602a. The fusion PCR was digested with BglII and cloned into pGL3 basic digested with BglII and treated with shrimp alkaline phosphatase (SAP) (Roche). To clone pGL3 -372/+166 M1, the plasmid pGL3 -372/+602 was partially digested with SmaI and religated. To generate pGL3 -372/+241 M1, the PCR product obtained with the primers M1 -372s and M1 +241a was digested with BglII and cloned into pGL3 basic digested with BglII and treated with SAP. pGL3 -372/+316 M1, pGL3 -372/+366 M1, pGL3 -372/+417 M1 and pGL3 -372/+504 were cloned in a similar manner by generating PCR products with the sense primer M1 -372 and the antisense primer M1 +316a, M1 +366a, M1 +417a or M1 +504a, respectively and cloned into the BglII site of pGL3 basic. pGL3 -372/+602M1 ΔPPY was constructed by fusion PCR with the two sets of oligonucleotides; fusion M1Δ+412/+485s//M1 +602a and M1 -372s//fusion M1Δ+412/+485a. To produce pGL3 Inr, the oligonucleotides Inr. HindIIIs and Inr.HindIIIa, containing the adenovirus major late promoter (AdMLP) initiator element, were annealed giving HindIII overhangs. This fragment was cloned into pGL3 basic digested with HindIII and treated with SAP. To
produce pGL3 +396/+569 M_i Inr, a PCR product obtained with M_i +396s.XmaI and GL2 was digested with XmaI and cloned into pGL3 basic digested with XmaI and SAP to make pGL3 +396/+569 M_i. This construct was digested with HindIII and treated with SAP. The annealed Inr.HindIII (s+a) oligonucleotides were then cloned into the HindIII site. To clone pGL3 +166/+602 M_i Inr, the construct H/P-1.4/+0.6pGL3 was partially digested with SmaI and religated to give pGL3 +166/+602 M_i. The Inr fragment was cloned into the HindIII digested plasmid.

To generate pMT G4 hNucleolin, a PCR product corresponding to the coding region of the human nucleolin gene was generated using the plasmid pNFor4 (generously provided by Nancy Maizels (Hanakahi et al., 1997)) as template and the primers hNuc 3s and hNuc 2124a containing EcoRI linkers (numbers are relative to the translation start site of nucleolin), and cloned into pMT G4 (Roopra et al., 2000) digested with EcoRI.

The plasmid pBM2389 +417/+166 M_i was produced as follows: A PCR product generated by using as template pGL3 +166/+602 M_i and the primers M_i +417a and RV3 (Promega) was cloned into pGem-T Easy (Promega). The fragment was excised with EcoRI and cloned into pBM2389 (Liu et al., 1993).

SHARP-1 coding region was generated by PCR with the primers SHARP-1.-11s and SHARP-1.762a (numbers are relative to the translation start site of SHARP-1) containing NcoI linkers and cloned in frame into the NcoI site of pCS2+MT (Turner et al., 1994) to give pMT SHARP-1. To generate pMT G4 SHARP-1, a PCR product obtained using the primers SHARP-1.4s and SHARP-1.763a, consisting of the SHARP-1 coding region flanked by EcoRI linkers was cloned into the EcoRI site of pMT G4 (Roopra et al., 2000). PCR products with EcoRI linkers were generated using the sense primer SHARP-1.4s and the antisense primers SHARP-1.519a,
SHARP-1.306a or SHARP-1.147a, and similarly used to generate pMT G4 NbHO-SHARP-1 (residues 1-173), pMT G4 NbH-SHARP-1 (residues 1-102) and pMT G4 N-SHARP-1 (residues 1-49), respectively. PCR products with EcoRI linkers containing SHARP-1 fragments between positions 520 and 762, 307 and 519, and 127 and 306 were cloned into the EcoRI site of pMT G4 to generate pMT G4 C-SHARP-1 (residues 174-253), pMT G4 O-SHARP-1 (residues 102-173) and pMT G4 bH-SHARP-1 (residues 43-102), respectively. The reporter plasmids pTRE UAS TATA, pGL3 UAS TRE TATA and pGL3 UAS TRE Inr have been described previously (Roopra et al., 2000).

All constructs were verified by sequencing on an Applied Biosystems Model 377A DNA sequencer using BigDye Terminator chemistry.

Primers (5'-3')

M₁ -372s.BglII - GAAGATCTTTGGGCAACACTCAGACATG
M₁ -396s.XmaI - CCCCCCCGGCCAGATGCTGCCTGAGGCTTCCC
M₁ -602a.BglII - GAAGATCTGCAGTGTCTCCTGGGACC
M₁ +241a.BglII - GAAGATCTCACCACCTGCACACAGAC
M₁ +316a.BglII - GAAGATCTCATGAGGAAGGGAGGACC
M₁ +366a.BglII - GAAGATCTCCTGTCCACACAAGGGCACC
M₁ +417a.BglII - GAAGATCTGGGAAGCCTAGGAGCCAGCATCTGG
M₁ +417a - GGGAAAGGCTCAGGCAGCATCTGG
M₁ +504a.BglII - GAAGATCTGGCCCTGGCACGCCAGCCTTGG
fusion M₁Δ+412/+485s - CCCAGATGCTGCCTGAGGAGGCTGCCG
fusion M₁Δ+412/+485a - CCCTGGCAGCCAGGCTCTCATTCAGGCAG
fusion M₁ΔSp1s - CTGGGTGGGTGGGTGGGTGGGTGGCAG
fusion M₁ΔSp1a - GGCCTAGAAGCTGTAGGGCTGTAGGGCA
Inr.Hinds - AGCTTCTAGAGCGTTCGCTCCTCACTCCTTCCTCCGCGGATTAC
Inr.Hinda - AGCTTGTAAACCGCGGGAAGAGAGTGAGGACGAACGCTCCTAG
Transient transfections - Qiagen drip column purified DNA was transfected into cells using Tfx™ 50 (Promega) according to manufacturer’s instructions. Briefly, cells were plated onto 10mm wells to a density of 50%. For M1-reporter transfections, 750 ng of reporter plasmid, 3 ng pRL-CMV (Promega) and 1.7 µl Tfx™ 50 were mixed and made up to a total volume of 200 µl with OptiMEM (Life Technologies Inc.) and overlaid on cells for 1 hour. Cells were fed with the appropriate media and harvested 2 days later into 100 µl Passive Lysis Buffer (Promega) of which 30 µl was used in Promega’s Dual-Luciferase™ Reporter (DLR™) assay system. Ga4-reporter transfections were carried out by transfecting IMR32 and 3T3 cells 750 ng of plasmid (for amounts of specific plasmids see Figure legends), 3 ng pRL-CMV and 1.7 µl Tfx™ 50 in a total volume of 200 µl with OptiMEM, which was applied to cells for 3-4 hours. For Neuro2a cells, 250 ng of total plasmid (for amounts of specific plasmids see Figure legends), 1 ng pRL-CMV and 0.57 µl Tfx™ 50 were used. Cells were
harvested 24 hours after transfection into 60 μl Passive Lysis Buffer (Promega) of which 30 μl was used in Promega's DLR™ assay system. Luminescence was measured using either a Turner TD-20E or Mediators PhL 1.8 luminometer, Firefly luciferase was normalised to Renilla luciferase and the results expressed relative to Renilla normalised luminescence driven from the promoterless pGL3 Basic or as a percentage of normalised expression in the presence of pMT G4. For transfections in the presence of Trichostatin-A (TSA) (Wako Chemical), cells were treated with the indicated concentrations of TSA for 24 hours prior to transfection, and fed with media containing TSA for 24 hours. Results were expressed as Renilla normalised luminescence, relative to expression with Gal4 alone and expressed as fold over cells untreated with TSA.

*Gel Mobility Shift Assays* - Nuclear protein was extracted from cell lines as described by Dent and Latchman (Dent and Latchman, 1993) and the concentration determined by a DC protein assay kit (Bio-Rad). The oligos M₁Sp1s 5'-GTACGGTGGTGGCTGGGT and M₁Sp1a 5'-GTACCAGCCCACC-CACC were annealed and labelled by Klenow fill-in using (γ-32P) dATP. For single stranded DNA binding experiments, oligonucleotides M₁ +412s (nucleotides from +412 to +485 of the M₁ gene), M₁ +486s (nucleotides from +486 to +569) and M₁ +569a (nucleotides from +569 to +486) were radioactively labelled using T4 polynucleotide kinase and (γ-32P) ATP. ETS G4 (see below) probe was also radiolabeled using T4 polynucleotide kinase. For all probes, 1-5 μg of nuclear protein was incubated on ice, with or without competitor DNA, for 20 min in 19 μl of binding buffer containing 20 mM HEPES (pH7.9), 100 mM KCL, 5 mM MgCl₂, 8% glycerol and 1 μg calf thymus DNA. Approximately 20,000 cpm of radioactive probe was added to each reaction and incubation continued at room temperature for 20 min. For super shift experiments, 1μl of Sp1 antibody (1 mg/ml, Santa Cruz) was added to the reactions and incubation continued for a further 20 min.
Reactions were run on 0.25 X Tris borate/EDTA 4% polyacrylamide gels at 150V for 1-2 hours, dried and exposed to X-ray film. For gel shifts including ETS-1 G4 probe, 20 ng of IMR32 nuclear protein were incubated on ice for 20 min, with or without competitor, in a 19 µl reaction containing 10 mM Tris, pH 7.4, 100 mM NaCl, 1 mM EDTA, 100 mg/ml bovine serum albumin and and 1 µg calf thymus DNA. Approximately 20,000 cpm of radioactive probe was added to each reaction and incubation continued at room temperature for 20 min. Glycerol was added to a final concentration of 5% (w/v), and complexes were resolved by gel electrophoresis on 0.5X Tris borate/EDTA, 6% polyacrylamide gel at 4°C at 70 V for 16 hours.

**Formation of four-stranded G4 DNA** - Formation of G4 was carried out as described by Hanakahi *et al.* (Hanakahi *et al.*, 1999). Briefly, a synthetic oligonucleotide corresponding to a region of the external transcribed spacer (ETS-1), 5'-TCTCTCGGTGGCCGGGGCTCGGGGGTTTTGGGTC-CGTCCC was incubated at 1 mg/ml in TE containing 1 M NaCl at 60°C for 48 hours. After incubation, samples were diluted 1:5 with 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 12.5 mM KCl, and 2.5% glycerol, and DNA resolved on an 0.5X Tris borate/EDTA, 8% polyacrylamide gel, run in 0.5 X Tris borate/EDTA containing 10 mM KCl at 4°C at 70 V for 15 hours. Bands corresponding to G4 DNA were identified by UV shadowing and excised. DNA was eluted from crushed gel slices by soaking in TE containing 50 mM NaCl and 20 mM KCl at room temperature for 8-12 hours, precipitated with ethanol, washed, and resuspended in TE containing 200 mM NaCl.

**Southwestern analysis** - Fifteen µg of nuclear protein was run on a 6% SDS-PAGE gel and blotted onto Hybond C+ Extra nylon membrane (Amersham Pharmacia Biotech). Proteins were renatured by incubating the filters in binding buffer (20 mM HEPES (pH7.9), 100 mM KCL, 5 mM MgCl2, 8% glycerol) followed by a series of 2 fold dilutions of guanidine
hydrochloride for 10 min starting at 6 M through to 0 M, blocked with binding buffer containing 3% nonfat milk and probed with radioactive probes at 1x10^6 dpm/ml (Vinson et al., 1988).

**UV crosslinking** - Gel mobility shift reactions were assembled as described previously and subjected to 60 min of UV exposure from an inverted transilluminator (Cooney et al., 1993). Samples were electrophoresed through a 10% SDS-PAGE gel which was dried and exposed to X-ray film.

**S1 nuclease sensitivity assay** - Twenty µg of the indicated plasmids were digested with 1 or 10 units of S1 nuclease (Promega) at 37°C for 15 min in a 100-µl reaction containing 50 mM sodium acetate, pH 4.5, 280 mM NaCl, 4.5 mM ZnSO₄. The reaction was stopped by phenol extraction and the DNA was precipitated. The DNA was then digested with SalI. In parallel reactions, 20 µg of DNA was first cut with SalI and then digested with 10 units of S1 nuclease. Equal amounts of DNA were separated on 1% agarose gels, stained with ethidium bromide and visualised with UV light.

**Single stranded mapping** - 1 µg of supercoiled pGL3 basic or pGL3 +396/+569 M₁ was partially digested with 0, 10 or 20 units of mung bean nuclease (Roche) for 15 min at 37°C in a 100 µl reaction containing 50 mM sodium acetate, pH 4.5, 280 mM NaCl, 4.5 mM ZnSO₄. Reactions were stopped with 10 µl of 10 µM EDTA, phenol/chloroform extracted, ethanol precipitated and resuspended in 30 µl of water. One tenth of each reaction was run in a 1% agarose gel to examine the extent of digestion and the samples containing mainly nicked DNA and no linear DNA was used for subsequent steps. In a parallel reaction, 1 µg of pGL3 +396/+569 M₁ was digested with either NcoI or ClaI (Roche), phenol/chloroform extracted, ethanol precipitated and resuspended in 30 µl of water. Positions of the cuts on the coding and template strand were mapped by primer extension.
using GL2 and RV3 primers (Promega), respectively. Primers were labelled using T4 polynucleotide kinase and (γ-32P) ATP and 1 pmol (5x10^6 dpm) was added to a reaction mix containing 14 μl of each digestion, 100 mM each of dATP, dTTP, dGTP and dCTP, 10 mM KCl, 20 mM Tris-HCl, pH 8.8, 10 mM (NH4)2SO4, 2 mM MgSO4, 0.1% Triton X-100 in a final volume of 18 μl. Reactions were incubated for 1 min at 95°C, 15 min at 55°C and then 2 units of Vent DNA polymerase (New England BioLabs) were added and incubated for 10 min at 76°C. The reactions were stopped on ice and the DNA precipitated. Samples were resuspended in 7 μl formamide-dye and 75,000 dpm of each reaction were run on a 4% polyacrylamide, 7M urea gel at 60 W for 1 hour 20 min. The gel was dried and expose to X-ray film overnight.

**Lambda ZAP II Library screening** - A rat hippocampus λcDNA expression library (Stratagene) was plated following manufacturers instructions with some modifications. Briefly, the host strain XL1-Blue MRF' was infected with the library and plated on 5 plates (24x24 cm) to a density of 30 plaque forming units (pfu)/ cm². Expression of the library and preparation of nitrocellulose filters were performed essentially as described by Vinson et al. (Vinson et al., 1988) with some modifications. Plates were incubated at 42°C for 4 hours to stimulate formation of plaques. Plates were then overlaid with 10 mM IPTG-impregnated Hybond-C Extra filters (Amersham Pharmacia Biotech) and incubated at 37°C for an additional 9-10 hours. Duplicate filters were prepared by overlaying a second IPTG-impregnated filter onto the plates after removal of the first filter and incubated for a further 4 hours at 37°C. Filters were left onto the plates at 4°C overnight and then removed. All subsequent steps were carried out at 4°C. Expressed proteins were denatured by incubating filters in 250 ml of binding buffer (same as for gel shift assay) containing 6M guanidinium hydrochloride and renatured as described under
"Southwestern analysis". Filters were blocked by incubating overnight with binding buffer containing 5% non-fat milk and then equilibrated with binding buffer for 5 min. Filters were probed for 2 hours with 100 ml of binding buffer containing 10^6 dpm/ml of ^32P-labelled oligonucleotide M1+412s (same probe as for gel shift assay). Excess probe was removed by three 15 min incubations with 250 ml of binding buffer. Filters were exposed to X-ray film for 3 to 6 hours at -8°C. Plaques which showed signal (30 plaques) were isolated, re-plated on 10 cm plates to a density of 10 pfu/cm and re-screened. Clones 3, 61 and 62, the only positives from the second screen, were subjected to a tertiary screening to ensure homogeneity. Three single plaques from each clone were isolated and the cDNAs excised as pBluescript phagemid by co-infection with ExAssist Helper Phage and rescued in the nonsuppressing strain SOLR following manufacturers indications (Stratagene). Sequencing was performed using BigDye Terminator chemistry on an Applied Biosystems Model 377A DNA sequencer.

**Yeast one-hybrid screening** – pBM2389 +417/+166 M1 was transformed into the yeast strain SFY526 (Bartel et al., 1993). This yeast strain was then transformed with DNA from an adult rat brain cDNA yeast expression library (Clontech), using the protocol of Schiestl and Gietz (Schiestl and Gietz, 1989), and transformants were grown on complete supplement mixture lacking histidine, leucine, and tryptophan (-His/-Leu/-Trp) (Bio 101, Vista, CA, U.S.A.), containing 5 mM 3-amino-1,2,4-triazole (3AT) (Sigma) to select for interactions. Library candidates were tested for their ability to specifically activate the M1 containing reporter plasmid by re-transforming back into SFY526. Library plasmids producing interacting proteins were sequenced for identification.

