Sheffield Hallam University

A Sheffield Hallam University thesis
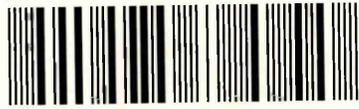
ProQuest Number: 10694518

ProQuest 10694518

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Protein factors and 5' flanking sequences involved in the expression of the mouse myelin basic protein gene.

Andrew Charles Fairclough

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the Degree of Doctor of Philosophy

JUNE 2001

Abstract

Myelin Basic Protein is a major structural protein of vertebrate myelin. The gene that codes for MBP is contained within the *golli-MBP* complex. This gene complex consists of two overlapping transcription units, *golli* and *MBP*, which are regulated by two distinct promoters.

The *golli* unit is expressed in cells of the oligodendrocyte lineage (Central Nervous System), neurons, B and T lymphocytes, testis and thymus. However, the *MBP* unit is expressed exclusively in oligodendrocytes and Schwann cells (Peripheral Nervous System).

The expression of the MBP unit is regulated mainly at the level of transcription by proteins that bind in a specific manner to DNA sequences located within its promoter region. The identification of these proteins and DNA sequences is essential to understanding the mechanisms that regulate the transcription of the MBP unit.

This project was initiated by the isolation of the putative promoter region of the mouse myelin basic protein (MBP) gene. To achieve this the *Hind III – Sac I* fragment of pEX1 plasmid was subcloned in the vector pBluescript. The cloned insert, which corresponds to the region between nucleotides – 1319 and + 227 relative to the transcription start site of the mouse MBP gene, was subsequently sequenced manually using the chain termination method. Sequence analysis revealed a number of putative binding sites for transcription factors. The region – 609 to – 577 was selected for further studies because work published by other groups suggested that it contains a cell-type specific (for oligodendrocytes) transcription activator. The presence of protein factors specifically binding to the region – 609 to – 577 was demonstrated by electrophoretic mobility shift assay (EMSA).

For this purpose, nuclear extracts were prepared from rodent brain or established glial cell lines e.g. C_6 glioma cells. Extracts from tissues and cell lines, which do not express myelin basic protein e.g., HeLa cells served as a control. Nuclei were isolated by Dounce homogenisation of cultured cells or brain tissue. The proteins were then isolated by high salt extraction of the nuclei followed by ammonium sulphate fractionation.

Putative protein(s) binding to the region located between nucleotides – 609 to – 577 of the myelin basic protein gene promoter were identified using the yeast one-hybrid system. This assay is based on the interaction between a specific protein DNA binding domain and the target DNA sequence. Proteins are expressed as fusions to the GAL4 activation domain (AD) in the yeast reporter strain in which the target sequence has been inserted upstream of the *HIS3* gene minimal promoter. Binding of AD fusions to the target sequence increases activity of the HIS3 promoter enabling growth on medium lacking histidine. In this work a yeast reporter strain containing four copies of the –609 to –577 region tandemly repeated upstream of the *HIS3* gene minimal promoter was constructed. A library containing rat brain cDNAs fused to the activation domain of GAL4 was screened using this strain as a host. Seven clones were obtained on medium lacking histidine in the presence of 30 mM 3-aminotriazole. DNA from these clones was automatically sequenced and analysed for sequence homology with known transcription factors by comparing the nucleotide and protein sequences to EMBL/Genbank and Swissprot/Swall databases using the

FastA and Blast search tools. From the results of the homology searches the clones were identified as follows: the activating transcription factor 2 (ATF-2), the pituitary specific positive transcription factor 1 (Pit-1) or general transcription factor 2i, the E2F family transcription factor, the PASK protein and two of the clones were identified as *c-jun*. One clone, however, remains unidentified and this could be a novel transcription factor.

Acknowledgements

I would like to dedicate the work in this thesis to the memory of my Mother and Father who sadly passed away before seeing this work completed, and also to thank them for their constant support and encouragement throughout this project.

I am most grateful to Dr Maria Blair for her supervision of this project, and helpful criticism whilst writing this thesis. I would like to thank Dr Barry Davis for his support throughout the project and comments on the thesis. My thanks also go to Dr Aruna Asipu and Dr Karin Wiebauer for their advice on methods, and friends and colleagues for their support during this project.

Finally I would like to acknowledge the School of Science and Mathematics for allowing me the opportunity to study this topic.

Contents

LIST OF FIGURES

xii

LIST OF TABLES

LIST OF ABBREVIATIONS

AIX plates   Ampicillin / IPTG / X-gal plates

3 – AT   3 – aminotriazole

ATP   Adenosine triphosphate

cDNA   Complementary DNA

CNP   2',3'-Cyclic nucleotide 3'-phosphodiesterase

Et-Br   Ethidium Bromide

EMSA   Electrophoretic mobility shift assay

GC   Galactosyl cerebroside

IPTG   isopropyl - β - D – thiogalactoside

kb   Kilobase

kDa   KiloDaltons

MEBA   Myelinating glia-enriched DNA binding activity

MRF-1   Myelin regulation factor-1

η Refractive index

ND-GE   Non-denaturing gel electrophoresis

NFI	Nuclear factor-1
nt	Nucleotide(s)
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase chain reaction
PLP	Proteolipid protein
SD	Synthetic Dropout
SDS	Sodium Dodecyl Sulphate
TBE	Tris Borate EDTA
TEMED	N,N,N',N' - Tetramethylethylenediamine
TRE	Thyroid hormone response element
U	Units (enzyme activity)
X-Gal	5 – bromo – 4 – chloro – 3 – indolyl - β - D - galactoside

COURSES, CONFERENCES AND RELEVANT EXPERIENCE

Teaching Experience and Responsibilities:

Undergraduate Teaching:

Lecturing on DNA sequencing to final year students.

Practical demonstrations on PCR, oligonucleotide synthesis and DNA sequencing to 2nd and final year students.

Co-supervision of HND and final year B.Sc students undertaking projects in the field of Molecular Biology.

Postgraduate Teaching:

An annual lecture on DNA sequencing to students on the M.Sc Pathological Sciences Course.

Co-supervision of MSc students undertaking research projects in the field of Molecular Biotechnology.

Short Courses:

Lecturing and practical demonstrations on DNA sequencing, oligonucleotide synthesis and PCR on short courses on DNA technology / PCR techniques, run by the Division of Biomedical Sciences for senior employees of L.E.P. Scientific, Knoll Pharmaceuticals and members of other academic and industrial institutions.

Research presentations:

Presentation of a research seminar on "Myelin Basic Proteins and their importance in Multiple Sclerosis and Alzheimer's Disease", to the academic and research staff at Sheffield Hallam University.

Presentation of a research seminar on my PhD work, "Protein factors and 5' flanking sequences involved in the expression of the mouse myelin basic protein gene", to the academic and research staff at Sheffield Hallam University.

Consultancy Work:

The successful running of an oligonucleotide synthesis service – the customer base includes, Sheffield Hallam University, Leeds University, Sheffield University and the Northern General Hospital, Sheffield

Courses and Conferences attended:

626th Meeting of the Biochemical Society (in particular the Education Group Workshop relating to Teaching Materials for Molecular Biology.)

Pharmacia DNA Sequencing Workshop (a one day workshop covering all the aspects of DNA sequencing).

Recombinant DNA Technology a refresher course.

Seminar on RT-PCR presented by Scientific Director Clontech Laboratories Inc.

Seminar on New Developments in DNA Sequencing presented by Associate Director United States Biochemical Corporation.

Seminar on expression systems with relevance to *Pichia pastoris* presented by R and D Systems.

Course on the Theory, Optimisation and Application of PCR presented by Perkin-Elmer/Applied Biosytems.

In-situ PCR one day workshop and practical session presented by Fraser Lewis and Perkin-Elmer/Applied Biosytems.

Chapter 1

Introduction

The principal role of the myelin sheath is to allow the faster transmission of the nerve impulse along the axons, which it surrounds, by acting as an insulator. The membrane of the myelinated axon expresses several physiologically active molecules that intervene directly in the transmission of the nervous impulse e.g. the voltage-sensitive sodium channels are responsible for the propagation of the action potential along the axon and are concentrated at the non-insulated axonal segments known as the nodes of Ranvier (Figure 1), and the fast and slow potassium channels responsible for various secondary electrophysiological activities involved in the recovery to resting-state conditions following the passage of the action potential (nerve impulse) (Waxman and Ritchie, 1993; Salzer, 1997). The myelin sheath, through its electrical resistance and low capacitance, allows the depolarisation of the internodal axonal membrane with a minimal consumption of energy (Sutcliffe, 1987). Due to the segmental nature of the electrically insulating myelin membrane, the action potential appears to jump from one node of Ranvier to the next: hence the term 'saltatory conduction'. By imposing a saltatory conduction on the nerve impulse, myelin not only allows an increased speed of conduction, it also saves energy and space, e.g. in order to achieve an equal conduction velocity, an unmyelinated axon would need to have a 40 times greater diameter than a myelinated one and would consume 5000 times more

energy (Garbay *et al.*, 2000). The myelin sheath is produced in the central nervous system (CNS) by oligodendrocytes and in the peripheral nervous system (PNS) by Schwann cells. At the outset it consists of loosely spirally wrapped extensions of the oligodendrocyte or Schwann cell plasma membrane. During maturation of myelin, compaction at both the apposed cytoplasmic and extracellular membrane surfaces occurs to form the tight multilamellar structure characteristic of myelin (Lemke, 1986). This compacted, spirally wrapped multilayer membrane system consists of approximately 80% complex lipids and cholesterol, and 20 % protein (Streicher and Stoffel, 1989). The decision to myelinate is critical both for the functioning of the nervous system and for the physiology of the sheath cell. For the cell, the formation of myelin entails a radical reorganisation of both morphology and metabolism, the most obvious being the massive increase in plasma membrane biosynthesis and surface area, and in the induction and high-level expression of a set of genes and proteins unique to the myelinating glia (Lees and Brostoff, 1984).

Myelination is a major event in the development of the nervous system of higher animals. In the central nervous system of rats and mice it occurs postnatally and follows a period of rapid proliferation of oligodendroglial cells. Myelination generally proceeds from the top of the spinal column and the base of the brain towards the frontal areas of the brain.
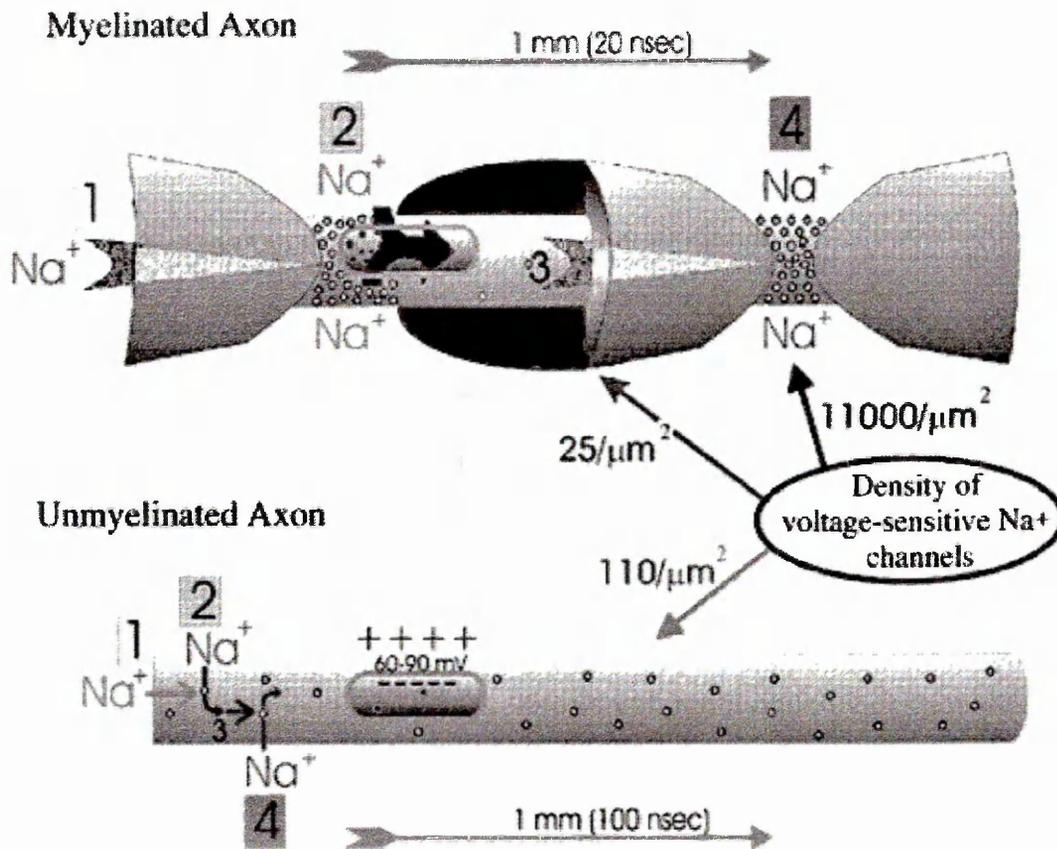
Figure 1. A diagrammatic representation of nerve conduction in a
myelinated and unmyelinated axon.