**Immunoprecipitation assay** – Neuro2a cells were plated onto 10 cm plates to a density of 50%. Cells were incubated for 3-4 hours with 10 μg of
DNA and 22.5 μl Tfx™ in a final volume of 4.8 ml with OptiMEM. Cells were harvested after 2 days into 1 ml of 1xPBS containing 0.5% Nonidet P40 and protease inhibitors Block (Roche), sonicated for 90 s and centrifuged at maximum speed for 10 min at 4°C. Supernatant was precleared for 2 hours at 4°C with 40 μl of Protein G Sepharose (Pharmacia Biotech). For each immunoprecipitation, half of the total sample was incubated with Gal4 DBD antiserum (Santa Cruz Biotechnology) overnight at 4°C. Beads were added and samples were incubated for a further 2 hours. Samples were washed 4 times with 20 mM Tris pH 8.0, 1 mM EDTA, 100 mM NaCl, 0.5 Nonidet P40, 10% glycerol and 0.1% SDS. Proteins were eluted with 15 μl of loading dye. Samples were run on a 10% SDS-PAGE gel and blotted onto Hybond C+ nylon membrane (Amersham Pharmacia Biotech) and results were visualized using c-myc antiserum (Santa Cruz Biotechnology).

**Binding site selection** - The random oligonucleotide selection was performed according to Pollock and Treisman protocol (Pollock and Treisman, 1990) with a few modifications. Briefly, a 62-mer sequence (5'-CAGGTCAGGAATTCTTCAGCTGCGNNNNNNNNNNNGAGGCAGTGCAACGAATrC-GCAGC) with a central random stretch of 12 base pairs was used as a probe. This sequence was duplexed by incubation for 20 min at 37 °C with 20 pmols of the primer R62 (5'-GCTGCAGAATTCGTTGCACTGCCTC), 10 units Klenow fragment (New England BioLabs), 10 mM Tris, HCl, pH 7.5, 5 mM MgCl₂, 7.5 mM DTT, and 0.5 mM dNTP in a 20 μl reaction. Double stranded DNA was gel purified using a 8% acrylamide/bisacrylamide gel, ethanol precipitated and resuspended in 10 μl of TE, of which one hundredth was used in a first round of site selection. For this, 1-5 mg of nuclear protein from IMR32 or Neuro2a cells transfected with either pMT SHARP-1 or pSHARP-1 myc were incubated with the double stranded oligonucleotide in a 25 μl reaction containing 20 mM Hepes, pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.1% Nonidet P40, 0.5 mM DTT, 1 mg/ml BSA, 40 ng/μl dI:C
and proteinase inhibitors Block (Roche) for 30 min in ice, and then added to 15 µl of blocked Protein G Sepharose (Pharmacia Biotech) pre-incubated with c-myc antibody, and left to incubate for a further 1 hour at 4°C. Beads were washed twice with 250 µl of the pre-cold buffer containing 20 mM Hepes, pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.1% Nonidet P40, and protein with its bound DNA eluted at 45°C for 2 h using 200 µl of buffer containing 5 mM EDTA, 0.5% SDS, 100mM NaOAc and 50 mM tris, pH 8. The eluted DNA was extracted with phenol/chloroform followed by ethanol precipitation. One tenth of the resulting DNA was amplified by 16 cycles of PCR using primers corresponding to the linker sequences, gel purified and resuspended in 10 µl of TE, of which 1 µl was directly used as a probe for the next round of binding site selection. In total, five selection cycles were performed. After PCR amplification, DNA from the 5th round was digested with EcoRI, ethanol precipitated, and radiolabelled by Klenow fill in. About 30.000 dpm were used in a gel shift assay following the same protocol as described under “Gel shift binding assays” using the same nuclear protein extracts as for the site selection cycles. DNA:protein complexes were cut out from the gel, eluted and DNA was amplified by 20 cycles of PCR, cloned into pGemT-easy (Promega) and sequenced. Sequences were analysed using a non-gapped algorithm (ClustalW).
CHAPTER 3

RESULTS (PART 1)
RESULTS (PART 1)

INTRODUCTION

In order to identify the mechanisms that regulate expression of the M1 muscarinic acetylcholine receptor gene, a cosmid clone containing the M1 gene and flanking regions has previously been isolated (Pepitoni et al., 1997). To date, the gene structures for three other members of the muscarinic receptor gene family have also been identified; the rat M4 (Mieda et al., 1996; Wood et al., 1995), the chicken M2 (Rosoff et al., 1996) and the chicken M3 (Creason et al., 2000). All of these genes appear to have a similar gene structure consisting of a single coding exon and one or more 5' exons containing the 5' untranslated region (5'UTR). The rat M1 gene consists of a single coding exon of 2.54 kb and a single non-coding exon of 657 bp separated by an intron of 13.5 kb (Pepitoni et al., 1997; Wood et al., 1999) (Figure 1). RT-PCR analysis suggests a similar structure for the human M1 gene, but indicates the use of an alternatively spliced exon in mouse cortex (Wood et al., 1999).

In common with several other members of the GPCR gene family, including the M4 (Mieda et al., 1996; Wood et al., 1995), V1a vasopressin (Murasawa et al., 1995) and D1a dopamine (Minowa et al., 1992; Zhou et al., 1992), the M1 gene lacks a TATA or CAAT box (Figure 1). Inspection of 2.2 kb of the sequence of the M1 gene around of the transcription start site reveals a number of sequence motifs that resemble consensus binding sites for different transcription factors, such as AP-1, NZF-1, AP-2, Oct-1 and NFκB (Figure 1) (Pepitoni et al., 1997). In common with the M4 gene (Wood et al., 1996), the M1 gene contains a GC box in the vicinity of the transcription start site. Most of these sequence motifs are very small, of only 6 to 10 base pairs, and they may randomly appear throughout the
Figure 1 - Organisation of the rat M₁ muscarinic receptor gene. 
a) shows the relative sizes of the coding (Ex2) and non-coding (Ex1) exons. 
b) shows an enlargement of the 5'-flanking sequences. 
B, BamHI; E, EcoRI; H, HindIII; K, KpnI; N, Ncol; P, PstI; S, SalI; Sa, SacI. 
c) shows the sequence of 1.6 kb of 5'-flanking region and the 657 bp non-coding exon. 
Positions of restriction sites are underlined. Positions of consensus regulatory elements Sp1, NZF1, AP1, AP2, E box, NFkB, Oct1 and 
of a polypyrimidine/polypurine tract are boxed. Transcription start site is indicated by an asterisk and is assigned position +1. 
5'-flanking sequence defined from assigning the guanine at +1 is in uppercase while the non-coding exon is shown in lowercase. 
Single stranded nucleotides identified by mung bean nuclease digestion and primer extension in the coding and template strand are indicated by arrows pointing down and up, respectively.
genome with high frequency. For example, a 6 base pair sequence is present on average every 4 kb, so in itself, identification of such motifs cannot be considered proof of a bona fide regulatory site. Such proof requires a functional analysis of their effect on M1 transcription. In addition to an examination of these recognisable motifs, identification of any novel sites important for expression of the M1 gene must also be considered.

In this chapter, data from experiments performed to identify the minimal region of the M1 gene necessary to drive cell specific expression are presented.

RESULTS

A neuroblastoma cell line expresses the endogenous M1 gene – PCR analysis on cDNA obtained from four different cell lines showed that one of these cell lines, IMR32 (human Caucasian neuroblastoma), expressed the M1 gene (Figure 2). This was used as a model cell line for M1 expression. Two neuroblastoma cell lines, Neuro2a (mouse neuroblastoma) and NB4 1A3 (mouse C-1300 neuroblastoma), and one non-neuronal cell line, 3T3 (mouse swiss albino embryo fibroblast), that did not express the M1 gene were also identified (Figure 2). As M1 gene expression in vivo is restricted to a subset of neuronal cells (Bonner et al., 1987; Buckley et al., 1988), both non-expressing neuroblastoma and non-neuronal cells were used in order to distinguish potential transcriptional mechanisms restricting expression to the nervous system from those directing expression to particular subsets of neuronal cells.
RNA from IMR32, 3T3, Neuro2a and NB4-1A3 cells was extracted and reverse transcribed as described under “Materials and Methods”, and the resultant cDNA was amplified using M₁ and hypoxanthine-guanine phosphoribosyl transferase (HPRT) specific primers. PCR products were electrophoresed through a 2% MetaPhor agarose gel. Lane 1 shows 1kb DNA marker ladder (Life Technologies Inc) and lane 5 shows the water negative control. Lanes 1-4 show amplification with HPRT specific primers for all the cell lines and lane 1 shows the presence of M₁ transcripts only in IMR32 cells.

Figure 2 – Expression of the M₁ gene by different cell lines.
977bp of the rat M1 gene is capable of driving reporter gene expression - In order to identify the regulatory elements important for expression of the M1 gene, a region of the M1 gene between -1602 and +602 was cloned in front of the Luciferase reporter gene, and its ability to drive expression analysed in the different cell lines (Figure 3). This construct drove 4 fold expression relative to the empty vector, pGL3 Basic, in the M1 expressing cell line IMR32, but failed to drive significant expression above pGL3 Basic in the two neuronal cell lines used that do not express the M1 gene, Neuro2a and NB4 1A3. This same construct drove a low level of expression (2 fold over pGL3 basic) of the reporter gene in the non-neuronal cell line used, 3T3. Sequential deletions of the region between -1602 and -372, did not cause any significant effect on expression of the reporter gene (Figure 3). These data suggest that the region between -372 and +602 of the M1 gene is sufficient to drive expression in M1 expressing neuronal cell lines and to be silent in non-expressing neuronal cell lines.

A GC box can recruit Sp1 and at least one other DNA binding protein - The region of the M1 gene between -372 and +605 contains 3 sites that resemble consensus binding sites for known transcription factors (a GC box at +17 to +23; an Oct-1 site at +59 to +66; an NFκB site at +139 to +149) (Figure 1). In order to investigate the ability of the GC box to recruit proteins to the M1 gene, a gel mobility shift assay was performed. Figure 4 demonstrates that the region encompassing the GC box can form a single DNA:protein complex with nuclear extracts from IMR32 cells (complex "A"). This interaction can be competed by a specific, but not by a non-specific competitor. However, use of a Sp1 antibody did not result in a supershifted band, suggesting that the DNA binding protein is not Sp1. When this region of the M1 promoter was incubated with nuclear extracts from 3T3 cells, (Figure 4), a retarded band was also observed that could be
Figure 3 – Deletional analysis of the rat M₁ gene. IMR32, 3T3, Neuro2a and NB4 1A3 cells were transfected using 750 ng of the reporter plasmids represented on the left, which contain distinct fragments of the M₁ gene driving expression of the luciferase reporter gene. Bars express results of at least 3 independent experiments each performed in triplicate, normalised to Renilla luciferase activity and expressed as -fold over expression of the promoterless pGL3 basic vector.
Normalised luciferase activity relative to pGL3 basic
Figure 4 - Binding of nuclear proteins to the M₁ GC box.
Annealed oligonucleotides containing the GC box spanning from +10 to +30 relative to the transcription start site of the M₁ gene were labelled by Klenow fill in. Nuclear protein extracts from IMR32 (lane 1-4) or 3T3 (lanes 5-8) cells were incubated with either specific competitor (lanes 2 and 6), non-specific competitor (lanes 3 and 7) or Sp₁ antibody (lanes 4 and 8) for 20 min on ice.³²P-labelled probe was added and incubated for a further 20 min at room temperature. DNA and DNA:protein complexes were resolved on 4% non-denaturing polyacrylamide gel.
977bp of the rat M1 gene is capable of driving reporter gene expression. In order to identify the regulatory elements important for expression of the M1 gene, a region of the M1 gene between -1602 and +602 was cloned in front of the Luciferase reporter gene, and its ability to drive expression analysed in the different cell lines (Figure 3). This construct drove 4 fold expression relative to the empty vector, pGL3 Basic, in the M1 expressing cell line IMR32, but failed to drive significant expression above pGL3 Basic in the two neuronal cell lines used that do not express the M1 gene, Neuro2a and NB4 1A3. This same construct drove a low level of expression (2 fold over pGL3 basic) of the reporter gene in the non-neuronal cell line used, 3T3. Sequential deletions of the region between -1602 and -372, did not cause any significant effect on expression of the reporter gene (Figure 3). These data suggest that the region between -372 and +602 of the M1 gene is sufficient to drive expression in M1 expressing neuronal cell lines and to be silent in non-expressing neuronal cell lines.

A GC box can recruit Sp1 and at least one other DNA binding protein. The region of the M1 gene between -372 and +605 contains 3 sites that resemble consensus binding sites for known transcription factors (a GC box at +17 to +23; an Oct-1 site at +59 to +66; an NFκB site at +139 to +149) (Figure 1). In order to investigate the ability of the GC box to recruit proteins to the M1 gene, a gel mobility shift assay was performed. Figure 4 demonstrates that the region encompassing the GC box can form a single DNA:protein complex with nuclear extracts from IMR32 cells (complex “A”). This interaction can be competed by a specific, but not by a non-specific competitor. However, use of a Sp1 antibody did not result in a supershifted band, suggesting that the DNA binding protein is not Sp1. When this region of the M1 promoter was incubated with nuclear extracts from 3T3 cells, (Figure 4), a retarded band was also observed that could be
Figure 3 – Deletional analysis of the rat M_{1} gene. IMR32, 3T3, Neuro2a and NB4 1A3 cells were transfected using 750 ng of the reporter plasmids represented on the left, which contain distinct fragments of the M_{1} gene driving expression of the luciferase reporter gene. Bars express results of at least 3 independent experiments each performed in triplicate, normalised to Renilla luciferase activity and expressed as -fold over expression of the promoterless pGL3 basic vector.
Figure 4 - Binding of nuclear proteins to the M1 GC box. Annealed oligonucleotides containing the GC box spanning from -160 to +60 relative to the transcription start site of the M1 gene were labelled by Klenow fill in. Nuclear protein extracts from IMR32 (lanes 1-4) or 3T3 (lanes 5-8) cells were incubated with either specific competitor (lanes 2 and 6), non-specific competitor (lanes 3 and 7) or Sp1 antibody (lanes 4 and 8) for 30 min on ice. 32P-labelled probe was added and incubated further 20 min at room temperature. UVa and IMR32 protein complexes were formed.

- IMR 32
- 3T3
- N2a

Normalised luciferase activity relative to pGL3 basic
Figure 4 - Binding of nuclear proteins to the M₁ GC box.
Annealed oligonucleotides containing the GC box spanning from +10 to +30 relative to the transcription start site of the M₁ gene were labelled by Klenow fill in. Nuclear protein extracts from IMR32 (lane 1-4) or 3T3 (lanes 5-8) cells were incubated with either specific competitor (lanes 2 and 6), non-specific competitor (lanes 3 and 7) or Sp1 antibody (lanes 4 and 8) for 20 min on ice. $^{32}$P-labelled probe was added and incubated for a further 20 min at room temperature. DNA and DNA:protein complexes were resolved on 4% non-denaturing polyacrylamide gel.
<table>
<thead>
<tr>
<th>Protein</th>
<th>IMR32</th>
<th>3T3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 Sp1 competitor:</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Non specific competitor:</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sp1 antibody:</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Image of gel electrophoresis](image.png)
competed only by specific competitor and in this case, the band was supershifted by Sp1 antibody demonstrating that this DNA:protein complex did contain Sp1. UV crosslinking of proteins bound to the GC box of the M<sub>i</sub> gene and SDS-PAGE analysis, (Figure 5), showed the protein responsible for the DNA:protein complex “A” from IMR32 cells in Figure 4 was approximately 20-25 kDa in size. UV crosslinking of DNA:protein complexes from 3T3 cells showed a single specific protein band of approximately 100 kDa, a size consistent with the protein being Sp1.

**Removal of the GC box does not have any discernible effect on transcription** - To test the functional relevance of the GC box in the M<sub>i</sub> gene, 7 bp encompassing the GC box (Figure 1) were deleted from the M<sub>i</sub> promoter (see pGL3 -372/+602 M<sub>i</sub> ΔSp1 in Figure 3). In transient transfection analysis, no significant change in the expression of the reporter gene was observed in any of the cell lines tested (compare pGL3 -372/+602 M<sub>i</sub> and -372/+602 M<sub>i</sub> ΔSp1 in Figure 3). Hence, although the GC box of the M<sub>i</sub> gene is able to bind Sp1 and at least one other protein in 3T3 and IMR32 cells respectively *in vitro*, this binding does not appear to effect the expression of the gene in transient transfection assays.