KEY:

1.   The arrival of an action potential in the form of a wave sodium ions.
2 and 4. Entry of sodium ions by the activation of sodium channels.
3.   The passive diffusion of the wave of sodium ions along the axon
            following the concentration gradient.

(After Garbay et al., 2000)

In the myelin of the central nervous system the major proteins are the basic proteins and the proteolipid protein (representing 70 - 80% of the protein content of the membrane). The proteolipid protein plays both a structural role in the association of apposed extracellular myelin membrane surfaces, and functions as an ionophore. Both the myelin basic protein and the proteolipid protein consist of multiple polypeptide chains derived through alternative splicing of single genes (deFerra *et al.* 1985). Less abundant CNS proteins include myelin – associated glycoprotein (MAG) and 2', 3' – cyclicnucleotide 3'-phosphodiesterase (CNP). Schwann cells form the myelin of the peripheral nervous system; with the major proteins being P_0 glycoprotein and myelin – associated glycoprotein (MAG) (Campagnoni, 1988).

P_0 is an integral membrane glycoprotein serving as a bi-functional structural element linking adjacent lamellae and therefore stabilising the myelin assembly. The myelin-associated glycoprotein has been localised to the periaxonal regions and is involved in the association of the myelin membrane with the axon, with its structure related to a cell adhesion molecule (Salzer *et al.*, 1987).

The myelin basic proteins (MBPs) represent a family of structurally related proteins, varying in molecular mass from 14 to 21.5 kDa (deFerra *et al.*, 1985, Newman *et al.*, 1987a; 1987b). The mouse myelin basic protein isoforms are referred to by their apparent molecular weights on SDS polyacrylamide gels or by their masses obtained through cDNA or protein sequencing studies. There are four isoforms in humans (Kamholz *et al.*, 1986, Roth *et al.*, 1986, Roth *et al.*, 1987).

The myelin basic proteins are hydrophilic extrinsic membrane proteins with isoelectric points greater than 10.6 (more basic than the histones), which have been localised to the major dense line of myelin that is formed by apposition of the cytoplasmic surfaces of the extruded oligodendroglial plasma cell membrane during myelinogenesis (Morell *et al.*, 1994). Unlike the proteolipid proteins, which are believed to be transmembrane proteins traversing the unit bilayer, the myelin basic proteins appear to remain associated with the cytoplasmic side of the unit bilayer. The myelin basic proteins undergo a number of post-translational modifications, including *N*-terminal acetylation, phosphorylation and methylation. Evidence suggests that methylation of the myelin basic protein may be important for the compaction of the membrane during myelin maturation (Amur *et al.*, 1986, Kim *et al.*, 1997).

1.1. Myelin Basic Protein gene structure

Cell-free translation of rat and mouse brain mRNA indicated that the major isoforms of the myelin basic proteins were encoded by separate MBP mRNAs (Yu and Campagnoni, 1982 and Colman *et al.*, 1982) and this understandably led to investigations into whether or not more than one gene codes for the multiple myelin basic protein isoforms. Southern blot analysis showed the presence of only a single gene in the mouse and the human, suggesting these isoforms, which are homologous over most of their sequences, are produced from a single primary MBP transcript through alternative patterns of RNA splicing (Takahashi *et al.*, 1985 and deFerra *et*

5

al., 1985). *In situ* hybridisation and somatic cell hybridisation studies have mapped the gene to the distal end of chromosome 18 in both the mouse (Roach *et al.*, 1985) and the human (Saxe *et al.*, 1985 and Sparkes *et al.*, 1987).

The mouse *MBP* gene is large and part of a complex genetic locus, termed the *golli-mbp* locus, (Figure 2). The term *golli-mbp* signifies the fact that most of the products of the **g**ene are expressed in the **ol**igodendrocyte **li**neage in the brain as **m**yelin **b**asic **p**roteins. Structural analysis has shown this locus to be conserved among most species (Pribyl *et al.*, 1993). The murine *golli-mbp* is distributed over a length of 215 kb and comprises 11 exons, with exons 5b – 11 encoding the myelin basic proteins. The human locus is smaller (179 kb) with 10 exons. The exons are interrupted by introns some of which are quite long (Takahashi *et al.*, 1985 and deFerra *et al.*, 1985).
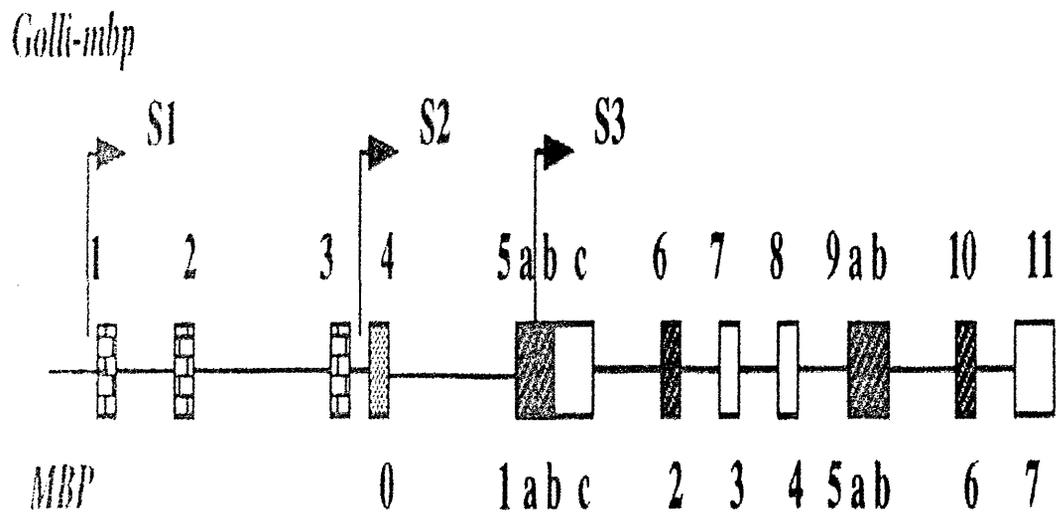
Figure 2. A diagrammatic representation of the mouse *MBP* gene and its relationship to the *golli-mbp* gene.

Exons 1 to 11 represent the *golli-mbp* gene (marked above the horizontal line representing the introns).

MBP represents the 'classical' *MBP* gene comprising exons 0 to 7 (marked below the horizontal line representing the introns).

Boxes indicating the exons are interrupted by introns indicated by the solid lines. Window-type boxes represent exons unique to *golli-mbp* transcripts. The stippled box represents an embryonic exon; transcripts containing this exon are initiated at the S2 site. Cross-striped boxes represent the exons shared by 'classical' MBP isoforms.

S1 represents the transcriptional start site of the *golli-mbp* gene transcripts.
S2 corresponds to the transcriptional start site for embryonic *MBP* gene transcripts.
S3 represents the transcriptional initiation site of the 'classical' *MBP* gene transcripts.
Alternative splicing of the primary transcripts results in a diversity of golli-mbp and MBP isoforms.

(After Asipu and Blair, 1997).

7

The MBP mRNA coding unit of the *golli-mbp* gene spans 32 kb. Exon 5a is a region of high transcriptional complexity, functioning as a promoter for the transcription of the *MBP* gene. Exon 5a is also the site for the alternative splicing of several of the *golli-mbp* transcripts. Not unexpectedly, this region is highly conserved among sharks, mice and humans (Fors *et al.*, 1993). Several laboratories have isolated cDNAs of mRNAs encoding a wider variety of MBP isoforms than had at first been anticipated, (deFerra *et al.*, 1985, Kamholz *et al.*, 1986, Roth *et al.*, 1986, Roth *et al.*, 1987, Newman *et al.*, 1987a, Kitamura et al., 1990, Aruga et al., 1991, Nakajima et al., 1993) for example, cDNAs of mRNAs encoding twelve forms of the mouse MBP have now been isolated. These cDNAs are derived from mRNAs that arise through the alternative splicing of the second, fifth and sixth exons of the 'classical' *MBP* gene primary transcript. The relationship between their molecular masses and exon usage is outlined as follows: the 14 kDa isoform contains exons 1b, 3, 4, 5b and 7, the 17 kDa – 1 isoform utilises exons 1b, 2, 3, 4, 5b and 7, the 17 kDa – 2 isoform utilises exons 1b, 3, 4, 6, and 7, the 18.5 kDa isoform contains exons 1b, 3, 4, 5b 6 and 7, and the 21.5 kDa isoform contains all 7 exons (Figure 3). Immunoblot studies have indicated the existence of additional minor myelin basic proteins such as a 13 kDa form (containing exons 1b, 3, 4 and 7), a 15 kDa form (utilising exons 1b, 2, 3, 4, and 7) and a 20 kDa form (containing exons 1b, 2, 3, 4, 6 and 7) in the mouse CNS, although it has been suggested that these two isoforms are predominantly embryonic (Nakajima et al., 1993). It is also possible that the minor MBPs are posttranslationally modified molecules or degradation products. Other data suggest that there may be even more alternatively

spliced forms of MBP mRNAs than have been reported. For example in a study carried out by Aruga *et al.* (1991) two myelin basic protein mRNAs were identified that had a novel internal exon (5a) that could putatively encode a 19.7 kDa and a 21 kDa myelin basic protein utilising exons 1b, 2, 3, 4, 5a, 5b and7 and exons 1b, 3, 4, 5a, 5b, 6 and 7 respectively. Exon 5a encodes 22 amino acid residues, which contain six valines, and forms a hydrophobic domain; therefore, the 21 kDa myelin basic protein (including exon 5a) might form a similar structure to that of the 18.5 kDa MBP. Kitamura et al., (1990) also identified an isoform utilising exons 0, 1a, 1b, 3, 4, 5, and 7, which encodes a 14 kDa myelin basic protein.

Complementary DNAs have been isolated that encode two different 17 kDa MBP isoforms that are so close in molecular mass that the corresponding proteins would be indistinguishable by SDS polyacrylamide gel electrophoresis. However, from the abundance of the two 17 kDa MBP cDNA clones in a cDNA library, it appears that one of the isoforms is present in greater abundance than the other at 18 days of postnatal development. Of five cDNA clones encoding a 17 kDa mouse MBP isolated in one study one was missing exons 2 and 5 and four were missing exon 6 (Newman *et al.*, 1987a).
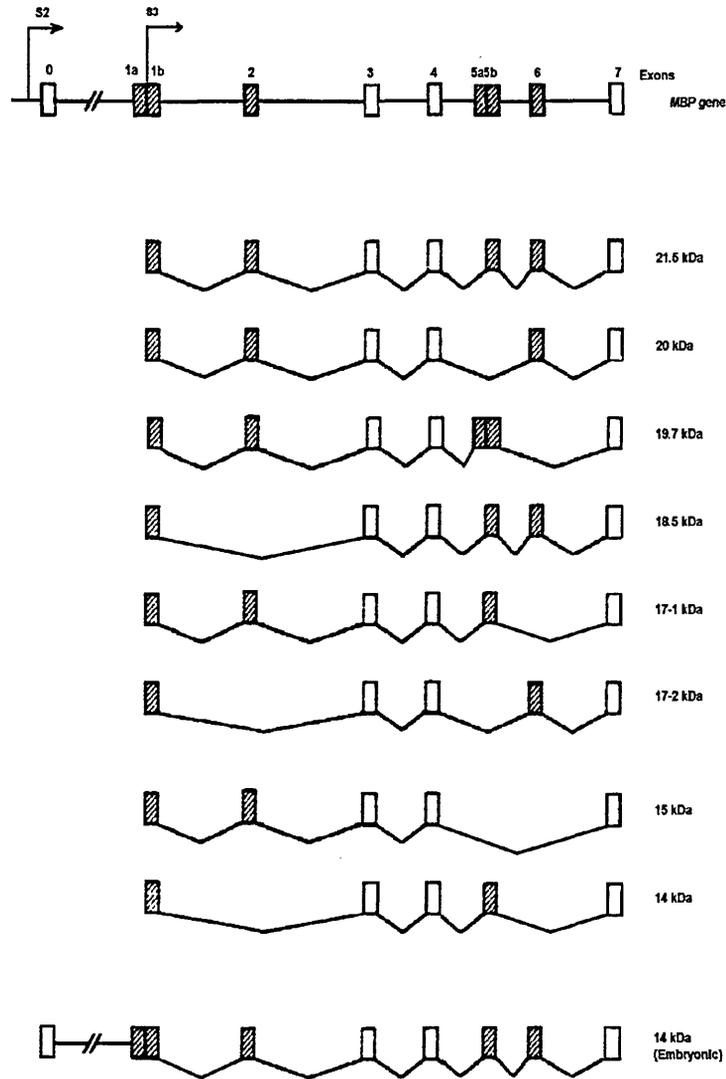
Figure 3. *MBP* gene and transcript structures.

The alternatively excised exons are shaded.
S2 corresponds to the transcriptional start site for embryonic *MBP* gene transcripts.
S3 represents the transcriptional initiation site of the 'classical' *MBP* gene transcripts.

(After Zelenika *et al.*, 1993).

The sizes of the exons are exaggerated and are not necessarily proportional.

10

All the MBP mRNA structures published to date contain a relatively short 5' untranslated region (<48 nucleotides) and a very long 3' untranslated region (>1 kb). On Northern blots, the 'classical' MBP mRNAs migrate as a broad band of approximately 2.0 to 2.4 kb, presumably reflecting the heterogeneity of the MBP mRNA population. There are two polyadenylation signals near the 3' end of both mouse and human MBP mRNAs. In addition, there is a second AUG codon in the 5' untranslated region 5 bases upstream of the initiator codon, which is immediately followed by a termination codon. This structural feature has been postulated to explain the poorer translation initiation efficiencies exhibited by MBP mRNAs relative to brain mRNAs as a whole (Campagnoni *et al.*, 1987).

1.2. Myelin Basic Protein gene transcription

1.2.1.Transcriptional control of gene expression

The first step in gene expression is RNA synthesis (which takes place in the nucleus). RNA polymerase II catalyses transcription of mRNA from the 5' to the 3' direction and uses the non-coding (antiparallel) strand of DNA as a template. Transcription can be divided into three stages, initiation, elongation and termination. In order for a gene to be expressed it must be placed under the control of a promoter. Promoters are relatively short sequence elements usually located in the immediate upstream region of the gene, relative to the transcription start site, and serve to initiate transcription

11

i.e. the site at which RNA polymerase initially binds to the DNA. Initiation of

transcription requires the binding of protein factors (*trans*-acting transcription

factors) to *cis*-acting regulatory nucleic acid sequences.

The structure of a typical eukaryotic promoter comprising up to four distinct

transcriptional control regions, the core promoter, upstream promoter

elements, regulatory elements and enhancer elements is shown

schematically in Figure 4.

RNA polymerase II is responsible for transcribing the genes encoding

polypeptides and certain species of snRNA genes (important for splicing).

RNA polymerase II is dependent on general transcription factors such as

TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. In promoters containing a

TATA box TFIID binds to this element and protects a region from −35 to −19

bases upstream of the transcription start site. The binding of TFIID to the

TATA box is the earliest step in the formation of a stable transcriptional

complex, such binding being facilitated by TIIFA. TFIIA may also play a role

in the response to transcriptional activators, acting as a co-activator

molecule linking DNA-bound activators and the basal transcriptional

complex. TFIIB joins the complex by binding to TFIID. This binding of TFIIB

is an essential step in initiation complex formation since TFIIB can also bind

to RNA polymerase II; hence, it acts as a bridging factor allowing the

recruitment of RNA polymerase to the complex in association with TFIIF.

Following the binding of RNA polymerase II, three other transcription factors

(TFIIE, TFIIH and TFIIJ) associate with the complex. At this point TFIIH,

which has DNA helicase activity, unwinds the double-stranded DNA allowing

it to be copied into RNA. Subsequently, the kinase activity of TFIIH

phosphorylates the C-terminal domain of RNA polymerase II, which converts

it to the phosphorylated form, which is capable of transcriptional elongation

to produce the RNA product. This complex of polymerase and general

transcription factors is known as the basal transcriptional complex, therefore,

TFIIH via its helicase and kinase activities plays a crucial role in allowing the

basal transcriptional complex to initiate transcription. Genes are

constitutively expressed at a minimum rate determined by the core promoter,

unless altered by additional positive or negative regulatory elements. Core

promoter elements are usually located close to the transcription initiation

site, from nucleotide positions –40 to +40, and include elements such as the

TATA box (consensus sequence <u>TATA</u>(A/T)A(A/T) located at approximately

position –25). If a promoter does not contain a TATA box (e.g. housekeeping

genes) an order of bases in the region covering the start site is recognised

by the transcriptional machinery as the transcription initiator region.

Upstream promoter elements are typically located from nucleotide positions

–50 to –200 relative to the transcription start site, and include multiple

recognition sites for some sequence-specific ubiquitous transcription factors

e.g. GC or Sp1 boxes (consensus sequence GGGCGG) and CCAAT boxes

(consensus sequence GG<u>CCAAT</u>CT typically located at nucleotide position

–75 and recognised by the protein factor CTF or NF-1). GC and CCAAT

boxes serve to up-regulate basal transcription of the core promoter.

Regulatory elements function to modulate transcription in response to

specific external stimuli (inducible gene expression) and are located

upstream of the promoter usually within 1 kb of the transcription start site

e.g. glucocorticoid response elements (GRE) or cyclic-AMP response

elements (CRE). Small hydrophobic hormones e.g. the steroid hormones diffuse through the plasma membrane of the target cell and bind to intracellular receptors in the cytoplasm or the nucleus. These receptors are inducible transcription factors. Following binding of the ligand, the receptor protein associates with a specific DNA response element e.g. the glucocorticoid response element (GRE) located in the promoter of the target gene and activates transcription. In the absence of the ligand, the receptor is inactivated by direct repression of the DNA-binding domain by the ligand-binding domain or by binding to an inhibitory protein as in the case of the glucocorticoid receptor.
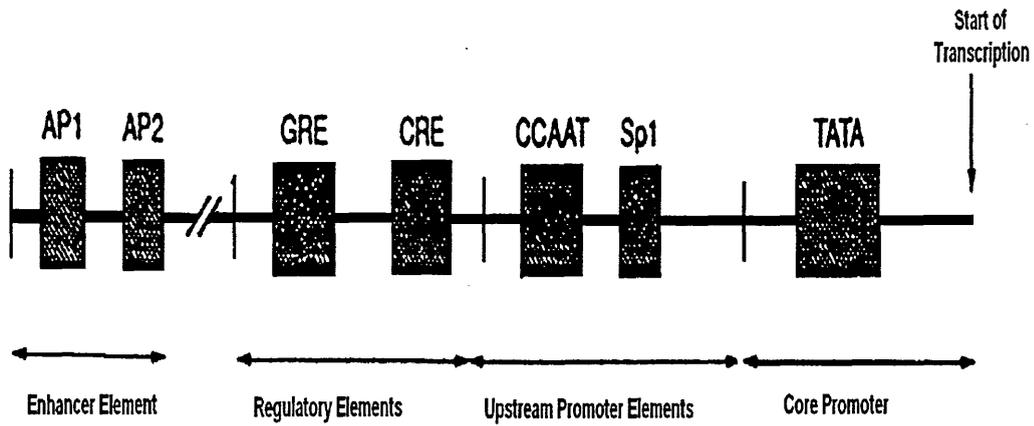
Figure 4. Schematic representation of a typical eukaryotic promoter showing distinctive transcriptional control elements.

(After Latchman, 1998)

The second messenger cyclic-AMP is synthesised from ATP by adenylate cyclase (a membrane bound enzyme). Binding of a hormone e.g. epinephrine to the cell surface receptor promotes the interaction of the receptor with a G protein. This results in the activation of the α subunit of the G protein causing it to dissociate and activate adenylate cyclase. The increase in intracellular cAMP activates the transcription of specific target sequences containing the cAMP response element (CRE). This function of cAMP is mediated by protein kinase A. Cyclic AMP binds to protein kinase A and activates it by releasing two catalytically active subunits which enter the nucleus and phosphorylate a specific transcription factor CREB (CRE-binding protein). Activated CREB activates the transcription of genes with the cAMP response element.

Enhancers are positive regulatory elements serving to increase the basal level of transcription that is initiated through the core promoter irrespective of their position and orientation.

Some genes encoding polypeptides are housekeeping genes, however, a large percentage of genes transcribed by RNA polymerase II show tissue-specific expression patterns. Since the DNA in different nucleated cells of an individual is essentially identical, the identity of a cell is defined by the proteins made by the cell. Therefore, tissue-specific transcription factors regulate the expression of many genes that encode polypeptides by recognising and binding specific cis-acting sequence elements. Tissue-specificity and developmental-stage specificity of gene expression are frequently conferred by enhancer and silencer sequences

16

and a variety of *cis*-acting sequences have been identified which are specifically recognised by tissue-specific transcription factors. In addition to actively promoting tissue-specific transcription, some *cis*-acting silencer elements confer tissue or developmental stage specificity by blocking expression in all but the desired tissue e.g. the neural restrictive silencer element (NRSE) represses the expression of several genes in all tissues other then neural tissue (Schoenherr *et al.*, 1996). A transcription factor binding to NRSE and called the neural restrictive silencer factor (NRSF) or RE-1 silencing transcription factor (REST) is ubiquitously expressed in non-neural tissue and neuronal precursors during early development but specifically not expressed in postmitotic neurons.

Transcription factors recognise and bind to short nucleotide sequences generally as a result of very specific molecular recognition between the surface of the protein and surface features of the DNA double helix. In eukaryotic transcription factors two functional domains have been identified, one for DNA-binding and one for transcriptional activation, that retain their function when removed from their natural context. The DNA-binding domains of transcription factors fall into several conserved structural motifs e.g. helix-turn-helix, helix-loop-helix, zinc finger and basic DNA binding domain. Each of the motifs use α-helices (or occasionally β-sheets) to bind to the major groove of DNA. Transcription factors bind to DNA as either homodimers or heterodimers, with the DNA-binding region of the protein distinct from the region responsible for forming dimers. The activation domain functions in activating transcription of the target gene once the transcription factor has bound to the promoter.

17

Zinc finger proteins are so called because of their structure, in which an atom of zinc tetrahedrally coordinates the repeated cysteine and histidine (or four cysteine) residues that form the base of a loop of twelve amino acids containing conserved leucine and phenylalanine residues, as well as several basic amino acids. The zinc finger motifs have been shown to constitute the DNA-binding domain of the protein, with DNA-binding being dependent on their activity. Most notable among the zinc finger proteins are *Sp1* and the steroid receptor family.

The helix-turn-helix motif is a common motif found in homeoboxes and a number of other transcription factors. It consists of two short α-helices separated by a short amino acid sequence that induces the turn. It is thought that while the HTH motif in general mediates DNA binding, the more C-terminal helix acts as a specific recognition element fitting into the major groove of the DNA and controlling the precise DNA sequence that is recognised.

Leucine zipper proteins derive their name from a helical stretch of amino acids rich in leucine residues precisely seven amino acids apart. However, unlike the zinc finger or helix-turn-helix motifs the zipper itself is not the DNA-binding domain of the molecule and does not directly contact the DNA. Rather it facilitates DNA binding by an adjacent region of the molecule, which is rich in basic amino acids and can therefore interact directly with the DNA. The leucine zipper is believed to serve an indirect structural role in DNA binding, facilitating dimerisation, which in turn results in the correct positioning of the two basic DNA-binding domains in the dimeric molecule for DNA binding to occur. Hence the DNA-binding

18

specificity of the leucine zipper-containing transcription factors is determined by the sequence of their basic domain with the leucine zipper allowing dimerisation to occur. As expected with this notion, the basic DNA binding domain can interact with DNA in a sequence-specific manner in the absence of the leucine zipper, if it is first dimerised via an inter-molecular disulphide bridge. The helix-loop-helix (HLH) motif plays a similar role to the leucine zipper family allowing dimerisation of the transcription factor molecule (as either homodimers or heterodimers), and thereby facilitating DNA binding by the basic domain.

1.2.2. Tissue specific expression of the MBP gene

Non-typical TATA and CAAT boxes are present in the MBP promoter, located at positions -34 and -85 bp respectively and a non-typical GC box sequence has been identified at position -92 bp from the transcriptional initiation site at $+1$. The TATA box-like sequence is indispensable for the promoter function. Tamura *et al.*, (1988) suggested that the GC box functions co-operatively with upstream sequences including the myelin basic protein transcription element (MBTE) a *cis*-acting element located between nucleotide positions -130 to -106 relative to the transcription initiation site. The MBTE was crucial to direct maximal transcription, and also functioned with a heterologous promoter in a direction-dependent manner. Tamura *et al.*, (1988) identified a ubiquitous binding factor, which interacted specifically with the MBTE and activated transcription. Intensive footprinting studies (Tamura *et al.*, 1988) demonstrated that the MBTE had an NFI-binding

19

sequence. The MBTE was considered to be one of the strongest NFI-binding motifs among known cellular genes. Interestingly, similar strong NFI-binding motifs are present in the enhancer of JC virus whose genes are expressed, like the MBP gene in the nervous system.