**DISCUSSION**

The human and rat M<sub>i</sub> genes contain a single coding exon and a single 5’exon containing the 5’UTR (Figure 1 and (Pepitoni et al., 1997; Wood et al., 1999)). In mice, an alternatively spliced non-coding exon has been identified (Wood et al., 1999). In a transient transfection analysis, the region between -372 and +602 of the M<sub>i</sub> gene has been found to be sufficient to drive expression in a M<sub>i</sub> expressing cell line (IMR32) and to silence it in the neuronal M<sub>i</sub> non-expressing cell lines (Neuro2a and NB4
Figure 5 - Sp1 crosslinking. Gel mobility shift reactions with M₁ Sp1 probe were assembled as described in Figure 4 and subjected to 60 min of UV exposure from an inverted trans-illuminator. DNA:protein complexes were resolved in a 10% SDS-polyacrylamide gel. Cell nuclear protein extracts and the use of competitors is indicated in the top of the figure.
<table>
<thead>
<tr>
<th>Protein: M₁ Sp1 competitor:</th>
<th>IMR32</th>
<th>3T3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

- 200 kDa
- 97.6 kDa
- 69 kDa
- 46 kDa
- 30 kDa

Sp1

A
1A3) (Figure 3). However, low levels of reporter gene expression are still observed in the non-neuronal M, non-expressing cell line (3T3) is not absolutely silent. One possible explanation is that the element(s) that silence expression of the M, gene in 3T3 cells are not present in any of the constructs used in this study, but exists further upstream or downstream. There are many examples of distal silencing elements, for instance, silencing of L1-CAM gene expression is conferred by an RE-1 within the second intron (more than 10kb downstream of the transcription start site) and by sequences in the first intron (Kallunki et al., 1997). Alternatively, the expression seen in 3T3 may be an artefact of transiently transfected DNA. In some cases, a region containing the promoter and regulatory elements identified by transient transfection assays in cell lines for a particular gene appears to be sufficient to drive tissue specific expression of a reporter gene in transgenic mice. However, this is not always the case, as has been shown for the 68-kDa neurofilament (NF-L) gene. It has been found that 1.7 kb of sequence 5' to the transcription start site of the NF-L gene is capable of driving expression of a reporter gene in a tissue specific manner in transgenic mice, but is active in both NF-L expressing and non-expressing cell lines in transient transfection assays (Nakahira et al., 1990). Although transient transfection assays represent a very simple and useful protocol to identify regulatory elements, they have a number of limitations. One potential problem is the number of copies of cis-elements introduced into the cell by transfection. Due to a high copy number of transfected elements, transcription factors may become limiting and only partial effects observed. Reporter gene expression driven by the 1.7 kb fragment detected in NF-L non-expressing cells by transient transfection assays could be due to a de-repression effect due to limiting amounts of a factor required to repress transcription in these cells. However, stable cell lines produced using the same construct and
containing only one copy of the reporter gene integrated into chromosomal DNA, still showed ectopic expression of the reporter gene (Nakahira et al., 1990). In this case, it is possible that expression of the NF-L gene is controlled by selective de-activation of its promoter during development, which may involve modifications of the DNA such as methylation and occur only during a certain stage of development. Therefore, introducing a promoter that has not been de-activated, into cell lines that have already passed this stage of development, will allow the promoter to drive expression of its gene, or a reporter gene, in a non-specific manner. It will thus be important to analyse any putative regulatory region identified by transient transfection in the context of transgenic mice.

Within the region between -372 and +602 of the $M_1$ gene, there is a GC box at position +17 to +23, which is able to bind Sp1 in 3T3 cells and a 20-25 kDa protein in IMR32 cells (Figure 4 and 5). However, this element does not appear to be required for the expression of the $M_1$ gene in transient transfection assays in the cells analysed (Figure 3). Sp1 plays an important role in the control of cell cycle-regulated genes and in the transcriptional regulation of tissue specific and ubiquitous genes (reviewed in (Lania et al., 1997)). It is likely that Sp1 acts in combinatorial manner with other transcription factors, so the possibility of the existence of an additional factor, with a binding site in a region of the $M_1$ gene further upstream or downstream of the fragment analysed, acting in conjunction with Sp1 and/or the other factor identified to bind the GC box cannot be ruled out. In the case of the $M_4$ gene, a tandem of three GC boxes was found to bind several proteins, including Sp1, but no effect was observed when the GC box tandem was deleted (Wood et al., 1996). The potential functional role of the Oct-1 and NFkB binding elements present in the $M_1$ gene remains to be determined. NFkB is found in the cytoplasm...
bound to the inhibitory molecule IκB. When IκB is phosphorylated, it releases NFκB, which moves to the nucleus to function as a transcription factor (May and Ghosh, 1997). Diverse stimuli produce the release of NFκB by IκB in different tissues, including heat shock, ionising radiation, lipopolysaccharide and cytokines (Basu et al., 1998; Deshpande et al., 1997; Maraldi et al., 1997; Pogliaghi et al., 1995). It is not yet determined whether the NFκB element binds NFκkB or whether this element has any functional role in regulating M₁ expression. However, the construct pGL3 -372/+169 M₁ which contains the GC box, Oct-1 and NFκB sites did not drive expression of the reporter gene in the M₁ expressing cell line, IMR32 (See Chapter 4 and Figure 6a), suggesting that these elements do not have an effect on M₁ transcription.
CHAPTER 4

RESULTS (PART 2)
RESULTS (PART 2)

INTRODUCTION

In prokaryotes, DNA sequences within individual genes that are essential for either basal or regulated gene expression are found immediately upstream of the transcription start site (for review see (McClure, 1985)). Specific σ factors are able to recognise DNA elements present in specific promoters (Burgess et al., 1969; Grossman et al., 1984; Haldenwang and Losick, 1980) allowing the positioning of the RNA polymerase complex (α2ββ'σ) at the transcription start site. Initially, it was thought that a similar mechanism would be at play at eukaryotic promoters, therefore, early studies of cloned eukaryotic genes concentrated on the regions immediately upstream of the transcription start site, where by analogy, sequences involved in transcription and its regulation should be located. Most of the cis-acting elements identified in eukaryotes map in the 5'-flanking gene region. However, an increasing number of regulatory elements are being identified in other gene regions. For example, nuclear DEAF-1-related (NUDR) protein represses transcription of the hnRNP A2/B1 and hNUDR8 genes through DNA binding sites located 3' proximal to the transcription start site (Michelson et al., 1999). Furthermore, this repression is dependent on the position of the NUDR binding sites downstream of the transcription site, since NUDR was able to repress transcription when the NUDR sites where cloned downstream from the thymidine kinase (tk) promoter, but was unable when the sites where cloned upstream (Michelson et al., 1999). Another example is seen with the collagen type IV (COL4A2) gene, where a potent silencer has been characterised in the third intron (Haniel et al., 1995). Since deletions within the 5'-flanking region of the M1 gene did not appear to give any
changes in expression and considering the increasing number of genes containing regulatory sequences outside of the 5' flanking region, we decided to look at the sequences within the non-coding exon.

RESULTS

Two regions highly conserved across species appear to be important for cell specific expression of the M₄ gene. Sequence analysis of the region within the non-coding exon of the M₄ gene did not identify any consensus element for DNA binding elements, except for the presence of a stretch of pyrimidines in the coding strand (nucleotides +412 to +485). Sequence comparison between human, mouse, rat (Wood et al., 1999) and pig (Kubo et al., 1986a) showed extensive homology for the human, mouse and rat genes within the polypyrimidine/polypurine (PPY/PPU) tract. Furthermore, the region immediately downstream of the PPY/PPU tract (nucleotide +486 to +602) was also conserved in pig. Any region of the M₄ gene that is important for functional expression would be expected to be conserved across species. In order to test the functional significance of these regions, internal deletions of the 5' UTR were made and their effects on reporter activity determined (Figure 6a). Deletion of the region between nucleotides +505 to +602, (to give pGL3 -372/+504 M₄), that is conserved between all 4 species results in approximately 50% decrease of luciferase activity in IMR32 cells with respect to pGL3 -372/+602 M₄ and a decrease of 17% in 3T3 cells. This deletion does not appear to affect the levels of luciferase activity in Neuro2a or NB4 1A3 cells. Deletion of the PPY/PPU tract between nucleotides +412/+485, (to give pGL3 -372/+602 ΔPPY M₄), results in a decrease of approximately 60% and 20% in the levels of luciferase activity in IMR32 and 3T3 cells, respectively, with respect to pGL3 -372/+602 M₄, but does not greatly affect
Figure 6 – Deletional analysis within the non-coding exon of the \( M_1 \) gene. Cells were transfected with 750 ng of the reporter plasmids represented on the left. *Bars and numbers* express results of at least 3 independent experiments each performed in triplicate, normalised to *Renilla* luciferase activity and expressed as –fold over expression of the promoterless pGL3 basic vector (a) or pGL3 Inr (b). a) Deletions within the non-coding exon of the \( M_1 \) gene were produced, and the ability of the \( M_1 \) promoter to drive expression of the luciferase gene measured. b) Fragments identified to be important for expression driven by the \( M_1 \) promoter were cloned in front of the Adenovirus Major Late promoter (Inr) and the ability to control expression of the luciferase reporter gene by this promoter analysed.
a) Normalised luciferase activity relative to pGL3 Basic

- pGL3 -372/+602 M1
- pGL3 -372/+504 M1
- pGL3 -372/+602 APPY M1
- pGL3 -372/+417 M1

b) Normalised luciferase activity relative to pGL3 Inr

- pGL3 +396/+569 M1 Inr
the luciferase activity in the neuronal \( M_i \) non-expressing cell lines. Deletion of both regions together (see pGL3 -372/+417 \( M_i \)), does not lead to any additional fold increase in repression in IMR32 or 3T3 cells. Taken together, these data suggest that the pyrimidine rich region of the \( M_i \) gene and the adjacent region (conserved between human, mouse, rat and pig) contribute to the expression of the \( M_i \) gene and appear to act selectively in the \( M_i \) expressing cell line, IMR32. Moreover, results suggests that the PPY/PPU tract and the conserved region act as a unique enhancer \( cis \)-element, and that the integrity of the whole element is required to drive full levels of expression.

**The \( M_i \) putative transcription elements can activate heterologous promoters** - As all the deletions that affect expression of the reporter construct occur downstream of the transcription start site (Figure 6a), it is possible that levels of reporter gene expression driven by the \( M_i \) promoter reflect differences in stability of the RNA transcripts. However, if the elements identified in transient transfection assays work by recruiting transcription factors rather than affecting mRNA stability, we reasoned that they should be able to regulate a heterologous promoter. Accordingly, the region of the \( M_i \) gene between +396 and +569 was cloned upstream of the Adenovirus Major Late promoter (AdMLP) Initiator (Inr) (Pengue and Lania, 1996) to test its effect on a heterologous promoter. The AdMLP was chosen because, like the \( M_i \) gene (Pepitoni et al., 1997), it contains a TATAless promoter. The ability of this fragment to enhance the expression of the Inr promoter was measured by transient transfection analysis. Figure 6b shows the results obtained as fold activity over pGL3 Inr. All cell lines show similar activity for pGL3 Basic. The fragment between +396 and +569, containing the PPY/PPU tract and the conserved region, enhanced reporter gene expression by 25 fold and 6 fold in IMR32 and 3T3 cells, respectively. These results suggest the possibility that the
factor(s) that activates transcription in the M_i expressing cell line IMR32 may be present in 3T3 cells, but in lower concentration, or showing a lower affinity due to the presence of different cofactors. Three-fold activation was seen with the fragment between +396 and +569 in the two neuronal M_i non-expressing cell lines Neuro2a and NB4 1A3. In summary, these results show that the PPY/PPU tract and the conserved region act as enhancing elements, suggesting that changes of reporter expression observed by deletions of the M_i promoter are not due to changes in mRNA stability.

The PPY/PPU tract is sensitive to S1 nuclease – PPY/PPU tracts are found in many promoters forming a H-DNA structure composed of intramolecular triple helix and single stranded DNA required for its regulatory function (Goodridge and Xu, 1996; Potam et al., 1996). In order to check if the PPY/PPU tract in the M_i gene is able to form this non-B-DNA structure, the sensitivity of this element to S1 nuclease, an enzyme that preferentially digests single stranded DNA, was examined. A fragment containing the PPY/PPU tract and the conserved region across species (nucleotides from +396 to +569) cloned into pGL3 basic was treated with S1 nuclease and then cut at a SalI site 2kb downstream of the M_i sequence (Figure 7a). Two bands of 2 and 3 kb were observed, consistent with the presence of single stranded DNA in the M_i fragment (Figure 7b, lanes 6 and 7). The intensity of both bands increased as the amount of S1 nuclease in the reaction was increased. These two bands did not appear when a control plasmid that lacked the M_i sequence was treated in the same way (Figure 7b, lanes 2 and 3). Neither the 2 kb nor the 3 kb bands were seen when the plasmid was first linearised with SalI, followed by digestion with S1 nuclease (Figure 7b, lane 8), indicating that the M_i PPY/PPU tract can form a single stranded structure in a supercoiled state but not when relaxed. Further experiments to see if both the PPY/PPU tract and the
Figure 7 - S1 nuclease sensitivity. a) Diagram representative of the digestion performed with S1 nuclease. b) Supercoiled pGL3 basic (lanes 1-4) or pGL3 +396/+569 M₁ (lanes 5-8) were either cut with SalI (lanes 1 and 5) or treated with increasing amounts of S1 nuclease before (lanes 2, 3, 6 and 7) or after digestion with SalI (lanes 4 and 8). c) pGL3 containing the M₁ sequence between either from +412 to +485 (lane 1) or from +486 to +569 (lane 2) was treated with 10 units of S1 nuclease and digested with SalI.
a) 

"Single stranded DNA"

b) 

<table>
<thead>
<tr>
<th>Reporter plasmid</th>
<th>pGL3 basic</th>
<th>pGL3 +396/+569 M₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 nuclease (units)</td>
<td>- 1 10 - 1 10 -</td>
<td>+ + + + + + + +</td>
</tr>
<tr>
<td>Sal I</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
</tr>
</tbody>
</table>

c) 

<table>
<thead>
<tr>
<th>pGL3 +412/+485 M₁</th>
<th>pGL3 +486/+569 M₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 10</td>
<td>+ +</td>
</tr>
</tbody>
</table>

\[ \text{5 Kb} \]
conserved region were able to form single stranded structure independently were performed (Figure 7c). A construct containing only the PPY/PPU tract (nucleotides from +412 to +485) again generated the two bands indicative of single stranded structure but a plasmid containing only the conserved region (nucleotides from +486 to +569) was unable to generate the 2 and 5 kb bands. Thus, the PPY/PPU tract, but not the conserved region, is able to form a supercoil-dependent single stranded structure.

The conserved region forms non B-DNA structure when it is next to the PPY/PPU tract - To further determine the extension of the region sensitive to single stranded DNA nucleases, a construct containing both the PPY/PPU tract and conserved region (pGL3 +396/+569 M1) and pGL3 basic were treated with increasing amounts of mung bean nuclease. This nuclease has a higher specific activity for single stranded DNA than S1 nuclease, and does not digest across nicks. Therefore, it is expected that any sequence that is not sensitive to S1 nuclease will not be sensitive to mung bean nuclease. Digestion conditions were adjusted to obtain only nicked DNA but not linearised DNA. The positions of the nicks were mapped by a primer extension method using primers flanking the M1 region (Figure 8). To control for the generation of partial extension products, plasmids were treated with a restriction enzyme that digested DNA at a known distance from the priming site in the absence of mung bean nuclease. The extension product of this positive control gave a single full-length product (Figure 9, lanes 7 and 14), indicating the absence of partial extensions. When supercoiled plasmid containing the M1 sequence was treated with increasing amounts of mung bean nuclease, a series of bands of increasing intensity were seen (Figure 9, lanes 4-6 and 11-13). Again, these bands did not appear with the control plasmid pGL3 basic, which lacks the M1 sequence (Figure 9, lanes 1-3 and 8-10). The position of
Figure 8 – Single stranded mapping protocol. Diagram representative of a digestion of supercoiled plasmid with mung bean nuclease followed by primer extension using radiolabelled primers.
Mung bean Nuclease

RV3 primer

GL2 primer

Vent DNA polymerase
**Figure 9 - Single stranded mapping.** Supercoiled plasmid containing both the PPY/PPU tract and the conserved region across species (pGL3 +396/+569 M₁) was digested with either *NcoI*, a restriction enzyme which digests downstream of the M₁ sequence (lane 7) or *ClaI*, enzyme with site upstream of the M₁ sequence (lane 14). Digested plasmid was used in a primer extension protocol with ³²P-labelled RV3 (lane 7) or GL2 (lane 14). Supercoiled pGL3 basic (lanes 1-3 and 8-10) or pGL3 +396/+569 M₁ (lanes 4-6 and 11-13) was treated with increasing amounts of mung bean nuclease for 15 min at 37°C and primer extended with either RV3 (lanes 1-6) or GL2 (lanes 8-13). Samples were resolved in a 4% polyacrylamide/7 M urea gel. *Arrows* and *numbers* indicate position and size of full-length product.
template strand strand (RV3 primer)  coding strand (GL2 primer)

<table>
<thead>
<tr>
<th>reporter plasmid</th>
<th>pGL3 basic</th>
<th>pGL3 basic</th>
<th>pGL3 +396/+569 M1</th>
<th>pGL3 +396/+569 M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>mung bean nuclease (units)</td>
<td>0 10 20</td>
<td>0 10 20</td>
<td>Nco I</td>
<td>0 10 20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>G</th>
<th>A</th>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

338 bp → 296 bp
the nucleotides, subsequently, the extension of the single stranded DNA region was determined (Figure 1 indicated by arrows) and showed that on the polypurine tract there is a single stranded region, not found on the opposite polypyrimidine strand. The single stranded region is not limited to the PPY/PPU tract, but extends both downstream into the conserved region and upstream into vector sequence. Thus, the PPY tract appears to influence the structure of the flanking sequence. In the context of the M<sub>i</sub> gene this would result in the partial opening of the downstream conserved region, which does not contain an autonomous single stranded structure (see above and Figure 7c).