By fusing segments from different regions of the 1.3 kb 5' flanking region to the chloramphenicol acetyltransferase (CAT) reporter gene it has been observed that the proximal regulatory sequence from −50 to −14 bp (which includes the TATA box-like sequence) constitutes the basal transcription element (Asipu and Blair, 1994; Devine-Beach et al., 1992, and Tamura et al., 1990). The myelin basic protein promoter also contains a number of DNA specific cis-acting elements, which interact with specific trans-acting factors (DNA binding proteins) present in both oligodendrocytes (OL) and Schwann cells (SC). The interactions of these factors, with each other and with the RNA polymerase II initiation complex, directly regulate the tissue-specific and developmental expression of MBP mRNA. Proof that specific DNA sequences within the MBP gene can support tissue-specific, developmentally regulated MBP gene expression has been demonstrated in vivo by the construction of transgenic mice carrying an MBP transgene. Kimura et al. (1989) used an MBP transgene consisting of 1.3 kb of the MBP 5' flanking sequence fused to an MBP cDNA encoding the 14 kDa MBP isoform to demonstrate that the transgene was expressed predominantly in the brain in a tissue-specific manner. Tamura et al. (1989) reported specific transcription from the MBP promoter in brain nuclear extracts. They also observed that there are no differences between mouse and rat brain nuclear extracts in specific transcription of the MBP gene. Tamura et al. (1989)

20

divided the MBP promoter into two domains referred to as the distal

promoter from –253 to –54, and the proximal promoter from –53 to +60. The

distal region was able to direct tissue-specific transcription when linked to

the heterologous adenovirus type 2 major late core promoter, suggesting

that the distal region contains some tissue-specific *cis*-acting elements and

that the brain is rich in the cognate *trans*-acting factors. The MBP distal

promoter region appears to function in a direction-dependent manner. Only a

native orientation of the distal region linked to the major late core promoter

or the MBP proximal promoter can enhance transcription in brain-derived

nuclear extracts. Cooperative interaction of transcription factors in the distal

and proximal regions may be required for tissue specific transcription.

Alternatively, inhibitory sequences in an upper part of the distal region may

contribute to the different profiles of transcription activation. Transcriptional

inhibitory sequences associated with the distal region may be dominant in

non-expressing tissues. Wrabetz *et al.* (1993) transfected primary cultures of

both oligodendrocytes and Schwann cells, as well as a number of cell lines

not expressing MBP, with a series of human MBP promoter-CAT reporter

gene constructs and monitored CAT activity as a measure of MBP promoter

activity. They demonstrated in both OL and SC that the 150 bp region

upstream of the MBP cap site is enough to promote CAT expression in a cell

type-specific manner, and that this expression is mediated by multiple

positive and negative acting modules. Wrabetz *et al.* (1993) also confirms

the results of Tamura *et al.* (1989) that the MBP distal promoter region

functions in an orientation-dependent manner. These transfection studies

suggested that there are cell type-specific *trans*-acting factors present in

both OL and SC that interact with specific DNA sequences within the first 150 bp of the MBP promoter region to increase MBP transcription thereby contributing to the tissue-specific developmentally regulated expression of MBP. Work carried out by Asipu and Blair (1994 and 1998) showed that cell-type specific transcription of the *MBP* gene in primary oligodendrocytes (OL) and Schwann cells (SC) is regulated by *cis*-acting regulatory elements both upstream and downstream of the TATA-like box region of the MBP promoter. The cell-type specific regulatory elements were defined using constructs containing varying lengths of the 5' flanking region and exon 1 of the *MBP* gene linked to the CAT reporter gene followed by transfection into several cell types e.g. primary cultures of differentiated OL, SC, and neuronal and non-neuronal cell lines. Several upstream cis-acting regulatory elements were identified, with sequences located between nt –655 to –397(including the region between nt –609 to –577 studied in this project) and nt –394 to –54 showing enhancer and repressor effects respectively in oligodendrocytes. Whereas, in Schwann cells the sequence located between nt –394 to –253 showed a positive regulatory effect whilst those between nt –655 to –397 and –253 to –54 showed negative regulatory effects. Downstream cis-acting regulatory elements were identified between nt positions +20 to +70 and +70 to +200 showing strong silencer and enhancer activities respectively, specifically in oligodendrocytes (Asipu and Blair 1994). The expression of *MBP* is therefore controlled by basal positive regulatory elements which function in different cell types, but both positive and negative regulatory elements located on either side of the basal promoter play an essential role in regulating the transcription of *MBP* in

22

myelin-forming oligodendrocytes (Asipu and Blair 1994). To identify the cell-type specific factors binding to the downstream (relative to the transcription initiation site) regulatory elements in more detail Asipu and Blair used DNase I footprinting analysis and gel retardation assays (EMSA). Two regions from nt −17 to +17 and +47 to +58 were protected specifically in oligodendrocytes whereas three regions nt +17 to +22, +43 to +49 and +58 to +64 were protected only with C6 nuclear extracts. Analysis of the protected regions for homology with known transcription factor binding sites showed the sequence between +47 to +58 and +56 to +68 had extensive homology to the negative regulatory element of the mouse rennin gene and to the interferon (IFN) consensus sequence of the MHC class I genes respectively. Several other *trans*-acting factors have been identified binding to the *cis*-acting regulatory elements (both upstream and downstream of the TATA-like box region of the MBP promoter) and these include: the thyroid hormone α-receptor (TRα) at nt −209 to −169 (Farsetti *et al.* 1992 and Asipu and Blair 1997), the M1 factor from mouse brain and rat liver extracts at nt −110 to −97 (Aoyama et. al. 1990), the Ets-consensus sequence at nt +31 to +39 (Farsetti *et al.* 1992), the viral (adenovirus, polyoma) enhancer sequences and the human β-interferon regulatory element between nt +41 to +48 (Goodbourn *et al.* 1986). Other regulatory sequences have been identified which show homology to previously identified prokaryotic and eukaryotic transcription factor binding sites e.g. the PEA3-consensus site (AGGAGG) at nt position −283, the *E coli* MalT binding site and the PER3 site of the mouse peripherin promoter at nt −216 (Desmaris *et al.* 1992) and the AP1 transcription factor binding site at nt −93.

The *golli-mbp* gene has multiple promoters; as a result the developing brain appears to make use of these multiple promoters to control the transcriptional initiation site, to confer tissue-specific and developmentally regulated expression of the *MBP* gene. Expression initiated by one promoter also appears to regulate the expression from a different promoter (Schibler and Sierra, 1987). The promoter controlling the transcriptional initiation site S3 (Figure 2) located between *golli-mbp* exons 5a and 5b is active in the brain during myelination. Evidence currently available from studies in mice suggest the *golli-mbp* gene is transcribed into at least three primary transcripts (Campagnoni *et al.*, 1993), with each primary transcript undergoing alternative splicing resulting in three sets of mRNA transcripts. Each transcript is of a different length and also differs in spatial and temporal expression. For example, the G_1 primary transcript undergoes alternative splicing of *golli-mbp* exons G5a, 5b, 6, 9 and 10 with the resulting transcripts covering the size range 2.3 to 5.1 kb. These mRNA transcripts are generally expressed from the embryonic to the neonatal period. The 5.1 kb mRNA transcript is expressed from the embryonic state to the early post-natal period. The 2.3 to 2.6 kb mRNA isoforms are expressed post-natally and then follow the developmental profile of the 5.1 kb mRNA. Expression of these mRNA transcripts is not solely confined to the brain but they are also expressed in several other tissues including the bone marrow, thymus, spleen, heart, kidney and lung. It is thought that in the case of myelin basic protein gene transcription, inhibition may be relieved at the onset of myelination when a specific trans-acting factor(s) is activated from the S3 initiation site.

1.3. Developmental changes in Myelin Basic Protein expression

With development, myelin isolated from the brains of mice and rats is increasingly enriched in the major myelin proteins relative to the higher molecular weight polypeptides. In addition it has been consistently noted that the 14 : 18.5 kDa ratio of MBP isoforms increased considerably with age (Campagnoni and Macklin, 1988). No clear consensus has emerged with respect to changes in the ratio of proteolipid proteins to basic proteins with development (Campagnoni and Macklin, 1988).

Developmental studies on the expression of myelin proteins in whole brain have largely fallen into two categories:

(a) those measuring changes in the myelin proteins either through biochemical isolation of the protein or immunoblot analysis;

(b) those measuring synthesis of the myelin proteins *in vivo* or *in vitro* (i.e. cell-free synthesis) or by titering the mRNAs with cDNA probes. Clearly all these techniques reveal important information generally related to the expression of the myelin genes, but individual techniques measure different aspects of expression.

Investigators have used different criteria to define the order of expression of the myelin proteins. Some workers have attempted to rely on the earliest age at which a protein could be detected to determine its order of expression, however, differences in the sensitivities of the antibodies, or the specific activities of cDNA probes could lead to misleading results. Another complication is the finding that alternative splicing mechanisms may generate related but not identical mRNAs containing common sequences but

25

encoding different proteins. Therefore, early detection of mRNAs hybridising to myelin protein cDNAs may be detecting structurally related but not identical mRNAs. The same problems exist with antibodies that might react with related but separate proteins.

The extent of the multiplicity of the MBP isoforms was not fully appreciated until immunochemical and recombinant DNA approaches were used to study the problem. In the late 1970s, immunochemical techniques revealed the presence of four major mouse MBPs with apparent molecular masses of 14, 17, 18.5 and 21.5 kDa (Barbarese et al., 1977). Subsequent work using immunoblots of the whole brain to measure the different MBP isoforms, and cell-free synthesis of brain mRNA to measure the levels of individual MBP mRNAs indicated that each isoform exhibited its own developmental pattern of expression and accumulation (Carson et al., 1983).

Sequence analysis has proved that the rat 14 kDa MBP was structurally similar to the 18.5 kDa bovine and human MBPs except for the deletion of a 40- amino acid sequence within the interior of the molecule. In retrospect, there was evidence for the presence of other MBP variants (Agrawal et al., 1986, and Sorg et al., 1986) but it was difficult to rule out the possibility that these were degradation products of genuine MBPs.

From studies carried out during the late 1980s other MBP isoforms have been identified for example, a second mouse 17 kDa MBP has been identified (Newman et al., 1987a). This second 17 kDa MBP is apparently the only form present in humans (Kamholz et al., 1986; Roth et al., 1986; Roth et al., 1987 and Deibler et al., 1986). A 20 kDa MBP has been identified in foetal human spinal cord, and immunoblots suggest it exists in

26

other species as well (Kerlero de Rosbo *et al.*, 1984) although it may be more prevalent in some regions of the CNS than others (Newman *et al.*, 1987b, Nakajima *et al.*, 1993). In rat brain, Agrawal *et al* (1986) have identified a 23 kDa phosphorylated protein that is immunologically related to MBP. Immunoblots of whole brain homogenate with both polyclonal (Carson *et al.*, 1983) and monoclonal (Bansal *et al.*, 1987) anti-myelin basic protein antibodies indicate the presence of higher molecular weight immunoreactive proteins. The structural relationship among all these proteins and the better-characterised MBP isoforms is not clear at the present time.

In the course of studying promoter activity in primary cultures of brain cells Nakajima *et al.* (1993) found that the myelin basic protein gene was expressed in brain cells obtained from embryonic mice at a very early stage of neural development, i.e., as early as embryonic day 12 (E12). The predominant myelin basic protein isoforms in the embryonic stage were identified as exon 5-deleted forms, and peaked during the late embryonic stage (days E15 to E18) and decreased to very low levels at postnatal day 18 (P18). The cell type producing the early embryonic myelin basic proteins (days E12 to E14) remains to be clarified. If, however, the myelin basic proteins are produced in cells of the same cell lineage as those producing in the later stage, there may be a group of cells that are committed to generate oligodendrocytes later and produce the MBPs. The existence of this type of committed cell at this early stage of development has been suggested by Luskin *et al.* (1988).

1.4. Sites of myelin protein synthesis and transport to the membrane

The majority of evidence favours the view that the myelin basic proteins are synthesised on free polysomes, (Campagnoni *et al.*, 1980, Campagnoni, 1985, Brophy *et al.*, 1993) and that the proteolipid proteins are synthesised on membrane bound ribosomes (Benjamins and Morell, 1987, Colman *et al.*, 1982). Neither the MBP nor the PLP are synthesised with "signal" (or "leader") sequences.

The subcellular trafficking pathway for MBP mRNA in the oligodendrocyte has been defined using micro-injection experiments (Ainger *et al.*, 1993) and involves three discrete steps: (a) assembly of the RNA into granules in the perikaryon, (b) transport along cellular processes and (c) localisation to the myelin compartment. Ainger *et al.*, (1997) have identified specific regions in the 3' untranslated region (UTR) of myelin basic protein mRNA that are required for the latter two steps. The first region is the RNA transport signal (RTS), which is a 21-nucleotide sequence that is required for transport of mRNA along oligodendrocyte processes. The second region is the RNA localisation regions (RLR) and is necessary for localisation of MBP mRNA to myelin. This region may be implicated in stabilising or anchoring MBP mRNA in the myelin compartment of the oligodendrocyte (Ainger *et al.*, 1997). The presence of the RTS in oligodendrocyte process localised isoforms of myelin oligodendrocyte basic protein (MOBP) 18A RNA suggests the RTS functions to target the expression of other proteins, in addition to MBP, in oligodendrocytes (Holz *et al.*, 1996). If the RTS

28

represents a *cis*-acting transport signal, which is conserved in different RNAs, it most probably interacts with a cognate *trans*-acting factor that is expressed in different cell types. *In vitro* binding experiments indicate that the RTS sequence specifically interacts with hnRNP A2, a member of a family of heterogeneous nuclear ribonucleoprotein (hnRNP) proteins that shuttle between the nucleus and cytoplasm (Hoek *et al.*, 1998, Kwon *et al.*, 1999, Shan *et al.*, 2000). This suggests that hnRNP A2 binding to the RTS is required for the transport of RTS containing RNA in oligodendrocytes (Ainger *et al.*, 1997). Localisation involves translocation from the oligodendrocyte processes to the myelin; this apparently involves dissociation from the transport apparatus in the processes and in the case of coding mRNAs association with the translational machinery in the myelin compartment. Translocation across the boundary between these two compartments is RLR dependent for coding RNA and RLR independent for non-coding RNA. A possible explanation for this could be that coding RNA associates with ribosomes, making translocation RLR dependent, while translocation of ribosome-free non-coding RNA is RLR independent (Bacher *et al.*, 1996). The RNA localisation regions contain a predicted stable secondary structure at a conserved position in rat, mouse and human MBP mRNA (Ainger *et al.*, 1997). Endogenous MBP RNA-containing granules in oligodendrocytes contain many components of the translational machinery, suggesting they also represent the unit of RNA translation. Therefore, the granule represents a common unit for transport, localisation and translation of RNA (Barbarese *et al.*, 1995).

1.5. Translational and Post-Translational Events

Cell-free synthesis studies have shown that the myelin basic protein mRNAs are poorly initiated during translation relative to the other brain mRNAs as a whole, and this appears to be related to the presence of an additional translational initiation signal within the 5' untranslated region of the MBP mRNAs which is immediately followed by a termination signal (Campagnoni *et al.*, 1987a). The role of post-translational modification such as acylation, phosphorylation, methylation, glycosylation, acetylation, deamidation, and attachment of fatty acids has been shown to play a significant role in directing the function of various proteins. In CNS myelin basic protein, these post-translational modifications include: the methylation of a specific arginine residue, the phosphorylation of specific serine and threonine residues, the acylation of the N-terminal amino acid, the deamidation of a glutamine residue and the conversion of arginine residues to citrulline by the enzyme peptidyl arginine deiminase.

1.5.1. Methylation

Myelin basic protein has an unusual amino acid at residue-107, a mixture of N^G-monomethylarginine and N^G, N'^G-dimethylarginine. The formation of these methylarginine derivatives is catalysed by one of the protein methyltransferases, utilising *S*-adenosyl-L-methionine as the methyl donor. It is now apparent that each protein methyltransferase with a specificity towards any given amino acid residue can be further subclassified

according to the specific methyl acceptor protein (Kim *et al.*, 1997). Protein methylase I is one such enzyme, however, it is only recently that unequivocal molecular evidence for the presence of subclasses of the enzyme e.g. MBP-specific and nuclear protein/histone-specific has been obtained (Ghosh *et al.*, 1988, Rajpurohit *et al.*, 1994). The N^G-methylarginines occur in nature in a limited number of highly specialised proteins such as MBP, nuclear and contractile proteins. The reaction is very specific for the amino acid sequence around the methylation site, e.g. among 18 arginine residues in MBP, only the arginine at residue-107 (in the sequence Lys-Gly-Arg-Gly-Leu) is present as a mixture of MMeArg and Di(sym)MeArg. Park *et al.*, (1986) demonstrated that the histone specific protein methylase I was incapable of methylating MBP and was inhibited by MBP. The activity of the MBP-specific methylase has been observed to be abnormal in both the *jimpy* and *shiverer* mouse mutants (Kim *et al.*, 1984, and Kim *et al.*, 1986).

Multiple sclerosis is one of the demyelinating diseases in humans and as a result of this myelin basic protein is dissociated from the membrane. MBP fragments formed by intracellular proteolysis and free amino acids find their way into the body fluids (Park *et al.*, 1986). Since methylarginines are not reutilised for *in vivo* protein synthesis, it would appear practicable to assess myelin abnormalities by analysing urinary concentrations of these methylarginines and MBP-like material in multiple sclerosis patients (Whitaker et al., 1993, 1994).

The specific location of methylarginine in MBP (residue-107), together with the chemical nature of hydrophobicity enhanced by methylation, has

been implicated in helping to stabilise the myelin structure (Park *et al.*, 1986). Cells treated with a methylation inhibitor e.g. sinefungin produced a membrane structure lacking compactness which resembled vitamin B_{12}-deficient myelin with a concomitant reduction in MBP-protein methylase I activity (Kim *et al.*, 1997).

It is of interest that the only methylated amino acid residue (residue-107) in MBP identified to date lies within the region of the protein encoded by exon 5. Therefore, presumably the 20 and the 17 kDa MBP (missing exons 2 and 5) and present in the very early stages in the developing human spinal cord, are incapable of being methylated.

1.5.2. Phosphorylation

Phosphorylation is a ubiquitous post-translational modification of proteins, which is particularly prominent in nervous tissue. Myelin has traditionally been regarded as metabolically inert with a wholly structural role. However, recent findings indicate it possesses biochemical components associated with signal transduction in which protein phosphorylation plays a significant part (Eichberg and Iyer, 1996). The myelin basic proteins possess multiple consensus sequences for phosphorylation by cyclic AMP-dependent protein kinase (PKA) and protein kinase C (PKC). In fact MBP has been shown to be a substrate for Ca^{2+}-calmodulin-dependent protein kinase, glycogen synthase kinase-3 (known as protein kinase F_A) and mitogen associated protein (MAP) kinase (Yung, 1986, Yu and Yang, 1994, Hickson

et al., 1990). Variable phosphorylation is responsible in part for the charge

microheterogeneity of myelin basic protein e.g. the phosphorylation of thr[97]

in bovine MBP has been shown to increase conformational order for the

protein in aqueous solution which is consistent with the stabilisation of the β-

pleated sheet structure (Deibler et al., 1990). The dynamic state of the

phosphate groups suggests that phosphorylated MBP would be

concentrated in non-compacted areas of myelin with close proximity to the

cytoplasm (Des Jardins and Morell, 1983).

The introduction of negative charges by phosphorylation of the major

myelin proteins during deposition may create electrostatic repulsion between

adjacent proteins or between proteins and neighbouring acidic lipids.

Phosphorylation of MBP may in part serve this purpose, a notion supported

by the finding that protein phosphorylation activity is inversely proportional to

the degree of myelin compaction (Deibler et al., 1990).

In mature myelin, it would be expected that the protein kinases would

be concentrated in the regions of myelin closest to a supply of cytosolic ATP,

i.e. the outer, inner and lateral loops in CNS myelin and the paranodal loops

and Schmidt-Lanterman clefts in PNS myelin. This would suggest that the

fraction of myelin protein molecules undergoing phosphorylation would be

confined to a population resident in these regions, and that in the case of

MBP and P_0, this proportion would be small. However, evidence suggests

the turnover of phosphate groups in both these proteins occurs in compact

myelin, suggesting that the myelin structure may be less rigid than generally

assumed so that cytoplasmic components may gain at least limited access

throughout (Lowery et al., 1989, Ledeen, 1992). In fact, in mature myelin, the extent of protein phosphorylation may modulate the penetration of cytosolic constituents between myelin lamellae. An indication of a possible functional relationship between cell signalling mechanisms and myelin protein phosphorylation is provided by the finding that cultured oligodendrocytes subjected to hypoxic conditions show evidence of an 80% decrease in MBP and CNPase phosphorylation, as well as a failure of agonist stimulated PLC-mediated phosphinositide breakdown (Qi and Dawson, 1993). Nevertheless, the remaining phosphorylating ability retains its sensitivity to stimulation by phorbol esters, suggestive of PKC activity. Eichberg and Iyer (1996) therefore conclude that diminished MBP and CNPase phosphorylation may result from impaired signal transduction via heterotrimeric GTP binding proteins and as a result such a mechanism may contribute to the vulnerability of myelination to hypoxia.

1.5.3. Deimination

Deimination of myelin basic protein has been implicated in the chemical pathogenesis of multiple sclerosis. Degradation of bovine myelin basic protein by cathepsin D, a myelin-associated protease, was increased when 6 arginyl residues were deiminated and became even more rapid when all 18 arginyl residues were deiminated (Pritzker et al., 2000). Myelin basic protein containing methylated arginine has been shown to increase the association of lipid vesicles into dimers, signifying that methylated arginine

34

may increase the interaction between MBP and the myelin lipids (Young et al., 1987). In subacute combined degeneration (SCD) of the spinal cord in humans, demyelination results from a deficiency in vitamin B_{12}, which decreases the amount of S-adenosyl-L-methionine (the major methyl donor) with a subsequent decrease in the methylation of MBP (Surtees, 1993). The conversion of arginine to citrulline by the enzyme peptidylarginine deiminase has been implicated in demyelination in multiple sclerosis (Moscarello *et al.*, 1994). In addition, the severity of MS has been correlated with the number of arginine residues converted to citrulline in MBP (Wood *et al.*, 1996). Pritzker *et al.*, (2000) show that human MBP isolated from normal and MS white matter contained greater amounts of citrulline in MBP in which the proportion of unmethylated arginine-107 was increased.

1.6. Myelin Protein Gene Expression in Cultured Cells

Myelin protein gene expression has been studied in a variety of cultured cell systems. Primary cultured oligodendrocytes differentiate to produce galactosyl ceramide (GC), CNP, MBP, PLP, and other myelin markers in the apparent absence of any neural signals, their differentiation process *in vitro* is somewhat more "plastic" than *in vivo*, and differences among studies may result from this plasticity. An important early element of this plasticity is whether the oligodendrocyte will differentiate from its progenitor cell. Raff *et al.*, (1983) reported the existence of a cell type in rat optic nerve primary cultures, which can differentiate into a type 2 astrocyte in

35

the presence of foetal calf serum or into an oligodendrocyte in defined media with no foetal calf serum. This oligodendrocyte progenitor cell appears to stop dividing and begin to differentiate into oligodendrocytes or type 2 astrocytes within 2 to 3 days in culture whereas *in vivo*, these progenitor cells continue to divide (Raff *et al.*, 1983 Temple and Raff, 1985).

From studies on mixed glial cultures, it has been possible to determine that the rat oligodendrocyte progenitor cells (by the late foetal period) are committed to express myelin markers two or three weeks later in culture. Neurons are not required to induce the synthesis of MBP, PLP, CNP, GC, or cerebroside sulphotransferase (McCarthy and deVellis, 1980; Pfeiffer *et al.*, 1981b, Dubois-Delacq *et al.*, 1986). During the early stages of myelination *in vivo* studies have shown that the 14 kDa MBP is expressed at lower levels than the 18.5 kDa MBP. With development, the ratio of the 14 : 18.5 kDa MBP increases significantly until the 14 kDa isoform predominates. In contrast in culture the 14 kDa MBP is expressed first, and it remains the predominant isoform up to 39 days in culture. However, the other isoforms accumulate from 27 - 39 days in culture (Barbarese and Pfeiffer, 1981). Conflicting results have been reported on the timing of MBP expression in cultured rat oligodendrocytes (Pfeiffer *et al.*, 1981b, Dubois-Delacq *et al.*, 1986); however, a possible explanation for the differences may provide insight into the regulation of MBP expression.

Rat oligodendrocytes in mixed glial cultures express approximately 20-fold more MBP and CNP per oligodendrocyte than do isolated oligodendrocytes purified away from astrocytes and other cells in the mixed

culture (Bhat *et al.*, 1981). When the isolated oligodendrocytes are added back to the astrocyte-containing cultures, they express the higher levels of MBP and CNP. This suggests that the presence of cell(s) and / or factor(s) in mixed primary glial cultures plays some role in regulating the expression of these two myelin components. Bhat and Pfeiffer (1986) established that a soluble astrocyte factor(s), which is non-dialyzable, heat-labile, and trypsin-sensitive, enhances MBP and CNP expression in purified oligodendrocytes.

1.7. Gene Expression in Dysmyelinating Mutants

For two of the mutants, *shiverer* and *jimpy*, the mutation has been localised to the structural genes for the myelin basic protein and the myelin proteolipid protein, respectively (Roach *et al.*, 1985, and Sidman *et al.*, 1985).