**The conserved region and the polypyrimidine tract both bind single stranded DNA binding proteins** - Gel mobility shift assays were performed with regions of the M<sub>i</sub> promoter found to be functionally important for M<sub>i</sub> expression. A single stranded probe consisting of the polypyrimidine (PPY) tract was found to bind proteins from IMR32, 3T3 and Neuro2a cells (Figure 10a). This binding was specific for the single stranded probe and a probe synthesised from the antisense sequence did not show any DNA:protein interactions. Single stranded DNA probes from the conserved region between nucleotides +486 and +569 were also found to bind proteins from both IMR32, 3T3 and Neuro2a cells (Figure 10b and c). Sequences from both strands of this region were found to form specific DNA:protein interactions. All interactions were competed by a specific, but not by a non-specific competitor. A Southwestern analysis was performed to identify 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 Figure 11a, b and c. For each of these probes, 2 bands, of approximately 100 and 110 kDa, were identified using IMR32 protein and only a single band, of approximately 110 kDa using protein extracts from 3T3 and Neuro2a cells. Surprisingly each of the probes
Figure 10 - Gel shift assay with single stranded DNA. Single stranded oligonucleotides corresponding to a) the M₁ PPY tract (nucleotides from +412 to 485), b) the coding strand of the conserved region (nucleotides from +486 to +569) and c) the template strand of the conserved region (nucleotides from +569 to +486) were labelled using T4 polynucleotide kinase. For all probes, nuclear protein extracts from IMR32 (lanes 1-3), 3T3 (lanes 4-6) or Neuro2a (lanes 7-9) were incubated with no competitor (lane 1, 4 and 7), specific competitor (lanes 2, 5 and 8) or non-specific competitor (lanes 3, 6 and 9) for 20 min on ice. °P-labelled probe was added and incubated for further 20 min at room temperature. DNA and DNA:protein complexes were resolved on 4% non-denaturing polyacrylamide gel.
a) Protein: $M_1 +412s$
   competitor: non specific competitor:
   IMR32 3T3 Neuro 2a
   free probe
   PPY tract ($M_1 +412s$)

b) Protein: $M_1 +486s$
   competitor: non specific competitor:
   IMR32 3T3 Neuro 2a
   free probe
   Conserved region ($M_1 +486s$)

c) Protein: $M_1 +569a$
   competitor: non specific competitor:
   IMR32 3T3 Neuro 2a
   free probe
   Conserved region ($M_1 +569a$)
Figure 11 - Southwestern blot with single stranded DNA. Nuclear protein extracts from IMR32 cells, 3T3 cells, Neuro2a cells or hippocampus tissue were electrophoresed on a 6% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane. Blots were incubated with \(^{32}\text{P}\)-labelled single stranded oligonucleotides corresponding either to a) the \(M_1\) PPY tract, b) coding strand of the conserved region and c) the template strand of the conserved region. Independent blots were used with each of the probes.
IMR32
3T3
Neuro2a
rat hippocampus

105 kDa 160 kDa
75 kDa

Conserved region (M1 +569)

IMR32
3T3
Neuro2a
rat hippocampus

105 kDa 160 kDa
75 kDa

Conserved region (M1 +486)

IMR32
3T3
Neuro2a
rat hippocampus

105 kDa 160 kDa
75 kDa

PPV tract (M1 +412s)
showed binding to the same molecular weight proteins in both of the cell lines, raising the possibility that they may be interacting with the same DNA binding proteins. This suggests again that the PPY/PPU tract and the conserved region are actually acting as a single regulatory element. As freshly produced blots were screened each time, signals cannot be the result of any carryover or incomplete removal of a previously used probe. The two identified bands obtained from IMR32 protein may be due to different proteins or modifications of the same protein.

Both the polypyrimidine tract and the conserved region bind nucleolin - As an approach to identify the transcription factor(s) that bind the PPY tract and/or both strands of the conserved region, a rat hippocampus λZap cDNA expression library (STRATAGENE) was screened with a $^{32}$P-labelled probe corresponding to the PPY tract. The probe used was the same as that used in gel shift assays. Southwestern analysis showed that rat hippocampus contained the 110 kDa protein but not the 100 kDa protein (Figure 11), although this could be due to the much lower intensity of the band in rat hippocampus. However, since M₁ mAChR is particularly abundant in rat hippocampus and the identification of one of the two proteins may aid the identification of the second protein, this library was used. One hundred thousand plaques were screened and three positive clones were isolated (clones 3, 61 and 62). Figure 12a shows the filter containing the amplified positive clone 3 probed with the PPY tract probe. A second lift from the same plate was taken and probed with an oligonucleotide corresponding to the coding strand of the conserved region (Figure 12b). The same plaques were able to hybridise with both probes, supporting the notion that, both the PPY tract and the conserved region bind the same factor. The positive clones 3, 61 and 62, contained an insert of about 2 kb, 2.2 (1.9+0.3) kb and 1.5 kb, respectively (Figure 13a). These clones were sequenced with vector primers, and resultant
Figure 12 – λ Zap cDNA library screening. Three potential clones isolated in a first screening of a λ Zap cDNA library (clones 3, 61 and 62) were replated and expressed proteins transferred to a nitrocellulose membrane. Proteins were denatured and renatured as described under “Materials and Methods” and blots incubated with $10^6$ dpm/ml of $^{32}$P-labelled probe. a) shows the blot obtained with clone 3 hybridised with a probe corresponding to the PPY tract (nucleotides +412 to +485) and b) shows a blot obtained from a second lift from the same plate hybridised with a probe corresponding to the sense strand of the conserved region (nucleotides +486 to +569).
a) PPY tract (M1 +412s)

b) Conserved region (M1 +486s)
sequences of both ends compared with the database. The search revealed that all three positive clones correspond to rat nucleolin, a multifunctional protein very abundant in the nucleolus, of apparent molecular weight of 100 and 110 kDa, depending on phosphorylation state (reviewed in (Tuteja and Tuteja, 1998) and (Ginisty et al., 1999)). The molecular weight of nucleolin corresponds with that of the protein found to interact with the PPY tract and both strands of the conserved region in the Southwestern assay (Figure 11). Two bands of 100 and 110 kDa were observed for IMR32, the M<sub>i</sub> expressing cell line, suggesting that nuclear extracts from these cells contain both phosphorylated and unphosphorylated forms of nucleolin. Only the band at 110 kDa was observed for 3T3 and Neuro2a, the two cell lines that do not express the M<sub>i</sub> gene, suggesting that nuclear extracts from these cell lines only contain detectable levels of the phosphorylated form of nucleolin. A graphic representation of the domains of mammalian nucleolin is shown in figure 13b. Clone 3 spanned from nucleotide 640 (relative to the translation start site of nucleolin) and included the stop codon (Figure 13b). This clone was found in reverse orientation, suggesting that expression derives from cryptic promoter activity, rather than from the lacZ promoter. Clone 61 and 62 were in forward orientation, in frame with the lacZ gene and spanned from nucleotide 196 and 800 respectively, and similarly to clone 3, also contained the stop codon (Figure 13b). The isolation of three independent clones, all coding for nucleolin, suggests that the PPY tract and both strands of the conserved region interact with nucleolin with high affinity and specificity. Nucleolin contains four RNA binding domains (RBDs) (residues 284 to 647 in the human gene) that are also able to bind single stranded DNA (Barrijal et al., 1992; Sapp et al., 1986). All the clones isolated during the λ zap screening contain the four RNA binding domains (Figure 13b). It is possible that different RBDs are involved in the
Figure 13 – Nucleolin interacts with the PPY tract and conserved region in the λ Zap screening. a) A single plaque from the tertiary screen of each positive clone (clones 3, 62 and 63) was cut out and the cDNA excised as pBluescript phagemid. Plasmid DNA was digested with Acc65I/NotI and electrophoresed through a 1% agarose gel to determine the size of the insert. Outer lines show 1kb plus DNA marker ladder (Life Technologies Inc). b) Positive clones were sequenced from both ends and found to correspond to different fragments spanning the coding region and 3’UTR of nucleolin. Domains within nucleolin are represented: dotted boxes represent stretch of basic or neutral amino acids, black boxes represent stretch of acidic amino acids, circles represent RNA binding domains, striped boxes represent a glycine/arginine rich domain and two vertical lines represent a bipartite nuclear localisation signal. Numbers on the top represent nucleotide position.
a) lambda zap clone

Acc65 I/Not I

3 61 62

pBluescript
2.0 kb
1.5 kb
0.3 kb

b) nucleolin

clone 3

clone 61

clone 62

- Basic / neutral
- Acidic
- RNA binding domain
- GAR domain
- NLS

1 2142

640
196
800
binding of the PPY tract and both strands of the conserved region, explaining why all three DNA sequences are bound by the same protein.

**Endogenous nucleolin binds to the polypyrimidine tract** - Nucleolin binds with high affinity and specificity to four stranded G4 DNA formed from the external transcribed spacer (ETS) region of the rDNA transcription unit (Hanakahi et al., 1999). The coding strand of the ETS region is very rich in guanine and readily self-associates *in vitro* to form structures stabilised by G-G pairing, where guanines interact via Hoogsteen bonds to form planar rings called G-quartets, and the G-quartets stack upon each other to form higher order structures. Since nucleolin binds to ETS, we reasoned that the PPY tract should be able to compete this binding. To test this hypothesis, a nuclear protein extract from IMR32 cells was incubated with an ETS-1 G4 probe (see materials and methods) and protein:DNA complexes resolved by electrophoresis. As previously reported by Hanakahi et al (Hanakahi et al., 1999), nucleolin binds to ETS-1 G4 forming multiple complexes of different mobility (Figure 14a, lane 1). All complexes can be competed by a specific, but not by a non-specific competitor (Figure 14a, lanes 2 and 6, respectively). When a single stranded DNA probe comprising the PPY tract was used as a competitor, two of the three complexes were competed, suggesting that endogenous nucleolin does bind to the PPY tract. Interestingly, for the third complex no competition was observed, but instead a subshifted band appeared. A rationalisation, not an explanation, of this result is presented next. First of all, it needs to be considered that the PPY tract has a mobility much higher than the ETS-1 G4 probe. This can be seen by the fact that the gel using the ETS-1 G4 probe has been running for 16 hours, compared to only 3-4 hours for the gel using the PPY tract as probe. To explain the observed subshifted band, we could consider that in this complex, nucleolin binds to more than one probe of ETS-1 G4, through
Figure 14 – Endogenous nucleolin binds to the PPY tract. a) An oligonucleotide corresponding to the external transcribed sequence (ETS) found upstream of the rDNA transcription unit forming a G4 DNA structure was isolated as described under “Materials and Methods” and labelled using labelled using T4 polynucleotide kinase. Nuclear protein extracts from IMR32 (lane 1-6) were incubated with either no competitor (lane 1), specific competitor (lanes 2), the PPY tract (nucleotides +412 to +485)(lane 3), the coding strand of the conserved region (nucleotides +486 to +569)(lane 4), the template strand of the conserved region (nucleotides +569 to +486)(lane 5) or non-specific competitor (lanes 6) for 20 min on ice. $^{32}$P-labelled probe was added and incubated for a further 20 min at room temperature. DNA and DNA:protein complexes were electrophoresed through a 4% non-denaturing polyacrylamide gel at 4°C for 16 hours, gel dried and exposed to an X-ray film. Lanes 7 and 8 show reactions in the absence of protein containing either specific competitor or no competitor, respectively. b) Gel mobility shift reactions with a PPY tract probe (nucleotides +412 to +485) were assembled as described above. Competitors used are indicated at the top of the figure. DNA and DNA:protein complexes were electrophoresed through a 4% non-denaturing polyacrylamide gel at room temperature for 2 hours, gel dried and exposed to an X-ray film.
a) Competitor Lane

<table>
<thead>
<tr>
<th>IMR32</th>
<th>Lane</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETS-1 G4</td>
<td>1</td>
</tr>
<tr>
<td>M1 +412s (PPY)</td>
<td>2</td>
</tr>
<tr>
<td>M1 +486s (Cs)</td>
<td>3</td>
</tr>
<tr>
<td>M1 +569a (Ca)</td>
<td>4</td>
</tr>
<tr>
<td>non specific</td>
<td>5</td>
</tr>
<tr>
<td>ETS-1 G4</td>
<td>6</td>
</tr>
<tr>
<td>free probe</td>
<td>7</td>
</tr>
<tr>
<td>IMR32</td>
<td>8</td>
</tr>
</tbody>
</table>

- ETS-1 G10
- ETS-1 G8
- ETS-1 G4

b) Competitor Lane

<table>
<thead>
<tr>
<th>IMR32</th>
<th>Lane</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 +412a (PPY)</td>
<td>1</td>
</tr>
<tr>
<td>M1 +486s (Cs)</td>
<td>2</td>
</tr>
<tr>
<td>M1 +569a (Ca)</td>
<td>3</td>
</tr>
<tr>
<td>ETS-1 G4</td>
<td>4</td>
</tr>
<tr>
<td>non specific</td>
<td>5</td>
</tr>
<tr>
<td>ETS-1 G4</td>
<td>6</td>
</tr>
<tr>
<td>free probe</td>
<td>7</td>
</tr>
</tbody>
</table>

- PPY tract
independent RBDs. The PPY tract may only be able to compete for the binding to one of the RBDs, consequently giving rise to a complex that contains nucleolin bound to both ETS-1 G4 DNA and PPY tract. The mobility of the complex consisting of nucleolin bound to both an ETS-1 G4 and to a PPY tract probe will be lower than the mobility of a complex where nucleolin is bound to two units of ETS-1 G4 probe, hence giving rise to a subshifted band.

Only the two bands with lower mobility were competed by either the sense or antisense strand of the conserved region (Figure 14a, lanes 4 and 5 respectively). The highest mobility complex is not competed by any of the strands of the conserved region. One possible explanation could be that in this complex, nucleolin interacts with the ETS-1 G4 probe and the conserved region through different RBDs, and therefore through different surfaces of the protein (Bouvet et al., 1997).

To confirm these results, a similar experiment was performed, but using the PPY tract as a probe. The PPY tract forms two different complexes when incubated with an IMR32 nuclear cell extract (Figure 14b, lane 1). These two complexes can be competed by specific competitor, but not by non-specific competitor (Figure 14b, lanes 2 and 6 respectively). Only the antisense strand of the conserved region was able to compete one of the complexes (Figure 14b, lanes 3 and 4), suggesting again that nucleolin uses different RBDs to interact with the PPY tract and both strands of the conserved region at the same time. ETS-1 G4 was able to compete the lowest mobility band (Figure 14b, lane 5), indicating again that both ETS-1 and the PPY tract bind to the same protein, nucleolin. A supershift is observed for the highest mobility band. In this case, and in a similar manner to the subshifted band observed in the previous case, the supershift could be due to the competition of the ETS-1 G4 probe for
binding to one of the RBDs, resulting in a complex containing nucleolin bound to both the PPY tract and the ETS-1 G4 probe, instead of two units of PPY probe. Since the ETS-1 G4 probe has a lower mobility than the PPY tract, this results in a supershifted band.

**DISCUSSION**

The data presented in this chapter showed that deletion of a fragment constituted by a 74 bp polypyrimidine/polypurine tract followed by a 84 bp sequence conserved between pig, mouse, rat and human (Wood *et al.*, 1999) found in the non-coding exon of the $M_i$ gene reduced expression of a reporter gene in the $M_i$ expressing cell line, IMR32 (Figure 6a). Furthermore, the PPY/PPU tract was able to form single stranded DNA within a supercoiled plasmid, and opening of the PPY/PPU tract forced the opening of the adjacent conserved region (Figures 7 and 9).

Single stranded elements have been identified in an increasing number of promoters (Avigan *et al.*, 1990; Bai *et al.*, 1998; Goodridge and Xu, 1996; Hoffman *et al.*, 1990; Potaman *et al.*, 1996; Santra *et al.*, 1994). The characterisation of the PPY/PPU tract of the chicken malic enzyme (Goodridge and Xu, 1996) and human Na,K-ATPase $\alpha_2$ genes (Potaman *et al.*, 1996) in a supercoiled plasmid has been reported. The use of supercoiled plasmid represents a valid model to study DNA secondary structure, since plasmid DNA is thought to mimic torsions and packaging present in nucleosomal DNA. In both the chicken malic enzyme and the human Na,K-ATPase $\alpha_2$ genes, the PPY/PPU tract forms an H-DNA structure composed of intramolecular triple-helix (where a DNA strand from one half of the sequence folds into the major groove of the other half-duplex, forming Hoogsteen base pairs) and single-stranded regions. In the case of the
chicken malic enzyme gene, the polypyrondine strand has been found to remain single stranded, whereas in the case of the human Na,K-ATPase α2 gene the polypurine strand is single stranded. The mapping of the single stranded nucleotides within the region containing the M1 PPY/PPU tract and conserved region suggests the possibility of a similar H-DNA structure. Alternatively, the polypyrondine tract may form a hairpin leaving the polypurine tract as single stranded DNA, and this forces the partial opening of the downstream conserved region, which in isolation would be double stranded (see Figure 7c).

The deletion of the M1 PPY/PPU tract causes a partial loss of transcription efficiency, similar to that seen for other genes, including the c-myc (Takimoto et al., 1989), myelin basic protein (MBP) (Lashgari et al., 1990), c-Ki-ras (Hoffman et al., 1990) and human decorin (Santra et al., 1994) genes. PPY/PPU tracts are not always correlated with positive regulation of a gene. For example, in the vascular smooth muscle (VSM) α-actin gene, the binding of two single stranded DNA binding proteins to opposite strands causes the formation of single stranded DNA, precluding the binding to double stranded DNA of the activator TEF-1 (Sun et al., 1995). In the human Na,K-ATPase α2 gene, the formation of H-DNA causes an unwinding of an adjacent TATA box. Consequently, the TATA box becomes unrecognisable to TBP, causing down regulation (Potaman et al., 1996). Transcriptional regulation through the PPY/PPU tract is not always dependent on the formation of non B-DNA structure. In the c-Ki-ras gene the transcriptional activator recognises the PPY/PPU tract in its double stranded form (Hoffman et al., 1990).

Single stranded DNA probes corresponding to the PPY tract and both strands of the conserved region were able to bind to one protein of 110 kDa in M1 non-expressing cells and two proteins of 100 and 110 kDa in M1 expressing cells (Figure 11). Screening of a λ zap cDNA library
identified nucleolin as the binding protein (Figures 12 and 13). Nucleolin is a multifunctional phosphoprotein that runs at 100 or 110 kDa depending on phosphorylation state (reviewed in (Tuteja and Tuteja, 1998) and (Ginisty et al., 1999)) suggesting that the two bands seen for IMR32 in the Southwestern assay correspond to translational modifications of the same protein, rather than to two different proteins. Competition assays with ETS-1 G4 (Hanakahi et al., 1999), a probe known to bind nucleolin (Figure 14) confirmed that nucleolin is the transcription factor binding to the PPY tract and both strands of the conserved region.

Both the PPY tract and conserved region of the M1 gene bind to the same protein, nucleolin, and the deletion of either or both domains reduces expression. Nucleolin contains four RNA binding domains (RBDs) (residues 284 to 647 in the human gene) that are also able to bind single stranded DNA (Barrijal et al., 1992; Sapp et al., 1986). Individually, none of the four individual RBDs interact significantly with RNA, but a peptide that contains the first two RBDs is sufficient to account for nucleolin RNA-binding specificity and affinity towards the minimal RNA-binding site target U/G CCCGA (Serin et al., 1997). The structural basis for RNA recognition by these RBDs was studied using a genetic system in Escherichia coli (Bouvet et al., 1997), showing that two RBDs domains participate in a joint interaction with an RNA hairpin called NRE and that each domain uses a different surface to contact the RNA. It is possible that different RBDs are involved in the binding of the PPY tract and both strands of the conserved region, and that binding to both elements is necessary to activate gene expression. A similar case occurs in the c-myc gene, where FUSE-binding protein (FBP) contains four different domains. FBP1+2 binds to the upstream fragment of the positive regulatory element FUSE and FBP3+4 binds to the downstream fragment, but the native
coupling between units is necessary to open the FUSE region and generate the active site seen in vivo (Michelotti et al., 1996).

Nucleolin was first described by Orrick et al. in 1973 (Orrick et al., 1973), in the course of a study of nucleolar proteins from rat liver. Nucleolin can represent as much as 10% of total nucleolar proteins (Bugler et al., 1982), being one of the most abundant proteins of the nucleolus. Nucleolin contains a bipartite nuclear localisation signal (NLS), which is necessary and sufficient for nuclear targeting (Creancier et al., 1993). However, nucleolin can be translocated between the nucleus and cytoplasm (Borer et al., 1989). During Xenopus oocyte maturation, cytoplasmic localisation of nucleolin coincides with massive phosphorylation by cdc2 and CK2 of sites clustered at the N-terminal domain of nucleolin, and during development, nuclear translocation is accompanied by net dephosphorylation (Rossner et al., 1997). On the other hand, nucleolin phosphorylation is accompanied by increased rRNA transcription and cell proliferation (Schneider et al., 1986; Suzuki et al., 1987). Therefore, it appears that nucleolin localisation and function are regulated by post-translational modification. Transfection assays showed that the PPY/PPU tract and conserved region are able to drive high levels of expression from the Inr containing promoter in IMR32 cells, but low levels of expression were also seen in all M1 non-expressing cells (Figure 6b). Southwestern assays showed that both phosphorylated and unphosphorylated forms are able to bind to the PPY tract and both strands of the conserved region (Figure 11). However, it would appear that the unphosphorylated form of nucleolin, found to bind at detectable levels only in M1 expressing cells (Figure 11), is more active in enhancing M1 expression. The low levels of expression observed for all M1 non-expressing cell lines when the PPY/PPU tract and the conserved region are driving expression from the Inr (Figure 6b) could be due to low levels
of unphosphorylated form of nucleolin present in these cells, not detected in the Southwestern assay. For transfection analysis, supercoiled plasmid is used. If the transcription factor that binds to the single stranded element is present in both M₁ expressing and non-expressing cell lines, it will bind to the element in both cases as long as this is in single stranded conformation. S1 nuclease experiments have shown that the region between +399 and +572 of the M₁ gene is able to form single stranded DNA in a supercoiled plasmid (Figure 7b). An in vivo study of the chromatin structure of the human c-myc gene has shown that the nucleosome ladder at the positive FUSE element is disrupted in c-myc expressing cells and furthermore, FUSE is underwound in vivo when c-myc is expressed (Michelotti et al., 1996). The processes by which the FUSE element is opened in c-myc expressing cell lines or by which is maintained as duplex DNA are still unknown. However, it has been shown that an adjacent sequence can prevent the formation of single stranded DNA by interacting with a factor that recognises double stranded DNA (Tomonaga and Levens, 1996). In IMR32, a cell line that does not express c-myc, FUSE-binding protein is present, and can bind FUSE in vitro, but not in vivo since in this cell line FUSE is found as duplex DNA (Michelotti et al., 1996). A similar effect may be seen in the M₁ gene. For that reason we may see an activation of the Inr promoter in all cell lines (see Figure 6b). It will thus be important to look at the chromatin structure of M₁ in vivo.

Nucleolin has been involved in many activities. The ability of nucleolin to interact with both rRNA through the RBDs (Ghisolfi-Nieto et al., 1996) and ribosomal proteins through the C-terminal domain (Ginisty et al., 1998) suggests that this protein may be involved in pre-ribosome assembly. Nucleolin is also involved in rDNA transcription (Bouche et al., 1984; Egyhazi et al., 1988) and pre-rRNA processing (Ginisty et al., 1998). Interaction of nucleolin with histone H1 through the acidic domains
found at the N-terminus of the nucleolin induces chromatin decondensation by a mechanism in which nucleolin displaces the chromatin binding domain of H1 (Erard et al., 1990; Erard et al., 1988). Shuttling of nucleolin between the nucleus and cytoplasm (Borer et al., 1989) has raised the possibility that nucleolin is involved in the nucleocytoplasmic transport of ribosomal particles. Cloning of human DNA helicase IV showed that this protein is in fact nucleolin (Tuteja et al., 1995). Importantly, nucleolin has been shown to function as a transcription factor for several genes. Nucleolin and hnRNP D form the B-cell specific transcription factor LR1, which activates transcription of the c-myc and the Epstein-Barr virus EBNA-1 genes (Dempsey et al., 1998; Hanakahi et al., 1997). On the other hand, nucleolin acts as a transcriptional repressor of the alpha-1 acid glycoprotein (AGP) gene (Yang et al., 1994). It has recently been shown that nucleolin interacts with the oncoproteins A-MYB and c-MYB, but not B-MYB DNA binding domains, causing a down-regulation of MYB transcriptional activity (Ying et al., 2000). Recent findings indicate that substantial cross talk may exist between transcriptional and post-transcriptional processes (reviewed in (Ladomery, 1997)). Firstly, there are suggestions that specific promoters influence the post-transcriptional fate of transcripts. In a study to investigate the HIV-1 promoter, it was found that reporter transcripts driven by the HIV-1 promoter were stored in the cytoplasm, but not translated. This effect was dependent on the choice of promoter, since transcripts were exported to the cytoplasm and translated when the CMV rather than the HIV-1 promoter was used. Secondly, an increasing number of proteins appear to be multifunctional, participating in both transcriptional and post-transcriptional events. For example, TFIIIA is involved in 5S rRNA gene expression, but also binds to the gene product 5S rRNA. Also, hnRNP K, a factor identified initially by association with hnRNP complexes (Matunis et al., 1992), binds to the polypyrimidine
strand of a CT element in the c-myc gene and activates transcription (Takimoto et al., 1993; Tomonaga and Levens, 1995). WT1, a protein that can exist in four different isoforms through alternative splicing, plays a role in transcriptional processing of RNA as well as in transcription, its final function dependent upon the splicing variant (Larsson et al., 1995; Rauscher et al., 1990). Hence, nucleolin appears to be one more component of this group of multifunctional proteins.
CHAPTER 5

RESULTS (PART 3)
INTRODUCTION

Frequently, the promoter and regulatory sequences of a gene are represented as a series of binding sites. However, this linear view lacks the complexity and dynamic character of the functioning promoter in vivo. This map fails to show the interactions among transcription factors. In some cases, the looping of DNA, forced by the binding of factors, allows contact of transcription factors bound at distant sites. The affinity and specificity of DNA binding is sometimes augmented, and also regulated, by interactions with accessory proteins bound to adjacent DNA sites, thereby forming what has been termed an “enhanceosome” (Carey, 1998; Maniatis et al., 1998). This linear map also fails to show the dynamic character of chromatin. DNA is wrapped around nucleosomes (Arents et al., 1991; Kornberg, 1974), and modifications in histones causes a condensation/decondensation of chromatin, which has a direct relation with repression/activation of transcription (reviewed in (Grunstein, 1997; Kornberg and Lorch, 1999; Struhl, 1998)).

Most genes are regulated by multiple factors, and the balance between positively and negatively acting transcription factors which bind to the regulatory regions of a particular gene will determine the rate of gene transcription in any particular situation. Furthermore, cell specificity of gene transcription may result from the formation of a transcription complex containing a specific interacting set of nuclear protein factors. For example, multiple sequences present along 400 bp of the proximal promoter of the insulin gene contribute to its overall expression in β-cells (reviewed in (Ohneda et al., 2000)). These sequence
elements include, among others, binding sites for the transcription factors PDX1, E2A, BETA2, C/EBPβ, NF-Y, CREB, HMG1 and LMX (Ohneda et al., 2000). Another example is represented by the ubiquitous Cox genes, the transcription activity of which is regulated by factors binding to the 5' upstream Sp1, NRF1, GABP (NRF2), and YY1 sites (Lenka et al., 1998).

Neither deletion of the PPY/PPU tract or the conserved region alone or both together from the M1 promoter resulted in complete abolition of luciferase activity in IMR32 cells (Figure 6a). This suggests that, as in the case of many other genes, the M1 gene is regulated by multiple factors.

RESULTS

Multiple cis-elements are present in the region between +166 and +417 - In order to identify any possible cis-elements present in the M1 gene other than the PPY/PPU tract and the conserved region, more deletions were produced (Figure 15a). As shown earlier, deletion of the PPY/PPU tract and conserved region together resulted in a 58% decrease in reporter gene activity (compare constructs pGL3 -372/+417 M1 and pGL3 -372/+602 M1). Further deletion of the region between +366 and +417 to give the construct pGL3 -372/+366 M1 increased from 42% (2 fold over basic) to 75% (3.7 fold over basic), suggesting the presence of a repressor between +316 and +366. Further deletion between +316 and +366 to give the construct pGL3 -372/+316 M1 reduced expression to 47% (2.3 fold over basic). This reduction could be due to the presence of an additional enhancer element in this region. The deletion of the fragment between +241 to +316 reduced expression from 2.3 to 0.9 fold over basic. Expression was further reduced to 0.2 fold over basic when the fragment between +166 and +241 was deleted. It appears that between +166 and +316 there are
Figure 15 – More deletional analysis within the non-coding exon of the \( M_1 \) gene. Cells were transfected with 750 ng of the reporter plasmids represented on the left. Bars and numbers express results of at least 3 independent experiments each performed in triplicate, normalised to Renilla luciferase activity and expressed as -fold over expression of the promoterless pGL3 basic vector (a) or pGL3 Inr (b). a) Deletions within the non-coding exon of the \( M_1 \) gene were produced, and the ability of the \( M_1 \) promoter to drive expression of the luciferase gene measured. b) Fragments identified to be important for expression driven by the \( M_1 \) promoter were cloned in front of the Adenovirus Major Late promoter (Inr) and the ability to control expression of the luciferase reporter gene by this promoter analysed.
a) pGL3 -372/+602 M1
pGL3 -372/+417 M1
pGL3 -372/+366 M1
pGL3 -372/+316 M1
pGL3 -372/+241 M1
pGL3 -372/+166 M1
pGL3 -372/+602 Δ166+227 M1
pGL3 -372/+602 Δ166+227 ΔPPY M1

Normalised luciferase activity relative to pGL3 Basic

b) pGL3 +166/+603 M1 Inr
pGL3 +166/+316 M1 Inr
pGL3 +396/+569 M1 Inr

Normalised luciferase activity relative to pGL3 Inr
one or more elements important for expression of the M$_i$ gene in IMR32. To test the importance of the region between +166 and +316, this fragment was deleted, to give the construct pGL3 -372/+602 Δ+166/+327 M$_i$. Deletion of this fragment caused a decrease in expression of 47%, suggesting that there are one or more elements within this region acting as enhancer(s) of M$_i$ gene expression. Concurrent deletion of both the region between +166 and +327 and the PPY/PPU tract (see pGL3 -372/+602 M$_i$ Δ+166/+327 ΔPPY) reduced expression to the same level as that observed for the deletion of the PPY/PPU tract alone, suggesting that these two regions do not cooperate. 3T3 cells seem to exhibit a similar profile of reporter gene expression, albeit less marked than that seen with IMR32 cells. Indeed, luciferase activity driven by the M$_i$ promoter relative to pGL3 Basic in 3T3 does not reach the levels seen in IMR32 for any of the constructs except pGL3 -372/+366 M$_i$. None of the deletions seems to have any significant effect in the two neuronal M$_i$ non-expressing cell lines, Neuro2a and NB4 1A3, as no construct, except again pGL3 -372/+366 M$_i$, drives expression of the reporter gene.

As in chapter 4, all deletions were produced within the non-coding exon. Again, to check that changes in expression were due to deletion of cis-elements and not to changes in mRNA stability, different fragments were cloned in front of the Inr (Pengue and Lania, 1996) and its ability to drive expression examined by transient transfection (Figure 15b). The +166/+602 fragment of the M$_i$ gene enhanced transcription from the Inr promoter 9 fold in IMR32 cells, 2 fold in 3T3 cells and drove no expression in Neuro2a and NB4 1A3. Therefore, this fragment appeared to be sufficient to recapitulate the expression pattern seen for the construct pGL3 -372/+602 M$_i$, where expression derives from the M$_i$ promoter rather than the Inr promoter. The fragment between +166 and +316 enhanced the Inr promoter 3 fold in IMR32 cells, suggesting that this region
contains an enhancer element. Interestingly, a fragment containing the PPY/PPU tract and the conserved region (pGL3 +396/+569 M₁ Inr) enhanced expression to a much higher degree than the fragment between +166 and +602, which also contains the PPY/PPU tract and conserved region. This suggests that in the +166/+602 fragment, in addition to the PPY/PPU tract and conserved region, which act as enhancers of expression, there is more than one element with different effects on expression and that the final levels of expression are due to the balance of both positive and negative regulatory elements. In summary, there are multiple regulatory elements in the region between +166 and +602 of the M₁ gene, and the balance between the different positive and negative regulatory elements is responsible for the net final level of expression. The same pattern but less marked is seen for the three M₁ non-expressing cell lines.

**SHARP-1 binds to regulatory regions of the M₁ gene** - As observed in transient transfection assays, transcription of the M₁ muscarinic acetylcholine receptor gene is regulated by several domains within the first exon. In order to identify the transcription factors important for the regulation of the M₁ gene that bind to sequences other than the PPY/PPU tract and conserved region, the region between +166 and +417 (relative to the transcription start site) was used as bait to screen an adult rat brain cDNA library in a yeast one-hybrid screen assay (Chong et al., 1995; Wang and Reed, 1993). Two independent positive clones were isolated (Figure 16a, colonies 3 and 10). Digestion of the positive clones showed that they both contain an insert of 1.5 kb (Figure 16b). Sequencing of the inserts in these clones showed that they were identical and contained the entire open reading frame of the previously identified transcription factor SHARP-1 (Rossner et al., 1997). Growth of yeast on plates containing 10 mM 3AT was dependent upon the binding of SHARP-1.
Figure 16 - SHARP-1 binds to the $M_i$ promoter in a yeast one-hybrid screen. a) An adult rat brain cDNA library was screened using the region of the $M_i$ gene between +166 and +417 as bait and the yeast strain SFY526. Candidate clones were re-transformed into SFY526 – pBM2389 +417/+166 $M_i$ and plated onto -T/-H/-L +10 mM 3AT. Two positives clones were isolated (number 3 and 10), which failed to grow with either empty reporter vector, pBM2389, or with empty expression vector, pGAD10, showing that growth was dependent upon the interaction of the expressed library protein with the $M_i$ bait. b) Digestion of the isolated library clones with $Bgl$II, showed that both clones contain a cDNA of 1.5 kDa. Sequencing of the two clones showed that they both encompass the entire SHARP-1 open reading frame and also 130 bp of the 5’ UTR.
a) pGADIO ARB 3

b) pGAD10
Adult Rat Brain cDNA

SD-TRP-HIS-LEU
+10 mM 3AT

pBM2389

pBM2389 +417/+166 M₁

pGAD10 ARB 3

pGAD10 ARB 10

pGAD10

310 empty

SHARP-1 insert

1.6 kDa
to the $M_1$ sequence, since growth was not seen in the absence of either bait or SHARP-1 (Figure 16a).

**SHARP-1 has high homology to SHARP-2, STRA13 and DEC1** – SHARP-1 was originally isolated during the course of a search for mammalian bHLH proteins expressed in differentiated neurones (Rossner et al., 1997). Although SHARP-1 was isolated by homology to *hairy* and *enhancer of split*, sequence alignment with these proteins shows that they are quite distantly related, sharing only 37 to 42% homology within the bHLH domain (Rossner et al., 1997). To date, no function to SHARP-1 has been ascribed.