1.7.1. Shiverer (*shi*) and Myelin Deficient (*shi^{mld}*)

The *shiverer* mutation (*shi*) and its allele myelin deficient (*shi^{mld}*) (Doolittle and Schweikart, 1977 and Doolittle *et al.*, 1981) are autosomal recessive, and map to the mouse chromosome 18 (Roach *et al.*, 1985 and Sidman *et al.*, 1985). These mice exhibit tremors beginning approximately at day 12 (postnatally), which become progressively worse. Convulsions appear in older animals and these animals do not survive past 90 days

(Chernoff, 1981). Morphological analyses indicate that many axons have no myelin, and those that are myelinated have only low amounts of myelin. A striking feature of *shiverer* CNS myelin is the absence of the major dense line but the presence of the intraperiod line (Inoue *et al.*, 1981). This suggests there is a greater loss of myelin basic protein in these animals than of proteolipid protein or of other proteins expressed on the extracellular surface of myelin. Biochemical analysis of *shiverer* tissue indicates essentially the total absence of myelin basic protein in *shiverer* brain (Bourre *et al.*, 1980 and Barbarese *et al.*, 1983). As with *shiverer* animals, shi^{mld} mice appear to have a major deficit in MBP. Myelin is poorly compacted, and the major dense line of myelin is essentially absent (Matthieu *et al.*, 1980). However, more MBP (between 3% and 10% of normal levels) was observed in shi^{mld} brains than in *shiverer* brains. The level of MBP mRNA in shi^{mld} mice was quite low, approximately 2 - 5% of normal in young animals, (Ginalski-Winkelman *et al.*, 1983; Campagnoni *et al.*, 1984; Roch *et al.*, 1986 and Popko *et al.*, 1987) but the size of the shi^{mld} MBP mRNA was normal, 2.2 - 2.4 kb (Roch *et al.*, 1986 and Popko *et al.*, 1987).

It has been established that the *shiverer* and shi^{mld} mutations are within the structural gene for the myelin basic protein. The alleles map to the distal end of mouse chromosome 18, which contains the *MBP* gene (Doolittle and Schweikart, 1977; Roach *et al.*, 1985 and Sidman *et al.*, 1985). When the myelin basic protein gene in these mutants was analysed, clear alterations in the gene were identified. The *shiverer* mutation (Figure 5) results in a deletion of exons 3 to 7 of the *MBP* gene (Roach *et al.*, 1983;

Roach *et al.*, 1985; Kimura *et al.*, 1985 and Molineaux *et al.*, 1986), and the production of low levels of MBP-related RNA. This MBP-related mRNA is aberrantly spliced and poorly polyadenylated. In contrast, in shi^{mld} mice, there appears to be a tandem duplication of exons 3 to 7 with the more upstream set of exons inverted (Figure 5). There is one complete *MBP* gene, and at least one extra-partial or complete gene that is closely juxtaposed to the complete *MBP* gene (Popko *et al.*, 1987, Popko *et al.*, 1988). Thus, it would appear that the partial or complete duplication of the *MBP* gene near a normal gene can reduce the expression of the normal gene to 2% of normal and alter the normal developmental expression of the gene.

Readhead *et al.*, (1987) produced transgenic *shiverer* mice carrying one or two copies of the MBP transgene. The transgene is expressed in a tissue-specific manner with the correct developmental pattern. *Shiverer* mice carrying the MBP transgene produce more MBP mRNA and protein than the original *shiverer* mice, and they live longer, with a less severe disorder. However, *shiverer* mice carrying a single transgene produced approximately 12.5% of normal MBP mRNA and 8.5% of normal protein, whereas mice homozygous for the transgene produced roughly twice as much mRNA as protein. The heterozygous transgenic mice survived longer than the original *shiverer* mice but they still had convulsions and died at approximately 6 months. Homozygous transgenic mice appeared to be phenotypically closer to normal, for example, they had no tremors or convulsions, but some subtle behavioural abnormalities were observed. Despite the presence of only 25% of normal levels of MBP mRNA in the homozygous transgenic mice, the

myelination of the optic nerve, spinal cord and cerebellar white matter was significantly increased relative to the heterozygous transgenic mice.

These studies with transgenic mice suggest that insertion of a normal MBP gene into the genome of mice carrying either a deletion or a duplication of the normal gene can produce a tissue-specific and developmentally accurate expression of the transgene. However, the level of expression was not normal, since even with two copies of the transgene, only 25% of normal MBP mRNA was observed. The fact that the transgene was expressed at comparable levels in *shiverer* and *shi^{mld}* mice suggests that the deleterious effect of the MBP gene duplication on MBP gene expression in *shi^{mld}* mice is *cis*-acting, not *trans*-acting. Clearly the production of these transgenic mice, and the future use of these mice will provide new insight into myelin gene expression.

Golli-mbp

S1   S2   S3

1   2   3 4   5a b c   6

MBP   0   1a b c   2 ————————————
                              deletion   *Shi*

Golli-mbp

S1   S2   S3   inversion

1   2   3 4   5a b c   6   11   10   9a b   8   7   7   8   9a b   10   11

MBP   0   1a b c   2   7   6   5a b   4   3   3   4   5a b   6   7 *mld*

————————————
inversion
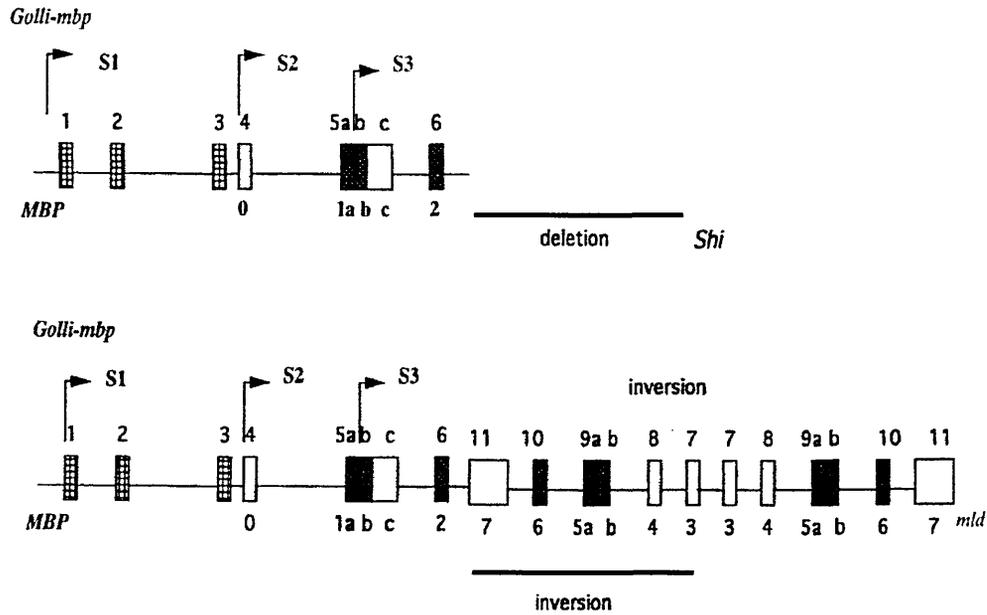
Figure 5. Diagrammatic representation of *golli-mbp* gene structure in dysmyelinating mutant mice.

Shi, shiverer mouse where there is a deletion of MBP exons 3 to 7 and *mld*, myelin deficient mouse, where MBP exons 3 to 7 are repeated with an inversion in the upstream gene, shown as exons 7 to 3.

(After Asipu and Blair (1997) Regulation of transcription and translation in the central nervous system – from Molecular Biology of Multiple Sclerosis.)

41

1.7.2. *Jimpy (jp)* and Myelin Synthesis Deficient (*jp^{msd}*)

The *Jimpy* mutation (*jp*) and its allele Myelin Synthesis Deficient (*jp^{msd}*) are recessive X-linked mutations. This is exclusively a central nervous system disorder with no PNS involvement. These mice exhibit tremors, which become progressively worse, beginning at postnatal day 11. Convulsions appear in older animals and these animals do not survive past 25 – 30 days. The MBP proteins are significantly reduced in *jimpy* mice, relative to normal animals (Delassale *et al.*, 1981, Kerner and Carson, 1984). Delassale *et al.* (1981), found that the deficit in MBP in *jimpy* mice ranged from 94 - 98% in different regions of the *jimpy* brain at 25 days of age. When the distribution of the four major MBPs was studied, it appeared that in young *jimpy* mice, the deficit of the 21.5 kDa MBP was less than the deficit in the other three MBPs. The overall deficit in MBP accumulation in animals in this study was 92% at all ages (Kerner and Carson, 1984). When MBP accumulation was compared in *jimpy* optic nerve and sciatic nerve, a significant reduction in MBP content was observed throughout development in the optic nerve, whereas no reduction was apparent in sciatic nerve (Jacque *et al.*, 1983). Thus, there are differences in the regulation of the *MBP* gene in *jimpy* oligodendrocytes and Schwann cells. The level of MBP mRNAs was also measured in *jimpy* mice but with contradictory results Carnow *et al.*, 1984).

In vitro translation studies by Campagnoni *et al.* (1984) indicated a reduction in the level of all MBP mRNAs. The mRNAs for the 17, 18.5, and 21.5 kDa proteins were approximately 43 to 48 % of normal, and the 14 kDa MBP mRNA was approximately 7% of normal. When the level of total MBP

42

mRNA was analysed by dot blots, it ranged from 19 to 42 % of normal age-matched controls during development (Roth et al., 1985), and these reductions were observed for both polysomal and nuclear RNA (Sorg et al., 1987). Despite differences in MBP mRNA levels measured in these studies, the level of detectable protein was significantly below the level of detectable mRNA, suggesting that whereas the MBPs may be synthesised to some extent, the amount that accumulates is quite low. This would suggest that the MBP that is not inserted into an appropriate membrane is "turned over". This would be consistent with the conclusions drawn from experiments on the quaking mutant where the synthesis of MBP is significantly above the level of accumulation.

1.8. Quaking (qk)

Quaking viable (qk^v) is an autosomal recessive mutation resulting in dysmyelination in the CNS and PNS (Hardy et al., 1998). The quaking^{viable} mutant mouse (qk) has a deletion on chromosome 17 (Ebersole et al., 1992). A candidate gene, qkI, has recently been identified and evidence suggesting abnormalities leading to diminished expression of the selective RNA-binding protein QKI in myelin producing cells are involved in the quaking phenotype (Ebersole et al., 1996, Hardy et al., 1996). QKI contains amino acid domains characteristic of RNA-binding and interaction with Src homology 3 (SH3)-containing signalling molecules, and therefore belong to a fast growing family denoted as signal transduction activators of RNA (STAR) (Vernet and

Artzt, 1997). Isoforms of the QKI proteins are expressed in myelin-forming cells, in other CNS glia, as well as in non-nervous tissue; it is therefore suggested that they are involved in RNA splicing or metabolism (Ebersole *et al.*, 1996, Kondo *et al.*, 1999, Wolf et al., 1999). The proteins might play roles not only in myelination but also in embryonic development (Ebersole *et al.*, 1996). These mice exhibit tremors, which continue throughout their lifetime beginning at approximately P12. Tonic seizures occur in these animals, although they have a normal lifespan. The morphologic phenotype includes thin myelin which often fails to compact, and is thrown into exuberant loops and whorls. Studies on the accumulation of MBP in different regions of the *quaking* nervous system showed that the developmental curve was different between *quaking* and normal mice and that there was a delay in the appearance of MBP in every CNS region tested. Furthermore, the level of MBP was severely depressed, ranging from 5 – 20 % of normal (Delassale *et al.*, 1981 and Jacque *et al.*, 1983). Interestingly the accumulation of MBP in spinal cord myelin did not stop at any particular time during development, as it normally does (Delassale *et al.*, 1981). Therefore at day 80 MBP content in the spinal cord was approximately 25 % of normal, whereas at 1 year it had risen to 35 % of normal. An early study on MBP biosynthesis in quaking mice suggested the synthesis of the MBPs occurred at the normal rate, but it was the incorporation of the myelin into the membrane that was deficient (Brostoff *et al.*, 1977). Both the stability and localisation of MBP mRNA are potentially coupled with its translation status (Ueno *et al.*, 1994a, 1994b, Ainger *et al.*, 1997) and QKI has recently been reported to act as a translation suppressor (Saccomanno et al., 1999),

44

therefore, interactions of MBP mRNAs with QKI may influence MBP

expression at multiple posttranscriptional levels, including mRNA turnover,

translation and subcellular localisation (Li et al., 2000).