A database search for proteins with homology to SHARP-1 identified three proteins; SHARP-2, STRA13 and DEC1 (Figure 17). SHARP-2 is a bHLH protein isolated in the same screen as used to identify SHARP-1 (Rossner et al., 1997), and its function is also unknown. STRA13 was isolated as a retinoic acid inducible gene in mouse P19 embryonic carcinoma cells and has been shown to be able to repress the tk promoter when fused to Gal4 DNA binding domain (DBD) (Boudjelal et al., 1997). Finally, DEC1 is a protein that was cloned by subtractive hybridisation to identify mRNAs expressed in cAMP differentiated human embryo chondrocytes (Shen et al., 1997). Again, no function for DEC1 has been reported. Inspection of amino acid sequence shows that SHARP-2, STRA13 and DEC1 contain 411 or 412 amino acids, of which 366 are conserved, showing a sequence identity between them of 89% suggesting that they are, in fact, rat, mouse and human homologues. SHARP-1 is more divergent and contains only 253 amino acids. Highest sequence identity is seen in the bHLH domain and in helices 3 and 4 (also called the Orange domain (Dawson et al., 1995; Giebel and Campos-Ortega, 1997)), whereas within the C-terminal domain only two stretches of 8 and 15 amino acids are conserved.
Figure 17 – SHARP-1 is closely related to SHARP-2/STRA13/DEC1. Sequence alignment of SHARP-1 with other bHLH proteins identified in a database search for proteins related to SHARP-1. SHARP-2 (Rossner et al., 1997), STRA13 (Boudjelal et al., 1997) and DEC1 (Shen et al., 1997) share 89% sequence identity between them and appear to be homologues derived from rat, mouse and human respectively. SHARP-1 is highly conserved within the bHLH domain with SHARP-2/STRA13/DEC1 (95% sequence identity) and within helix III and IV, also called orange domain (60% sequence identity). Outside of these domains, SHARP-1 diverges from the other three proteins, apart from two stretches of 8 and 15 amino acids within the C-terminal domain.
Expression of SHARP-1 in different cell lines – It has been shown previously that expression of SHARP-1 is largely restricted to differentiated neurones in the post-natal CNS, predominantly in the cerebellum and hippocampus, although lower levels are also detectable in heart, muscle and lung (Rossner et al., 1997). To test if SHARP-1 and M₁ are differentially expressed, we examined expression of SHARP-1 in the different cell lines and cerebellum using RT-PCR. As seen in Figure 18a, SHARP-1 is highly expressed in IMR32 cells, the cell line that also expresses M₁, Neuro2a cells and NB4 1A3, the two M₁ non-expressing neuroblastoma cell lines. Low levels of expression were detected in the fibroblast cell line 3T3. PCR was carried out using HPRT primers as a cDNA loading control (Figure 18b).

SHARP-1 controls levels of M₁ expression – In order to test the functional effect of SHARP-1 on M₁ expression, IMR32, 3T3 and Neuro2a cells were transfected with a reporter vector containing the region of the M₁ gene between -372 and +602 (relative to the transcription start site). Over-expression of SHARP-1 (pSHARP-1 myc) had no effect on expression of the promoterless reporter vector pGL3 basic, but reduced expression driven by the M₁ promoter by 55% (Figure 19a). A similar effect was seen in 3T3 cells (Figure 19b). No effect was seen in Neuro2a, the neuronal cell line that does not express M₁. Similarly, over-expression of SHARP-1 reduced expression of the reporter plasmid pGL3 +166/+316 M₁ Inr, which contains the region of the M₁ gene shown to be bound by SHARP-1 in a yeast 1-hybrid screening, by 35% in IMR32 cells (Figure 19a). However, no significant effect was seen in 3T3 cells (Figure 19b). These results show that SHARP-1 is able to downregulate expression of the M₁ gene in a M₁ expressing cell line.
Figure 18 – Expression of SHARP-1 in different cell lines. a) PCR was performed using a rat SHARP-1 sense primer within the basic domain (SHARP-1.125s) and an antisense primer within helix 3 (SHARP-1.441a) on cDNA derived from IMR32 (lane 1), 3T3 (lane 2), Neuro2a (lane 3), NB4 1A3 (lane 4), rat cerebellum (lane 5) and mouse cerebellum (lane 6). Lane 7 shows the water negative control and the two outer lanes correspond to 1 kb plus DNA marker ladder (Life Technologies Inc). PCR products were electrophoresed through a 2% MetaPhor agarose gel. b) PCR was performed on the same cDNA samples as above, using primers for hypoxanthine-guanine phosphoribosyl transferase (*HPRT*) gene hprt 231s and hprt 567a to verify that equivalent amount of cDNA had been assayed.
to the $M_1$ sequence, since growth was not seen in the absence of either bait or SHARP-1 (Figure 16a).

**SHARP-1 has high homology to SHARP-2, STRA13 and DEC1**—
SHARP-1 was originally isolated during the course of a search for mammalian bHLH proteins expressed in differentiated neurones (Rossner et al., 1997). Although SHARP-1 was isolated by homology to *hairy* and *enhancer of split*, sequence alignment with these proteins shows that they are quite distantly related, sharing only 37 to 42% homology within the bHLH domain (Rossner et al., 1997). To date, no function to SHARP-1 has been ascribed.

A database search for proteins with homology to SHARP-1 identified three proteins; SHARP-2, STRA13 and DEC1 (Figure 17). SHARP-2 is a bHLH protein isolated in the same screen as used to identify SHARP-1 (Rossner et al., 1997), and its function is also unknown. STRA13 was isolated as a retinoic acid inducible gene in mouse P19 embryonic carcinoma cells and has been shown to be able to repress the tk promoter when fused to Gal4 DNA binding domain (DBD) (Boudjelal et al., 1997). Finally, DEC1 is a protein that was cloned by subtractive hybridisation to identify mRNAs expressed in cAMP differentiated human embryo chondrocytes (Shen et al., 1997). Again, no function for DEC1 has been reported. Inspection of amino acid sequence shows that SHARP-2, STRA13 and DEC1 contain 411 or 412 amino acids, of which 366 are conserved, showing a sequence identity between them of 89% suggesting that they are, in fact, rat, mouse and human homologues. SHARP-1 is more divergent and contains only 253 amino acids. Highest sequence identity is seen in the bHLH domain and in helices 3 and 4 (also called the Orange domain (Dawson et al., 1995; Giebel and Campos-Ortega, 1997)), whereas within the C-terminal domain only two stretches of 8 and 15 amino acids are conserved.
Figure 19 - SHARP-1 represses M₁ expression. IMR32 cells (a) and 3T3 cells (b) were co-transfected with the reporter construct represented on the left, where Inr represents the Adenovirus Major Late promoter Inr and numbers represent the nucleotide position relative to the transcription start site of the M₁ gene, and either pMT (open bars) or a vector expressing SHARP-1 (shaded bars). Reporter and expression vectors are in a ratio of 1:5. Values are the averages of three independent transfections, each performed in triplicate, normalised to co-transfected CMV-Renilla, and expressed as -fold over luciferase activity obtained with the promoterless reporter vector, pGL3 basic co-transfected with empty expression vector, pMT.
a) IMR32 cells

pGL3 basic

Luciferase

pGL3 -372/+602 M1

Luciferase

pGL3 Inr

Inr Luciferase

pGL3 +166/+316 M1 Inr

Inr Luciferase

Fold over pGL3 basic / pMT

b) 3T3 cells

pGL3 basic

Luciferase

pGL3 -372/+602 M1

Luciferase

pGL3 Inr

Inr Luciferase

pGL3 +166/+316 M1 Inr

Inr Luciferase

Fold over pGL3 basic / pMT
**SHARP-1 homodimerises** – All bHLH proteins dimerise in order to bind DNA (Murre et al., 1989b). Since SHARP-1 was identified in a yeast one-hybrid screen, it seemed likely that SHARP-1 could either homodimerise or heterodimerise with a yeast partner. In order to distinguish between these possibilities we carried out an immunoprecipitation assay using differentially tagged recombinant SHARP-1. Neuro2a cells were transfected with either a combination of myc-tagged SHARP-1 (pMT SHARP-1) and myc-tagged Gal4 DBD (pMT G4) (Figure 20, lanes 1 and 3) or a combination of myc-tagged SHARP-1 (pMT SHARP-1) and a myc tagged fusion of Gal4 DBD and SHARP-1 (pMT G4 SHARP-1) (Figure 20, lanes 2 and 4). Cell extracts were immunoprecipitated with Gal4 DBD antiserum, subjected to PAGE and the results visualised by Western blot analysis using c-myc antiserum. An antibody to Gal4 was able to co-immunoprecipitate SHARP-1 only in the presence of a Gal4–SHARP-1 fusion protein (Figure 20, compare lane 4 with 3), demonstrating that SHARP-1 is able to homodimerise.

**SHARP-1 acts as transcriptional repressor** – Over-expression of SHARP-1 showed that this bHLH protein is able to downregulate expression driven by the M1 promoter. However, nothing is known about SHARP-1 function. bHLH proteins can act as transcriptional activators or repressors (reviewed in (Kageyama et al., 1997)). It is becoming evident that many transcription factors are able to modulate transcription by more than one mechanism. Such is the case of YY1, which represses the α-actin promoter by occluding an overlapping site for the serum response factor (SRF) (Lee et al., 1992), and actively represses the c-fos promoter through a protein/protein interaction with CREB, which is bound upstream of the YY1 site (Zhou et al., 1995). Furthermore, many transcription factors can act both as transcriptional repressors or activators depending on the nature of the target promoters and/or cell
Figure 20 - SHARP-1 homodimerises. Neuro2a cells were transfected with either myc-tagged Gal4 DBD (pMT G4) and myc tagged SHARP-1 (pMT SHARP-1) (lane 1) or myc tagged Gal4 SHARP-1 (pMT G4 SHARP-1) and myc-tagged SHARP-1 (pMT SHARP-1) (lane 2). Cells were harvested after 2 days and cell extracts were immunoprecipitated with Gal4 DBD antibody and analysed by Western blotting with c-myc Ab. SHARP-1 was not immunoprecipitated in the presence of the Gal4 DBD (lane 3), but G4-SHARP-1 was immunoprecipitated (lane 4), showing that SHARP-1 is able to homodimerise.
<table>
<thead>
<tr>
<th></th>
<th>Total input</th>
<th>G4 IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMT G4 SHARP-1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pMT SHARP-1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>pMT G4</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Western c-myc Ab

[Image of gel with markers at 75 KDa, 50 KDa, and 35 KDa]

- G4 SHARP-1
- SHARP-1
- G4
conditions. Again, this is the case for YY1 (reviewed in (Shi et al., 1997; Thomas and Seto, 1999)), which as a default mode acts as a repressor, but under certain situations, such as in the presence of E1A, an activation domain is unmasked and its function is switched from repressor to activator (Shi et al., 1991). In order to confirm a function of SHARP-1 as transcriptional repressor, or to find out if SHARP-1 is able to function in more than one manner as observed for many other transcription factors, the ability of SHARP-1 to control expression driven by different promoters was analysed. IMR32, 3T3 and Neuro2a cells were transfected with plasmids expressing SHARP-1 fused to Gal4 DBD with each of the reporter genes showed in Figure 21. pTRE UAS TATA contains 7 TRE (Tet-responsive element) and 5 Gal4 binding sites 21 bp upstream of the Elb TATA box. pGL3 UAS TRE TATA contains 5 Gal4 binding sites (placed 350 bp from the TATA box) and 7 TRE upstream of the TATA box. In pGL3 TRE UAS Inr, the TATA box from pGL3 TRE UAS TATA has been replaced by the AdMLP Initiator. Expression values of all reporter constructs are normalised to expression in the presence of Gal4 DBD alone. SHARP-1 was able to repress transcription of a TATA containing promoter when bound proximal to the transcription start site in all cell lines (Figure 21a, left). SHARP-1 was also able to repress activated transcription by Tet-VP16 (activation domain of the herpes simplex virus transcriptional activator VP-16 fused to the binding domain of the Tetracycline responsive factor) in all cell lines (Figure 21a, right). In addition, SHARP-1 was able to repress both basal and activated transcription from a TATA-containing promoter when bound distal to the transcription start site, although repression was not so marked in this case (Figure 21b). The ability of SHARP-1 to regulate transcription from a TATAless promoter was also tested. As can be seen in Figure 21c, SHARP-1 can repress basal and activated transcription from a TATAless promoter. These data indicate that
Figure 21 – SHARP-1 acts as transcriptional repressor. IMR32, 3T3 and Neuro2a cells were transfected with vectors expressing Gal4 DBD alone or fused to SHARP-1 together with or without a vector expressing Tet-VP16 and the reporter plasmid shown in each figure. Equal amounts of Gal4 DBD fusion, Tet-VP16/stuffer and reporter plasmids were used. Values are the averages of three independent transfections, each performed in triplicate, normalised to co-transfected CMV-Renilla, and expressed as a percentage of the values obtained with Gal4 alone for each cell line in the absence or presence of Tet-VP16 (open bar). a) shows repression by proximal recruitment of SHARP-1 to a TATA-containing promoter while b) shows the effect when distally recruited. c) shows repression of a TATA-less promoter. In each case the left-hand panel shows repression of basal promoter activity whilst the right panel shows repression of VP16 activated transcription.
SHARP-1 acts as a repressor of both basal and activated transcription of both TATA-containing and TATAless promoters. In the case of a TATA-containing promoter, repression is evident when SHARP-1 is bound either proximally or distally to the promoter, although the degree of repression is more marked when SHARP-1 is tethered proximally.

**SHARP-1 represses transcription through two independent domains** – In order to map the domain(s) responsible for the repression function of SHARP-1, deletion mutants of the Gal4–SHARP-1 fusion protein were generated. The ability of these fusion proteins to repress transcription was analysed using the reporter gene driven by a TATA containing promoter with 5 Gal4 binding sites proximal to the transcription start site (pTRE UAS TATA) in Neuro2a cells (Figure 22). Results were normalised to expression of the reporter gene in the presence of Gal4 DBD alone. Deletion of the C-terminal domain of SHARP-1 (to give pMT G4 NbHO-SHARP-1) slightly relieved repression by SHARP-1, but the C-terminal domain (residues 174 to 253) of SHARP-1 fused to Gal4 DBD (pMT G4 C-SHARP-1) was able to repress transcription as robustly as full length SHARP-1. Therefore, it would appear that SHARP-1 must contain at least two independent repression domains, one in the C-terminal domain, and another in the remaining fragment. To map the second repression domain of SHARP-1, more deletion mutants were examined. Deletion of the orange domain and C-terminal domain to give pMT G4 NbH-SHARP-1 still gave robust repression, but further deletion of the bHLH domain to leave only the N-terminal domain (pMT G4 N-SHARP-1) led to relief of most of the repression activity, suggesting that the bHLH domain also mediates repression. This was confirmed by analysis of two further constructs. Fusion of the orange domain and flanking sequence (residue 103 to 173) and the Gal4 DBD (pMT G4 O-SHARP-1) showed some degree of repression but fusion of the bHLH domain (residue 43 to 102)
Figure 22 – SHARP-1 represses transcription through two independent domains. Neuro2a cells were transfected with vectors expressing Gal4 DBD alone or fused to different deletion mutants of SHARP-1 along with a gene driven by a TATA containing promoter with 5 Gal4 binding sites proximal to the transcription start site (pTRE UAS TATA). Equal amounts of Gal4 DBD fusion and reporter plasmids were used. Values are the averages of three independent transfections, each performed in triplicate, normalised to co-transfected CMV-Renilla, and expressed as a percentage of the values obtained with Gal4 alone for each cell line.
and Gal4 DBD (pMT G4 bH-SHARP-1) indicated that the bulk of repression was mediated by the bHLH domain. In summary, SHARP-1 contains two independent repression domains, one in the bHLH domain and the other in the C-terminus.

**SHARP-1 represses transcription through two different mechanisms** – Recent studies have shown that many transcriptional repressors exert their action through recruitment of HDAC activity (see (Ayer, 1999) for review). To test if SHARP-1 represses transcription through such a mechanism, Neuro2a cells were treated with the histone deacetylase inhibitor Trichostatin A (TSA) (Nakano et al., 1995) and the effect on SHARP-1 mediated repression examined (Figure 23). For each concentration of TSA used, expression values of the reporter gene were normalised to those obtained in the presence of Gal4 DBD alone, and results were expressed as fold over untreated cells. Repression by full length SHARP-1 is partially relieved by TSA, since expression of the reporter gene was de-repressed 4 fold in the presence of 100 nM TSA. Deletion of the C-terminal domain of SHARP-1 (to give pMT G4 NbHLHO-SHARP-1) showed that repression mediated by the bHLH domains was much less sensitive to TSA, resulting in a 1.6 fold de-repression by 100 nM TSA. However, repression by the C-terminal domain alone was relieved by 6.5 fold with 100 nM TSA and 2-fold de-repression could be seen in the presence of 10 nM TSA. These results show that the C-terminal domain of SHARP-1 represses transcription via a mechanism that is likely to involve histone deacetylase activity, but that the bHLH domain represses transcription in a histone deacetylase independent manner.