1.9. **TAIEP mutant rat**

The name of this mutant is derived from the clinical signs observed in

affected animals, e.g. at 3 weeks old taiep rats show a tremor (t), especially

in the hind quarters, ataxia (a) is seen at 3 to 4 months of age, followed by

immobility (i), epilepsy (e) and a gradually worsening hind limb paralysis (p)

in animals over 6 months of age. Studies have confirmed an autosomal

recessive mode of inheritance (Duncan et al., 1992). All the myelin proteins

including MBP, PLP, MAG, and CNP are present in the mutant CNS albeit in

reduced levels. This agrees with the finding that taiep myelin sheaths show

positive immunostaining for both MBP and PLP (Duncan et al., 1992). The

microtubule accumulations in the taiep oligodendrocytes are a striking and

unique finding focussing attention on the role of the oligodendrocyte

cytoskeleton in the functioning of this cell. The cytoskeleton of the

oligodendrocyte is likely to influence the transport of both myelin proteins

and their message (Ainger et al., 1993). A close association between MBP

and CNP and the cytoskeleton has been observed in the oligodendrocyte,

and it would appear that MBP colocalises with tubilin, whereas CNP

associates with microfilaments (Wilson and Brophy, 1989). It has also been

known for some time that the myelin protein mRNAs are spatially segregated

within the oligodendrocyte (Trapp, 1990). The messages for PLP and MAG are found in the perinuclear region, whereas the MBP message is more diffusely distributed throughout the cytoplasm. The role of the cytoskeleton in the transport of MBP mRNA was demonstrated by Ainger et al., (1993) using digoxigenin-labelled MBP mRNA microinjected into cultured oligodendrocytes. The mRNA formed granules that were transported from the cell body along the oligodendrocyte processes. It would not be unreasonable to suggest this transport would be disrupted by the massive accumulations of microtubules in the taiep oligodendrocytes. As a result this could lead to the inefficient incorporation of myelin components into the myelin sheath with the resulting inadequate myelination of the developing CNS, and failure of myelin maintenance in the longer term (Lunn *et al.*, 1995).

1.10. Myelin basic protein is a zinc-binding protein in brain

Zinc is an abundant transition element in the brain that has been implicated in a variety of metabolic processes. In humans, zinc deficiency has been known to cause many neurological and neuropsychiatric disorders (Frederickson, 1989). In the brain zinc, like calcium, is another important divalent cation in the CNS, is unevenly distributed with the highest level in the hippocampus. As zinc in nervous tissue rarely exists as a free ion but usually associated with proteins, it is not known whether this uneven

distribution in different regions of the brain is related to different concentrations of zinc-binding proteins present (Tsang *et al.*, 1997).

Tsang *et al.*, (1997) showed that three major zinc-binding proteins (ZnBPs) of molecular weight 53, 42 and 21 kDa were present in porcine brain. The 53 and 42 kDa zinc-binding proteins were present in all subcellular fractions whereas the 21 kDa protein was mainly found in the particulate fractions. It was suggested that of the three major zinc-binding proteins in the brain only the 21 kDa protein appears to be brain-specific and was later identified as myelin basic protein (Tsang *et al.*, 1997). Since no cysteine residues are present in mammalian MBPs the probable zinc binding sites in myelin basic protein may consist of imidazole or carboxylate groups (Riccio et al., 1995). These groups are fairly abundant in MBP, about 6% each of aspartic acid, glutamic acid and histidine. Although the structural motif responsible for zinc-binding in MBP is still unknown, studies on other ZnBPs have revealed a common structural motif of His-$X_{2,3}$-His (Berg, 1986). From the amino acid sequence of porcine MBP, a structural motif of His-Ala-Arg-His has been found between amino acid residues 23-26, which may be responsible for zinc binding in MBP (Tsang *et al.*, 1997). The association of zinc with myelin and the fact that myelin basic protein is a zinc binding protein show that zinc plays a structural and/or functional role in myelin. Myelin basic protein is located on the cytoplasmic side of the multilamellar membrane at the major dense line and is responsible for the compaction of the two opposing membrane bilayers. The fact that MBP can bind zinc implies that zinc may be involved in myelin compaction by

47

interacting with MBP. There are developmental studies that showed that maternal zinc deficiency caused poor myelination in foetuses and decreased amounts of MBP associated with foetal myelin membranes (Liu *et al.*, 1992). Furthermore, recent fluorescence studies have confirmed the role of zinc ions in the formation of myelin basic protein aggregates (Cavatorta *et al.*, 1994). In myelin compaction, bimolecular interactions, MBP-membrane, as well as MBP-MBP interactions are involved in multilamellar myelin formation. Therefore, the effect of zinc on myelin compaction is probably by promoting or stabilising the MBP-membrane and/or MBP-MBP interaction (Berlet *et al.*, 1994). As zinc has been shown to stabilise the binding of macromolecules to plasma membranes in peripheral tissues, and that zinc has been shown to prevent the dissociation of MBP from membrane, it is not surprising that zinc is involved in the stabilisation of MBP in myelin sheath formation (Berlet *et al.*, 1994).

1.11. Aims of the project

The MBP unit is expressed in oligodendrocytes (Central Nervous System) and Schwann cells (Peripheral Nervous System) and is regulated at the level of transcription by proteins binding in a specific manner to DNA sequences located within its promoter region. The identification of these proteins and DNA sequences is essential for understanding the mechanism(s) that regulate the transcription of the *MBP* gene.

The major aims of this project were to clone and sequence the promoter region of the mouse myelin basic protein (*MBP*) gene and to identify a transcriptional activator(s) (transcription factor(s)) which interact with an oligodendrocyte-specific element(s) in the myelin basic protein gene promoter. To achieve this the following steps had to be taken:

- cloning and sequencing of the *Hind III – Sac I* fragment (from the plasmid pEX1). This fragment is located between nucleotides – 1319 and + 227 relative to the transcription start site and comprises the promoter region and exon 1 of the mouse myelin basic protein gene.

- analysis of the interaction of oligodendrocyte specific protein factor(s) with the region of the MBP promoter between nucleotides – 609 to – 577. This region was selected for further study because work published by other groups suggested that it contains an oligodendrocyte specific transcription activator (Asipu and Blair 1994). To achieve this electrophoretic mobility shift assay (EMSA) and competition binding with oligonucleotides corresponding to this region were performed.

49

- identification using the yeast one – hybrid system cDNAs encoding proteins, which bind to the –609 to –577 region of the mouse myelin basic protein promoter. The one-hybrid system is a functional *in vivo* genetic assay used for isolating and studying novel genes encoding proteins which bind to target DNA. In this system a reporter construct containing at least three tandemly repeated copies of a putative binding sequence are inserted upstream of a *HIS3* gene minimal promoter in a specially designed plasmid. This reporter construct is integrated by homologous recombination into a mutated *HIS3* gene in the yeast host genome to create a yeast reporter strain carrying a normal copy of the *HIS3* gene. A library of cDNAs encoding GAL4 AD fusions is expressed in the reporter strain. If an AD/library protein interacts with the target DNA sequence, *HIS3* promoter activity is increased resulting in increased amounts of histidine produced. Therefore, clones expressing histidine at a level higher than the reporter strain are selected by growth on medium lacking histidine in the presence of 3-aminotriazole (3-AT) for further study.

- sequencing of cDNAs encoding proteins, which bind to the –609 to –577 region of the myelin basic protein promoter.

- search of nucleotide databases (e.g. EMBL and Genbank) and protein databases like Swissprot or Swall for homology with other transcription factors using the FastA and Blast sequence alignment programs.

Chapter 2

MATERIALS AND METHODS

2.1. MATERIALS AND SUPPLIERS

2.1.1. BACTERIAL STRAINS

Escherichia coli strain HB101 (containing plasmid pEX1)

Donated by Arthur Roach, Division of Biology, California

Institute of Technology, Pasadena, California.

Escherichia coli strain (DH5α).

2.1.2. CLONING VECTORS

Bluescript SK$^+$	Promega
pHISi – 1	Clontech

2.1.3. ENZYMES

Hind III	20 U/μl	NEB (104 S)
Sac I	20 U/μl	NEB (156 S)
EcoR 1	20 U/μl	NEB (101 XL)
Mlu 1	10 U/μl	NEB (198 S)
Xho 1	20 U/μl	NEB (146 S)

51

T4-DNA Ligase (E.C.6.5.1.1.)   400 U/μl   NEB (202 S)

Alkaline. phosphatase (E.C.3.1.3.1.)   10 U/μl NEB (290 S)

Polynucleotide kinase   10 U/μl   NEB (201 S)

Lysozyme   (E.C.3.2.1.17.)   SIGMA (L 6876)

2.1.4. OLIGONUCLEOTIDES

The oligonucleotides used for chain termination sequencing,

electrophoretic mobility shift assays, PCR amplification were synthesised on

a Pharmacia Gene Assembler Plus using the phosphoramidite chemistry.

2.1.5. KITS

Mini-Prep Kit Plus   Pharmacia Ltd.

T7 Sequencing Kit   Pharmacia Ltd.

HYDRO-LINK™ DNA Sequencing System   A.T. Biochem.

Sequenase ™ DNA Sequencing Kit   United States Biochemical

                                            Corporation.

2.1.6. RADIOCHEMICALS

Deoxyadenosine triphosphate α-thiol (^{35}S)   500 Ci mmol^{-1} (10 mCi ml^{-1})

                              NEN (DUPONT) Hertfordshire.

Adenosine 5' triphosphate, tetra (triethylammonium salt) (γ - ^{32}P)

>4000 Ci mmol^{-1} (10 mCi ml^{-1})   ICN Biomedicals Ltd Buckinghamshire.

2.2. LARGE SCALE ISOLATION OF PLASMID DNA BY A MODIFIED CLEWELL AND HELINSKI TECHNIQUE

An overnight culture (4 ml) of *Escherichia coli* harbouring the pEX1 plasmid was inoculated into 400 ml of Luria broth (see Appendix II), and incubated at 37° C with vigorous shaking until the absorbance at 600 nm reached 1.00. Chloramphenicol (4ml of stock solution at a concentration of 10 mg/ml in 50% ethanol) was added and the incubation continued for a further 18 h.

Cells were harvested by centrifugation at 4 krpm for 20 min at 4° C using the Sorvall GS3 rotor, and the cell pellet was resuspended by gentle agitation in 3.0 ml of ice-cold Buffer I (see Appendix II). After transferring the bacterial suspension to a suitable centrifuge tube 500 µl of a freshly prepared ice-cold lysozyme solution (at a concentration of 10 mg/ml in buffer I) were added. Following incubation at 0° C for 4-5 min (for chloramphenicol amplified cultures) 1 ml of ice-cold 0.25 M EDTA pH 8.0 was added and the 4-5 min incubation at 0° C repeated.

Finally 4.0 ml of Triton X-100 lysis solution (see Appendix II) were added and the lysis of the cells was monitored by observing the viscosity of the mixture. Separation of the chromosomal DNA and cell debris from the plasmid DNA was achieved by centrifugation at 18 krpm for 1 h at 4° C using the SL50T rotor.

The plasmid DNA was stored at 4° C before purification by centrifugation to equilibrium on caesium-chloride density gradients.

2.3.   PURIFICATION OF PLASMID DNA BY CENTRIFUGATION TO EQUILIBRIUM IN CAESIUM CHLORIDE-ETHIDIUM BROMIDE GRADIENT

The volume of cleared lysate prepared as described in section 2.2. was accurately measured.

For every 10 ml of lysate, exactly 9.62 g of solid caesium chloride and 625 μl of a solution of ethidium-bromide (stock concentration 10 mg/ml in water) were added, and mixed gently by inversion until all the caesium chloride had dissolved, (the final density of the solution being 1.55 g η = 1.3680 and the final concentration of ethidium bromide approximately 600 μg/ml). It should be noted that in this procedure the furry purple aggregates that float to the top of the solution are complexes formed between the ethidium-bromide and bacterial proteins.

Centrifugation was carried out at 40 krpm for 40 h at $18° \pm 1°$ C using a Sorvall T865 rotor. Two bands of DNA were evident in both visible light and UV light, the upper band consisting of chromosomal DNA and the lower band of plasmid DNA. The lower band of DNA was collected with a hypodermic needle, and transferred into a fresh tube.

Ethidium bromide was removed by extraction with caesium chloride saturated isopropanol (discarding the top layer each time and re-extracting until the bottom layer was colourless, usually six times).

Following dialysis against three changes of TE buffer pH 8.0 (see Appendix II), the solution of DNA was extracted three times with phenol (transferring the top layer to a fresh tube each time).

The DNA was precipitated with 0.1 volume 3 M sodium acetate, 2.5 volumes of ethanol at -70° C for 15 min and pelleted by centrifugation using a Sorvall ST-micro rotor. The pellet was dried *in vacuo* and resuspended in 400 µl of sterile water. The concentration of the DNA was determined by measuring the absorbance at 260 nm (1 A_{260} unit of double stranded DNA = 50 µg/ml).

2.4. ANALYTICAL AGAROSE GEL ELECTROPHORESIS OF DNA

Gels were prepared by adding 300 mg of agarose to 30 ml of 1X Tris-acetate buffer (1% w/v) (see Appendix II) and heating in a microwave oven until all the agarose had dissolved. After cooling to 50° C in a water-bath, 3 µl of ethidium bromide were added to give a final concentration of 0.5 µg/ml ethidium bromide (stock Et-Br 10 mg/ml stored at 4° C). The agarose slurry was then poured into a pre-levelled gel former and allowed to set at room temperature for 30 minutes.