**SHARP-1 binds to class B sites** – SHARP-1 contains an Arg in position 13 (R13) of the basic region present in all class B bHLH proteins. This residue is essential for class B proteins to bind the consensus sequence
Figure 23 – Repression by SHARP-1 is through two independent mechanisms. Neuro2a cells were incubated with 1, 50 or 100 nM TSA for 24 hours before transfection. Cells were transfected with vectors expressing Gal4 DBD alone or fused to different deletion mutants of SHARP-1 along with a reporter vector containing 5 Gal4 DBD upstream of a TATA containing promoter. Equal amounts of Gal4 DBD fusion and reporter plasmids were used. Transfected cells were fed with media containing the indicated concentration of TSA for 24 hours before harvesting. Values are the averages of three independent transfections, each performed in triplicate and normalised to co-transfected CMV-Renilla. Results are normalised to expression with Gal4 alone and expressed as fold over cells untreated with TSA.
CA(C/T)GTG (Dang et al., 1992). To identify SHARP-1 binding sites, a binding site selection by PCR assay was performed (Pollock and Treisman, 1990). Nuclear protein extracts from IMR32 and Neuro2a expressing either SHARP-1 myc-tagged at the N-terminus (pMT SHARP-1) or SHARP-1 myc-tagged at the C-terminus (pSHARP-1 myc) were used to select for binding of specific DNA sequences from a library of oligonucleotides containing a central random stretch of 12 bp. Complexes containing SHARP-1 bound to specific DNA recognition sequences were separated from unbound oligonucleotide by 5 sequential rounds of immunoprecipitation using the c-myc antibody. Bound DNA was subsequently amplified using PCR and cloned into pGemT-Easy (Promega). DNA from the last round of immunoprecipitation was used in a gel shift assay. Protein-oligonucleotide complexes were excised and eluted from the polyacrylamide gel, the oligonucleotide sequence amplified by PCR and cloned into pGemT-Easy. Sequences of 24 oligonucleotides obtained from either the 5th round of immuno-selection or the gel shift when using SHARP-1 transfected IMR32 cells were determined. Figure 24a shows an alignment of the sequences obtained. As predicted by the presence of an arginine at position 13 of the basic region, SHARP-1 appears to bind preferably to class B sites, since 22 of the 24 sequences correspond to a class B site. However, 2 of the sequences belong to a class C site, showing that SHARP-1 is also able to bind this class. Figure 24b shows the sequence of the oligonucleotides obtained from the 5th round of selection with SHARP-1 transfected Neuro2a cells. In Neuro2a cells, SHARP-1 also binds preferentially to class B sites, although 2 out of 11 sequences corresponded to class C sites. The results shown correspond to oligonucleotides selected using either cells expressing SHARP-1 myc tagged either at the N- or C-terminus, demonstrating that the myc tag does not interfere with SHARP-1 binding. In summary, the site selection assay
Figure 24 – SHARP-1 preferentially binds to a B box. Nuclear protein extracts from IMR32 or Neuro2a cells transfected with either pMT SHARP-1 or pSHARP-1 myc were used for a binding site selection assay (see “Materials and Methods”). Oligonucleotides containing a central random stretch of 12 base pairs were submitted to five rounds of immunoprecipitation and the DNA obtained in the last round was used for gel shift assay. DNA:protein complexes were cut out from the gel and the DNA eluted, amplified by PCR, cloned into pGemT-easy (Promega) and sequenced. Lower case indicates oligonucleotide linker sequence, and upper case selected random sequence. Bold represents consensus sequence.
a) IMR32

5th round 1  tcACGTGACACAG  
2  tcACGTGAGCACTT  
3  tcACGTGACCATAC  
4  tcACGTGAGCCACC  
5  tcACGTGACCATAC  
6  tcACGTGAGCCACC  
7  tcACGTGAGCCACC  
8  tcACGTGACCCAC  
9  tcACGTGACCCAC  
10  tcACGTGACCCAC  
11  tcACGTGACCCAC  

5th round +  
Gel shift 11  tcACGTGACCAACGA  
12  tcACGTGACCAACGA  
13  tcACGTGACCAACGA  
14  tcACGTGACCAACGA  
15  tcACGTGACCAACGA  
16  tcACGTGACCAACGA  
17  tcACGTGACCAACGA  
18  tcACGTGACCAACGA  
19  tcACGTGACCAACGA  
20  tcACGTGACCAACGA  
21  tcACGTGACCAACGA  
22  tcACGTGACCAACGA  
23  tcACGTGACCAACGA  
24  tcACGTGACCAACGA  

b) Neuro2a

5th round 1  ...GAGACCCACGTC...  
2  ...TACGCCACGTC...  
3  ...GTCCGACGTC...  
4  ...GTCCGACGTC...  
5  ...GTAGCAGCAGTT...  
6  ...GTAGCAGCAGTT...  
7  ...tcACGTGACGAGG.  
8  ...tcACGTGACGAGG.  
9  ...tcACGTGACGAGG.  
10  ...tcACGTGACGAGG.  
11  ...tcACGTGACGAGG.  

......tcACGTGACGAGG.
shows that SHARP-1 preferentially binds to class B sites, but is also able to bind to class C sites.

**DISCUSSION**

In this chapter it has been shown that SHARP-1, a basic helix loop helix protein of unknown function, binds to the region of the M promoter between +166 and +417 in a yeast one-hybrid screen assay (Figure 16).

bHLH proteins are key players that regulate many aspects of development and differentiation in all tissues and phyla. Numerous vertebrate bHLHs (for reviews see (Anderson, 1999; Brunet and Ghysen, 1999; Kageyama *et al.*, 1997)) have been identified on the basis of homology to their *Drosophila* counterparts (reviewed in (Campos-Ortega, 1995; Campuzano and Modolell, 1992; Fisher and Caudy, 1998)). MASH1, a homologue of Achaete and Scute, is a transcriptional activator and acts as a positive regulator of neurogenesis. Disruption of *Mash1* in mice leads to the loss of specific subset of neurones (Guillemot *et al.*, 1993) and forced expression in rat neural crest stem cells leads to premature differentiation (Lo *et al.*, 1998). HES-1, a homologue of Hairy (h) and Enhancer of Split (E(spl)), is an example of a bHLH protein that acts as a transcriptional repressor and negatively regulates neuronal differentiation. Neural precursor cells infected with HES-1 transducing retrovirus fail to differentiate normally into neurones (Ishibashi *et al.*, 1994). Additionally, *Hes-1* null embryos exhibit severe neurulation defects and postmitotic neurones appear prematurely (Ishibashi *et al.*, 1994). These results suggest that HES-1 normally controls the proper timing of neurogenesis.
Structurally, bHLH proteins share a number of common features. The HLH domain mediates homommeric or heterommeric dimerisation and the adjacent basic region mediates DNA binding (Murre et al., 1989a). Three groups of bHLH proteins can be recognised, according to the target binding site they recognise (Blackwell et al., 1990; Dang et al., 1992; Ohsako et al., 1994). Class A and Class C bHLHs function as transcriptional activators and repressors respectively, while class B bHLH proteins can be either activators or repressors.

SHARP-1 is a bHLH protein that has recently been isolated on the basis of its homology to Hairy and Enhancer of Split (Rossner et al., 1997). However, sequence alignment showed that SHARP-1 is only distantly related to these proteins, exhibiting 37 to 42% sequence identity within the bHLH domain. Unlike all other h/E(spl)/HES proteins, SHARP-1 lacks the hallmark WRPW domain, which binds the co-repressor Groucho (or Groucho-like proteins) and is required for both transcriptional repression and suppression of neurogenesis. Absence of the WRPW motif suggests that SHARP-1 functions by recruiting transcriptional machinery other than Groucho.

Unlike most bHLH proteins, SHARP-1 is not expressed in neuronal progenitor cells or early differentiating neurones but is restricted to a subset of neurones of the post-natal CNS (Rossner et al., 1997) suggestive of a role in terminal neuronal differentiation rather than in neural determination. More specifically, most prominent expression of SHARP-1 at P40 is found in granule cells and Purkinje cells of the cerebellum, and in the medial habenulae (Rossner et al., 1997).

SHARP-1 and M1 genes have overlapping expression profiles, with both genes being expressed late in development, mainly in pyramidal neurones of the hippocampus, granule cells of the dentate gyrus,
striatum, and olfactory bulb (Buckley et al., 1988; Rossner et al., 1997), suggesting that SHARP-1 has a role in regulating expression of the M\(_i\) gene. SHARP-1 is also expressed at high levels in granule and Purkinje cells of the cerebellum, where expression of M\(_i\) is not detected. On the other hand, the M\(_i\) gene is highly expressed in the deeper cortical layers of the cerebral cortex, where expression of SHARP-1 is low. Over-expression of SHARP-1 reduced the levels of expression of a reporter gene driven by the M\(_i\) promoter in M\(_i\) expressing cells (Figure 19), suggesting that SHARP-1 functions as a transcriptional repressor to downregulate the levels of expression of the M\(_i\) gene in M\(_i\) expressing cells. Taken together, these results suggest that SHARP-1 does not have a role in specifying the expression profile of the M\(_i\) gene, but rather in controlling the levels of expression in a sub-population of M\(_i\) expressing cells.

In order to further investigate the function of SHARP-1, Gal4-SHARP-1 fusion proteins were produced, and their ability to modulate transcription from different promoters analysed. SHARP-1 was able to repress transcription when bound to both a TATA containing and a TATAless promoter (Figure 21) and furthermore that SHARP-1 is able to repress both basal and activated transcription (Figure 21). These data supports the idea that SHARP-1 functions as a repressor to regulate the levels of M\(_i\) expression in M\(_i\) expressing cells.

It is well documented that arrangement of basal promoter elements can profoundly influence the response of a promoter to different factors. The evolutionarily conserved Kruppel-associated box (KRAB) present in the N-terminal regions of most Kruppel-class zinc finger proteins specifically silences the activity of promoters whose initiation is dependent on the presence of a TATA box (Pengue and Lania, 1996) whilst initiator containing promoters are relatively unaffected. Similarly, Oct-2 isoforms only repress transcription from TATA-containing promoters
Here, we show that SHARP-1 is more promiscuous and can repress transcription driven from both TATA-containing and TATAless promoters (Figure 21). Repression action can also be influenced by the relative position of its DNA binding site to the promoter. This can be seen in the case of REST/NRSF (Chong et al., 1995; Schoenherr and Anderson, 1995) which represses transcription when recruited distally but can activate transcription when recruited proximally (Bessis et al., 1997) (Roopra & Buckley, unpublished observations). Here again, SHARP-1 appears to act as a more global repressor and is capable of repressing transcription from a TATA-containing promoter when recruited either proximally or distally. Nevertheless, repression is more marked when SHARP-1 is tethered proximal to the TATA box (Figure 21).

Transcriptional repression can take many forms (see Chapter 1 and (Maldonado et al., 1999)). Several repressors function by deacetylating the N-terminal tails of histones that are thought to render chromatin inaccessible to the transcriptional machinery. Such is the case of the MAD family of bHLH-ZIP proteins, whose members bind to mSIN3A, an adapter molecule which links histone deacetylases to DNA-bound transcription factors (Ayer et al., 1995; Laherty et al., 1997). Other repressors appear to interact directly with the transcriptional machinery itself, for example Kr, which at high concentrations homodimerises and becomes a potent repressor by interacting directly with TFIIE (Sauer et al., 1995; Sauer and Jackle, 1993). Others such as retinoblastoma (Rb) can interact with both general transcription factors or with histone deacetylase (Brehm et al., 1998; Dawson et al., 1995; Luo et al., 1998; Magnaghi-Jaulin et al., 1998).

Repression by SHARP-1 appears to be bimodal. One mode of repression is mediated via the bHLH domain and is TSA-insensitive while the other mode is mediated by the C-terminal domain and is TSA-sensitive (probably due to recruitment of HDAC). Interestingly, STRA13 has been shown to
repress expression of the c-myc gene through a HDAC-independent pathway and to negatively autoregulate its own promoter through a HDAC-dependent mechanism (Sun and Taneja, 2000). The region responsible for the HDAC-independent repression of STRA13 has not been mapped, but GST pull-down studies showed that residues 111-343 are required for interaction with HDAC-1, SIN3 and NCoR (Sun and Taneja, 2000) while functional analysis of mutated STRA13 demonstrated that residues 147-354 were required for STRA13 repression of VP16 activated transcription. As shown in Figure 17, this region contains the C-terminal domain and part of the Orange domain of SHARP-2/STRA13/DEC1. We have mapped the domain responsible for the TSA-sensitive component of SHARP-1 repression to the C-terminal domain (residues 174 to 253) (Figure 23). Within this region, SHARP-1 and STRA13 display pockets of homology, principally within two stretches of 8 (residues 204-211) and 15 amino acids (residues 222-236), suggesting that these may mediate the interaction with HDAC. The HDAC-independent repression domain of SHARP-2/STRA13/DEC1 has not been reported but the Groucho-independent mode of repression by Hairy/E(spl) requires the bHLH and Orange domains (Dawson et al., 1995). Although SHARP-1 and SHARP-2 are almost identical in their bHLH domains (95% over 59 residues) they diverge more markedly from Hairy (37% over 59 residues). It remains to be seen whether the same domains or mechanisms are involved in mediating HDAC-independent repression by SHARP-1, SHARP-2/STRA13/DEC1 and Hairy/E(spl) proteins.

Bimodal repression is not uncommon. Repression by REST/NRSF (Chong et al., 1995; Schoenherr and Anderson, 1995) is mediated by an N-terminal domain that recruits SIN3/HDAC-1 (Grimes et al., 2000; Huang et al., 1999; Naruse et al., 1999; Roopra et al., 2000) and a C-terminal domain that recruits a novel co-repressor Co-REST (Andres et al., 1999).
Repression by Rb is also mediated via an HDAC-1 dependent and HDAC-1 independent mechanism (Brehm et al., 1998; Dawson et al., 1995; Luo et al., 1998; Magnaghi-Jaulin et al., 1998). In the latter case, some promoters, such as the Adenovirus major late promoter, are repressed by the HDAC-1 dependent arm whilst transcription activated by the SV40 enhancer or driven by the tk promoter is repressed by the HDAC-insensitive arm. It may be that the breadth and selectivity of repressor action of bimodal repressors such as SHARP-1 is enhanced by their ability to bring one or both repression mechanisms to bear upon different promoters.

SHARP-1 contains an Arg in position 13 (R13) of the basic region present in all class B bHLH proteins. This residue is essential for class B proteins to bind the consensus sequence CA(C/T)GTG (Dang et al., 1992). A binding site selection assay showed that SHARP-1 binds preferentially to class B sites, as predicted by the presence of R13, but can also bind to class C sites (Figure 24). However, no class B or C site is present in the region of the M_i gene used in the yeast one-hybrid screen. SHARP-1 may recognise an element present in the M_i sequence distinct to either class B or C sites, but probably binds to this novel site with less affinity than it binds to class B or C sites, and for that reason it has not been isolated in the binding site selection assay. Other target genes for SHARP-1 remain to be identified, and the affinity of SHARP-1 for different regulatory elements may contribute to a mechanism to control the effect of SHARP-1 on different genes. The binding site for the transcription factor REST/NRSF, the RE-1/NRSE, contains 19-21 base pairs and up to five mismatches from the consensus sequence have been identified in different RE-1/NRSE containing genes (Schoenherr et al., 1996). Gel mobility shift assays using an RE-1/NRSE derived from the rat GABA-A rec. δ subunit gene, which contains five mismatches from the consensus sequence, yielded a complex with different mobility from a control RE-1/NRSE, suggesting that
REST/NRSF binds this probe with different affinity. Furthermore, when repression mediated by this site was compared to that mediated by the RE-1/NRSE derived from the SCG10 gene, it was found that repression mediated by the SCG10 RE-1/NRSE is at least 20 times stronger (Schoenherr et al., 1996). Gel shift binding assays using the region of the M1 region found to interact with SHARP-1 in the yeast one-hybrid screen have been unable to show binding to SHARP-1, possibly because this assay is not sensitive enough. However, co-transfections of SHARP-1 and a reporter vector containing the region of the Mi gene found to interact with SHARP-1 in the yeast 1-hybrid screen, showed that SHARP-1 is able to repress transcription in a manner that is dependant on the presence of this fragment (Figure 19), suggesting that SHARP-1 does bind to the Mi gene and regulates its transcription.
CHAPTER 6

GENERAL DISCUSSION
The mechanisms that control transcription of some pan-neuronal and some cell type specific genes have been determined. Whilst these studies do not provide a complete picture, they do provide insight into the mechanisms that control specification of neural phenotype. However, there are still many questions to be answered. For instance, what genes are involved in the determination of neuronal phenotype? Are there "master" regulators that can turn any cell into a neurone, or are there different regulators for each neuronal type?

Synapsins are a multigene family of neurone-specific phosphoproteins, which are the most abundant proteins in synaptic vesicles, and are involved in modulation of neurotransmitter release and synaptogenesis. *In situ* hybridisation assays have shown that each synapsin gene has a unique expression profile (Zurmohle *et al*., 1996). Synapsin I transcription is controlled by an RE-1/NRSE (Li *et al*., 1993). In contrast, a fragment containing 5.1 kb of 5'flanking region of the synapsin II gene, shown to be sufficient to drive tissue specific expression in transgenic mice, does not contain an RE-1/NRSE or any other recognisable motif (Chin *et al*., 1999). Therefore, these results represent an example where different mechanisms control transcription of two genes with largely overlapping expression profiles.

Tyrosine hydroxylase (TH) and dopamine β-hydroxylase (DBH) are two enzymes of the noradrenaline biosynthesis pathway, and expression of the two genes occurs during the same developmental stage (Ernsberger, 2000). It has been found that PHOX2 transcription factors are required for expression of the TH and DBH genes in the noradrenergic
system (Morin et al., 1997; Pattyn et al., 1999), showing that a common mechanism of transcription is responsible for the expression of two genes involved in defining the same neuronal phenotype. However, TH expression is not restricted to noradrenergic neurones, but is also expressed in dopaminergic neurones, suggesting that it is also regulated by mechanisms independent of DBH. These results represent an example of two genes that share a common transcriptional mechanism in those cells where the two genes are expressed.

REST/NRSF represents an example of a transcription factor involved in repression outside of the nervous system of some pan-neuronal genes, such as the sodium type II (Chong et al., 1995; Maue et al., 1990) and SCG10 (Mori et al., 1990; Schoenherr and Anderson, 1995) genes, and also neuronal-subtype specific genes, such as the M₄ mAChR (Mieda et al., 1997; Wood et al., 1996), β2 nicotinic AChR subunit (Bessis et al., 1995) and L1-CAM (Kallunki et al., 1997) genes. Transgenic mouse studies of the promoters of the β2 nicotinic AChR subunit (Bessis et al., 1997) and L1-CAM (Kallunki et al., 1997) genes have demonstrated that the RE-1/NRSE, as well as mediating repression in both neuronal and non-neuronal cells, can also mediate activation in subsets of neuronal cells. Therefore, these results represent an example of a single transcription factor regulating the expression of a battery of genes. Nevertheless, cell type specific expression must be achieved by the combinatorial action of REST/NRSF and other factors.

The principal means of intercellular communication within the mammalian brain is by the release of neurotransmitters from one neurone, which then interact with membrane receptors present on neighbouring neurones and consequently elicit a response. Each receptor is expressed in a specific cell population but the mechanisms that control these expression profiles are only beginning to be understood. The work
presented here has aimed to identify the mechanisms and factors that control transcription of one receptor subtype, the \( M_1 \) muscarinic acetylcholine receptor.

**mAChR gene structure**

The \( M_1 \) gene is one of the members of the muscarinic acetylcholine receptor gene subfamily (\( M_1 \) to \( M_5 \)), which themselves belong to the G-protein coupled receptor superfamily. To date, the gene structures for four members of the mAChR gene family have been identified; the rat \( M_4 \) (Mieda et al., 1996; Wood et al., 1995), the chicken \( M_2 \) (Rosoff et al., 1996), the rat \( M_1 \) (Pepitoni et al., 1997; Wood et al., 1999) and the chicken \( M_5 \) (Creason et al., 2000). All of these genes appear to have a similar gene structure consisting of a single coding exon and one or more 5' exons containing the 5'UTR. Both RNase protection analysis and RT-PCR on rat brain cortex cDNA have shown that the rat \( M_1 \) gene consists of a single coding exon of 2.54 kb and a single non-coding exon of 657 bp separated by an intron of 13.5 kb (Pepitoni et al., 1997; Wood et al., 1999) (Figure 1). RT-PCR analysis suggests a similar structure for the human \( M_1 \) gene, but indicates the use of an alternatively spliced exon in mouse cortex (Wood et al., 1999). However, studies on the structure of the rat \( M_1 \) gene by Klett and Bonner, suggest a different structure (Klett and Bonner, 1999). They found that the rat \( M_1 \) gene contains two non-coding exons, the first one spanning from \(-246\) to \(-119\) relative to the transcription start site identified by Pepitoni et al. (1997) (Figure 1) and a second one starting at position +489 and ending at the same position to that reported by Pepitoni et al. (1997). In order to determine the veracity of the transcription start sites mapped by the two labs, a RT-PCR assay was carried out on mRNA extracted from IMR32 cells, the neuronal \( M_1 \) expressing cell line used as model for the present study, mouse cerebral cortex and rat cerebral cortex (Wood et al., 1999). PCR using either of the sense primers RM, \(-212\) and
RM₁ -140 (numbers are relative to the transcription start site determined by Pepitoni et al. (1997)) in conjunction with an antisense primer in the coding exon did not produce amplified products from any of the cDNA samples used. However, PCR using the two primers RM₁ +299 and RM₁ +499 in conjunction with the same antisense primer did give amplified products. As a positive control, all sense primers were shown to produce amplified products from a genomic DNA template when used in conjunction with a second antisense primer. These results suggest that the exons described by Klett and Bonner (1999), are not used in the tissues and cell lines examined, but rather support the gene structure and exon usage proposed by Pepitoni et al. (1997). Therefore, the gene structure determined by Pepitoni et al. (1997)(graphically represented in Figure 1) has been assumed to be correct for the present study.

**The M₁, M₂ and M₄ genes are transcriptionally regulated through different mechanisms.**

In the central and autonomic nervous system, the expression patterns of the M₁ and M₄ 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 (Hassal et al., 1993; Wess, 1995) whereas in the striatum, over 80% of the neurones express M₁ and 40-50% express M₄ (Weiner et al., 1990). In addition, 85% of the large striatal interneurones and 90% of medium sized neurones co-express M₁ and M₄, resulting in a population of neurones that express one of the genes, both of the genes or neither (Bernard et al., 1992). Although the expression profiles of the M₁ and M₄ genes are largely overlapping, different mechanisms appear to control transcription of the M₄ gene. Cell specific expression of the rat M₄ gene is achieved, at least in part, through selective repression in M₄ non-expressing cells mediated by the binding of REST/NRSF to an RE-1/NRSE (Wood et al., 1996) (Mieda et al., 1997). Southern blot analysis using a
probe corresponding to the consensus RE-1/NRSE showed hybridisation to a cosmid containing the M4 gene, but not to a cosmid containing at least 9 kb of sequence 5' to the transcription start site of the M1 gene (Ian Wood, unpublished observations). Furthermore, sequence alignment of 32 kb of human genomic sequence containing the M1 gene plus 23 kb of flanking region (14 kb upstream and 9 kb downstream) with a consensus RE-1/NRSE failed to identify the presence of this site. These results suggest that the M1 gene does not contain an RE-1/NRSE and therefore the M1 and M4 genes appear to be regulated through distinct mechanisms.

Expression profiles of the M2 and M3 genes appear to be more disparate to that observed for the M1 gene, suggesting that although they are members of the same family, they are transcriptionally regulated through different mechanisms. In the case of the M2 mAChR, M2 transcripts are abundant in heart, but rare in the CNS, being only found in significant quantities in the medial septum and pons, with lower levels present in the thalamus (Buckley et al., 1988; Peralta et al., 1987). Both endogenous M2 AChR number and gene transcription have been found to be increased by CNTF and LIF in a neuronal cell line (Rosoff et al., 1996). Furthermore, basal levels of expression in heart have been found to be regulated by the GATA family of transcription factors (Rosoff and Nathanson, 1998). The vertebrate GATA family, which includes six members, is involved in diverse cell differentiation pathways during development (Simon, 1995). This family can be divided into two subgroups: GATA-1, -2 and -3 play a major role in the specification and differentiation of hematopoietic lineages (Orkin, 1996), whereas GATA-4, -5 and -6 are involved in the differentiation of heart and viscerae (Molkentin, 2000). GATA-2 and -3 are also expressed in the nervous system and have a key role in neuronal development (Nardelli et al., 1999). However, there is no evidence of GATA binding sites in the M1 gene (Pepitoni et al., 1997).
The transcriptional mechanisms that drive expression of other members of the muscarinic receptor gene subfamily have not been reported to date.

**Multiple elements within the non-coding exon control expression of the M₁ gene.**

In this study it has been shown that transcription of the M₁ gene is due to both positive and negative regulatory elements found within the non-coding exon. Within this region, the fragment between +169 and +605 of the M₁ gene appeared to contain more than one regulatory element and its deletion completely abolished expression (Figures 6 and 15). A similar profile of promoter activity was observed through the different deletions within this fragment for the M₁ expressing cell line (IMR32) and the non-neuronal M₁ non-expressing cell line (3T3), with the exception that changes in promoter activity were generally less marked in the M₁ non-expressing cell line. Therefore, no single element appears to be responsible for promoter specificity, but rather the fragment between +169 and +605 is required to recapitulate the expression profile of the endogenous gene. A similar effect is seen in the regulation of the GluR2 gene (Myers et al., 1998). The GluR2 promoter activity is 30-fold stronger in cultured cortical neurones than in primary glia, but none of the elements identified showed cell-type selectivity in transient expression assays.

**The PPY/PPU tract and a region conserved across species recruit nucleolin and act as cis-enhancer elements.**

Within the fragment between +166 and +602, multiple regions of the M₁ gene that are capable of driving the expression of a reporter gene in an M₁ specific manner have been identified. One of these regions is a polypyrimidine/polypurine sequence capable of forming single stranded
DNA in a supercoiled plasmid (Figure 7), and the other (found downstream of the PPY/PPU tract) is conserved across species, has no recognisable motif and is not able to form single stranded DNA by itself (Figure 7c), although it showed sensitivity to specific single stranded nucleases when located next to the PPY/PPU tract (Figure 9). Deletion of both the PPY/PPU tract and the conserved region reduced expression of a reporter gene in a M₁ expressing cell line, suggesting that they act as cis-enhancer elements (Figure 6). Screening of a cDNA library showed that both regions are bound by nucleolin (Figures 12 and 13), a multifunctional phosphoprotein that has been previously shown to be able to regulate transcription of other genes (Dempsey et al., 1998; Hanakahi et al., 1997; Yang et al., 1994).

**SHARP-1, a bHLH protein, downregulates expression of the M₁ gene in M₁ expressing cells.**

A second region important for expression of the M₁ gene has been identified to be bound by SHARP-1 (Figure 16), a bHLH protein of unknown function expressed in the adult nervous system (Rossner et al., 1997). Co-transfection assays showed that SHARP-1 downregulates expression of the M₁ gene in M₁ expressing cells (Figure 19). Consistent with these results, Gal4 fusion experiments showed that SHARP-1 is able to function as a repressor of both basal and activated transcription driven either by a TATA-containing or a TATA-less promoter in a position independent manner (Figure 21).

**Possible mechanisms by which nucleolin and SHARP-1 control expression of the M₁ gene.**

Using a sandwich-screening procedure that relies on protein-protein interactions to generate specific DNA binding activity, the ubiquitously expressed bHLH protein E47 has been shown to interact with
both a mouse homologue of SHARP-2, termed EIP1, and nucleolin (Dear et al., 1997). Although interaction between E47 and EIP1 was demonstrated by both immunoprecipitation and yeast two-hybrid assays, the interaction with nucleolin was not verified by any other means (Dear et al., 1997). SHARP-1 and SHARP-2 are highly conserved within the bHLH domain (only 3 residues out of 60 are not conserved) (Figure 18), suggesting that SHARP-1 may also interact with E47. Therefore, SHARP-1 may dimerise with E47 to bind to the M1 gene and interfere with the activity of nucleolin as a transcriptional activator by direct interaction through E47. This mode of repression is commonly used by many factors (see Chapter 1), as for example MYC-PRF, which binds to a DNA sequence in the c-myc gene adjacent to the site for the transcription activator MYC-CF1, masking its activation domain (Kakkis et al., 1989). E47 is a bHLH protein that can homodimerise or heterodimerise with multiple factors such as MyoD, MASH1, NGN1 and 2 and control different aspects of hematopoietic, muscular and neuronal development (Massari and Murre, 2000). Although E47 has always been related to transcriptional activation, it is not unusual for a transcription factor to have multiple functions depending on the nature of the promoter and cell conditions. YY1, as previously described, represents an example of a transcription factor with dual function. YY1 normally acts as a transcriptional repressor, but under certain circumstances, such as in the presence of E1A, it acts as an activator (Shi et al., 1991). In order to inhibit the transcriptional activation function of nucleolin, SHARP-1/E47 dimers would need to be bound proximal to nucleolin, therefore, this type of repression would only be observed at promoters where E47, SHARP-1 and nucleolin are all present. Consistent with this hypothesis, we find that a construct containing only the PPY/PPU tract and conserved region drives higher levels of expression than a construct which also contains the upstream sequence known to be bound by SHARP-1 (Figure 15, compare pGL3 +396/+569 M1 Inr and pGL3
However, without showing a direct interaction between SHARP-1 and nucleolin, the possibility that SHARP-1 acts as a classical repressor cannot be excluded.

Alternatively, or in addition to the mechanism just suggested, SHARP-1 may downregulate M₁ expression by altering chromatin structure and therefore preventing binding of the PIC or other factors important for expression, such as nucleolin. Deletion analysis has shown that SHARP-1 contains two independent repression domains, one at the C-terminus, which acts by a mechanism sensitive to the HDAC inhibitor Trichostatin A, and the other at the bHLH domain, which works through a TSA insensitive mechanism. Recruitment of histone deacetylases leads to a more compact, repressive state of chromatin (Kornberg and Lorch, 1999; Wolffe and Guschin, 2000). Chromatin structure may have an important effect on the binding of nucleolin, since this factor recognises single stranded DNA. Chromatin modifications induced by the binding of SHARP-1 or other yet unidentified factors may inhibit the opening of the PPY/PPU tract and conserved region, therefore preventing the binding of nucleolin. Although the PPY/PPU tract and conserved region are able to form single stranded DNA in vitro, it remains to be determined if this single stranded conformation is present in vivo both in M₁ expressing and non-expressing cells.

Role of nucleolin and SHARP-1 in specification of neuronal phenotype.

Nucleolin is a ubiquitously expressed protein, which can be post-translationally modified by phosphorylation (Borer et al., 1989; Rossner et al., 1997). The presence of the unphosphorylated form only in M₁ expressing cells suggests that only this form is able to activate expression of the M₁ gene. Studies of the M₁ promoter have therefore identified a
factor that is not neurone-specific, but nevertheless has a role in controlling the expression of a neurone-specific gene. Since its function depends on post-translational modifications, it will be important to study the signals that induce phosphorylation or dephosphorylation of nucleolin, in order to understand the role of these signals in regulation of \( M_1 \) expression.

The involvement of bHLH proteins in neuronal development became evident with genetic studies in *Drosophila*, which identified several bHLH proteins responsible for the proper development of the nervous system (reviewed in (Campuzano and Modolell, 1992; Fisher and Caudy, 1998)). Vertebrate homologues of those *Drosophila* factors have been identified and numerous studies have demonstrated the evolutionary conservation of their expression patterns and cellular functions (Guillemot, 1999; Shi *et al.*, 1997). However, in several cases, the developmental step regulated by the gene product shows marked differences across the phyla. Some genes, such as *Mash1* and *Math1*, are expressed in dividing cells, suggesting that they are involved in the specification of neural precursor populations, while others, such as *NeuroD*, are expressed mostly in postmitotic precursors or differentiated neurones, suggesting roles in the differentiation of neural precursors or the maintenance of the differentiated state. *SHARP-1* expression starts late in development (Rossner *et al.*, 1997), suggesting that this factor belongs to the second family discussed. Accordingly, SHARP-1 controls the levels of expression of the \( M_1 \) muscarinic acetylcholine receptor gene, which encodes one of the many neurotransmitter receptors that define neuronal phenotype.

In rats, very low levels of *SHARP-1* are first detected in brain at E18, and start to rise after P5 (Rossner *et al.*, 1997). Expression of the \( M_1 \) gene, however, is first observed at E12.5, and the adult spatial expression
profile is established by P2. However, postnatal protein concentration keeps rising for several weeks (Buckley, unpublished observations). The different temporal expression patterns observed for the two genes suggests that SHARP-1 does not have a role in demarcation of the spatio-temporal expression profile of \( M_i \), but rather may control absolute levels of expression.

Other target genes for SHARP-1 remain to be identified. In addition to controlling the levels of expression of neural specific genes, such as the \( M_i \) gene, SHARP-1 may also have a role on "switching off" transcription of transiently expressed neuronal genes, including those genes coding for transcription factors required at earlier stages of development. Most transcription factors involved in development of the nervous system present a very stringent spatio-temporal expression profile, which corresponds to defined phases of cell fate determination and differentiation. For example, expression of \( Mash1 \) precedes expression of \( Phox2a \), which precedes expression of the GDNF receptor \( c-ret \) (Lo et al., 1998). At the same time, in \( Hes-1 \) knock-out mice, \( Mash1 \) is prematurely expressed and upregulated in certain regions, suggesting that HES-1 has a role in controlling the spatio-temporal expression profile of \( Mash1 \) (Ishibashi et al., 1995).

It will thus not only be important to identify the target genes for SHARP-1, but also those factors that control expression of the \( SHARP-1 \) gene, in an attempt to draw a picture of the hierarchy of factors that control the establishment of neuronal phenotype.
REFERENCES


phosphatidylinositol 4,5-bisphosphate hydrolysis using purified m1 muscarinic receptor, Gq/11, and phospholipase C-beta 1. *J Biol Chem*, 267, 8081-8088.


Orrick, L.R., Olson, M.O. and Busch, H. (1973) Comparison of nucleolar proteins of normal rat liver and Novikoff hepatoma ascites cells by two-dimensional polyacrylamide gel electrophoresis. *Proc Natl Acad Sci U S A*, 70, 1316-1320.


Pengue, G. and Lania, L. (1996) Kruppel-associated box-mediated repression of RNA polymerase II promoters is influenced by the
arrangement of basal promoter elements. *Proc Natl Acad Sci U S A*, 93, 1015-1020.


Sun, H. and Taneja, R. (2000) Stra13 expression is associated with growth arrest and represses transcription through histone deacetylase (HDAC)-