Tris-acetate buffer (1X) (see Appendix II) was added to the electrophoresis chamber so that the gel was covered to a depth of 1 mm.

Samples (5 µl in loading buffer – see Appendix II) were loaded below the surface of the running buffer into the pre-formed sample wells. The gel was electrophoresed at a constant voltage (100 V), until the bromophenol blue in the loading buffer was 1 cm from the anode end of the gel.

After electrophoresis the gel was analysed on a UV transilluminator.

2.5. ANALYTICAL SCALE DIGESTION OF DNA WITH RESTRICTION ENDONUCLEASES

Plasmid DNA (2 µg in 5 µl) was mixed in a sterile Eppendorf tube with 1 µl of the appropriate 10X restriction buffer (supplied with each enzyme), and 3 µl of sterile deionised water.

For single restriction enzyme digests, 1 µl of enzyme (20 U/µl) was added and the digest incubated at 37° C for 1 h after mixing by gentle flicking.

For double endonuclease digestions (where the buffer conditions are identical), only 5 µl of sterile water were added followed by 1 µl (20 U/µl) of the second enzyme, the contents mixed and incubated as above.

The digestion was terminated by the addition of 2 µl of loading buffer (see Appendix II) and the products analysed by agarose gel electrophoresis.

2.6. CLONING OF THE *Hind III – Sac I* FRAGMENT OF pEX1 IN pBLUESCRIPT PLASMID

2.6.1. PREPARATION OF THE *Hind III - Sac I* FRAGMENT OF pEX1 PLASMID FOR CLONING

2.6.1.1. LARGE-SCALE DIGESTION OF THE PLASMID pEX1

Digestion with restriction endonucleases was performed essentially as described in 2.5, with the minor modifications indicated below.

Plasmid pEX1 (37.5µg in 95 µl) was mixed with 35 µl of the

appropriate 10X restriction buffer, 180 µl of sterile distilled water and 20 µl

(20 U/µl) of *Hind III* and 20 µl (20 U/µl) of *Sac I*. After an incubation for 2h at

37° C the progress of the digestion was monitored by analytical scale

agarose gel electrophoresis.

Preparative gel electrophoresis was performed after 3h 45 min of

digestion.

2.6.1.2. ELECTRO-ELUTION OF THE *Hind III - Sac I* FRAGMENT

The band containing the *Hind III-Sac I* fragment was excised from the

preparative agarose gel and fastened securely in "pre-wetted" dialysis tubing

for electro-elution at a constant voltage (100 V). The progress of the elution

was monitored every 10 min on the UV transilluminator. When the ethidium

bromide had migrated out of the gel slices the elution of DNA was ended and

the samples stored at -20° C for subsequent purification and concentration.

2.6.1.3 PURIFICATION AND CONCENTRATION OF THE *Hind III - Sac I* FRAGMENT

An ion-exchange column was plugged with sterile siliconized glass

wool and filled with approximately 2 ml of DE-52 cellulose matrix. The

solution containing the *Hind III-Sac I* fragment (produced as described in

section 2.6.1.2.) was loaded onto the column, the flow-through collected, and passed through the column a second time.

After washing the column with "low salt" solution (0.3 M NaCl), the DNA fragment was eluted in a small volume (1000 μl) of "high salt" solution (1.0 M NaCl), and collected in a Corex tube. Nucleic acid precipitation was carried out by the addition of 0.1 volume of 3 M sodium acetate and 2 volumes of ice-cold ethanol followed by incubation at $-20°$ C for 24h.

The DNA was pelleted by centrifugation at 10 krpm for 1h at $2°$ C using a Sorvall ST-micro rotor. After decanting the supernatant any remaining ethanol was evaporated under vacuum and the pellet dried.

Finally the pellet of DNA was resuspended in 170 μl of sterile distilled water and stored at $-20°$ C.

2.6.2. PREPARATION OF THE PLASMID pBLUESCRIPT FOR CLONING

2.6.2.1. DIGESTION OF pBLUESCRIPT WITH RESTRICTION ENDONUCLEASES *Hind III* and *Sac I*

The vector DNA (pBluescript) was cleaved with *Sac I and Hind III* using a similar protocol to that outlined in section **2.5.**, and the restriction enzymes inactivated by heating at $60°$ C for 10 min. The completeness of the digestion was established by running a 2 μl sample on an analytical scale agarose gel.

2.6.2.2. PHOSPHATASE TREATMENT OF pBLUESCRIPT

Digested plasmid pBluescript (20 µg in 20 µl) was treated with alkaline phosphatase to remove 5' phosphate groups. The reaction mixture consisted of 20 µl of cleaved vector, 2.5 µl of 1 M Tris pH 8.0 (to give a final concentration of 50 mM), 21.5 µl of sterile distilled water and 6 µl of alkaline phosphatase (10 U/µl). After mixing the tubes were incubated at 37° C for 45 min.

The alkaline phosphatase was inactivated by adding an equal volume (50 µl) of water-saturated phenol, (discarding the lower layer and repeating usually twice). Water-saturated diethyl ether was added, and after removing the organic phase (upper layer) the procedure was repeated a further six times.

Following the removal of as much of the ether as possible 5 µl of 3M sodium acetate and 150 µl of ethanol were added. The DNA solution was then incubated at -20° C for 72 h (approx) to precipitate the DNA. The dephosphorylated *Hind III-Sac I* cleaved vector DNA was pelleted by microcentrifugation at 13 krpm for 20 min at 4° C, the supernatant removed and the pellet dried *in vacuo*.

Sterile deionised water (40 µl) was added to reconstitute the pellet, and the recovery of the dephosphorylated vector was checked by agarose gel electrophoresis.

2.6.3. LIGATION

Ligation was performed as follows:

2 μl of dephosphorylated vector (0.5 μg/μl) and 5 μl (2 μg) of the

Hind III-Sac I fragment isolated from pEX1 were mixed in a sterile Eppendorf

tube. This was followed by the addition of 1 μl 10X ligation buffer, 1 μl ATP

(10 mM) and 0.5 μl (200 U) T4-DNA ligase. After mixing the contents by

gentle flicking an overnight incubation at 16° C was carried out.

2.6.4 INTRODUCTION OF PLASMID DNA INTO *E. coli* CELLS

2.6.4.1. PREPARATION OF COMPETENT CELLS

E.coli cells strain DH5α (5 μl) were spread over the surface of an M9

minimal medium plate (see Appendix II) and incubated overnight at 37° C.

If 5 μl proved too large an inoculum to give individual colonies the

procedure was repeated with a 1 μl inoculum. A single colony was then

inoculated into a tube containing 15 ml of L-broth and the culture incubated

overnight at 37° C with vigorous shaking.

Overnight starter culture (1 ml) was inoculated into L-broth (100 ml)

and incubated at 37° C with vigorous shaking to an absorbance at 650 nm of

0.5. Then 50 ml of the culture was centrifuged at 3 krpm for 5 min at 4° C

using the Sorvall SL50T rotor. Following aspiration of the supernatant the

cells were carefully resuspended in 7.5 ml of ice-cold 0.1 M $CaCl_2$ (made

from filter sterilized 2 M stock) and incubated on ice for 15 min before

centrifugation at 3 krpm for 5 min at 4° C using the Sorvall SL50T rotor. The cells were resuspended in 1.5 ml (0.1 volume) of ice-cold 0.1 M CaCl₂, and kept on ice until required.

2.6.4.2. TRANSFORMATION

Three transformation mixtures were prepared as follows:

TUBE 1 UNCUT VECTOR: 10 μl of uncut vector (0.1 μg/μl) was added to 20 μl of 1X SSC. This was followed by 200 μl of competent cells prepared as described in section 2.6.4.1. and the contents of the tube mixed well.

TUBE 2 VECTOR AND FRAGMENT: 10 μl of ligation mixture (dephosphorylated vector + pEX1 fragment) were added to 20 μl of 1X SSC. This was then followed by 200 μl of competent cells and the contents mixed well.

TUBE 3 LINEAR VECTOR: 10 μl of dephosphorylated vector (0.1 μg/μl) and 20 μl of 1X SSC were added to 200 μl of competent cells and the contents of the tube mixed well.

After incubation of all three tubes on ice for 30 min, with occasional shaking, a heat-shock at 42° C for 2 min was performed. This was followed by incubation on ice for a further 15 min.

L-broth (2ml) was added to each sample and incubated at 37° C for

1 - 1.5 h with continuous shaking to allow gene expression. The cells were

then pelleted by centrifugation at 3 krpm for 5 min at room temperature and

each pellet was resuspended in a final volume of 100 μl of L-broth.

2.6.4.3. PLATING-OUT OF THE TRANSFORMANTS

L-broth (80 μl) was pipetted into the centre of pre-dried AIX plate (see

Appendix II) followed by 20 μl of transformed cells. The cells were spread

over the whole surface of the plate using an "L-shaped" glass rod, kept on

the bench for a few minutes to let the transformed cells absorb into the agar,

and then incubated overnight at 37° C.

2.6.5. ANALYSIS OF THE TRANSFORMANTS

Transformants were analysed for the presence of the recombinant

plasmid by the method of Holmes and Quigley (1981) followed by agarose

gel electrophoresis. Restriction enzyme digestion (*Sac I – Hind III*) was used

to excise the cloned insert.

2.6.5.1. PREPARATION OF PLASMID DNA ON A SMALL-SCALE (HOLMES AND QUIGLEY (1981) METHOD)

White colonies from the vector plus DNA fragment plate were

resuspended in L-broth containing ampicillin at a final concentration of

50 μg/ml and incubated at 37° C overnight with constant shaking. Aliquots of each culture (2 x 1.5 ml) were pipetted into sterile Eppendorf tubes and the cells pelleted by centrifugation in a microfuge for 30 sec. Each pellet was resuspended by vortex mixing in 75 μl of STET buffer (see Appendix II) and the suspensions combined into one 150 μl suspension in one tube. Freshly prepared lysozyme (12 μl at a concentration of 20 mg/ml) was added to each tube and the contents mixed by vortexing.

Then the cell lysates were heated in a boiling water bath (40 sec) and immediately centrifuged in a microfuge for 15 min at room temperature. The supernatant (approximately 100 μl) was carefully removed from the fluffy pellet containing chromosomal DNA, and transferred to a fresh sterile tube. An equal volume of isopropanol (100 μl) was added to the supernatant and the nucleic acids precipitated in an alcohol dry-ice bath for 10 min.

The nucleic acids were pelleted by centrifugation in a microfuge at 12 to 14 krpm for 15 min in the cold-room. Following the careful removal of the supernatant with a sterile Pasteur pipette, the pellet was resuspended in 100 μl of sterile distilled water and the nucleic acids precipitated with 10 μl of 3M sodium acetate (pH 6.0) and 300 μl of absolute ethanol (stored at -20° C) in an alcohol dry-ice bath for 10 min.

The nucleic acid was pelleted by centrifugation in a microfuge at full speed, the supernatant discarded and the pellet washed carefully with 100 μl of 70 % ethanol (stored at -20° C). The nucleic acid pellets were dried under vacuum for 3 to 4 min, resuspended in 30 μl of sterile distilled water and stored at -20° C.

2.6.5.2. CONFIRMATION OF THE PRESENCE OF THE *Hind III-Sac I* FRAGMENT IN THE RECOMBINANT PLASMIDS

Plasmid DNA prepared according to the protocol described in section 2.6.5.1. was digested with a combination of *Hind III and Sac I* restriction endonucleases as described in section 2.5.

Agarose gel electrophoresis was used to analyse the products of the endonuclease digestion as outlined in section 2.4.

3.0. ISOLATION AND PREPARATION OF PLASMID DNA FOR SEQUENCING

3.1.1. ALKALINE LYSIS MINI-PREPARATION USING THE PHARMACIA KIT

SB medium (1.4 ml) (see Appendix II) containing the appropriate antibiotic was inoculated with a single colony of the recombinant clone of interest and incubated overnight at 37° C with constant shaking. The cells were pelleted by centrifugation in a microfuge for 1 min at full speed, and the medium removed by aspiration to leave the cells as dry as possible.

Solution I (100 μl) (see Appendix II) was added and the cells resuspended by vortex mixing. Following incubation on ice for 5 min, 200 μl of Solution II (see Appendix II) were added and the incubation on ice repeated. After the further addition of 150 μl of Solution III (see Appendix II) and incubation on ice for 5 min the cell lysates were microfuged (Sorvall T21 with ST Micro rotor) at full speed for 5 min at room temperature. The

supernatant was transferred to a fresh microfuge tube and 1 volume (450 µl) of isopropanol was added, mixed by inversion and incubated at room temperature for 5-10 min.

The DNA was pelleted in a microfuge at full speed at room temperature for 5 min. Following the removal of the supernatant by aspiration the DNA pellet was gently washed in 250 µl of isopropanol, and air-dried.

Since the DNA was to be used for sequencing by the dideoxy chain termination method the pellet was redissolved in 25 µl of Column Buffer (see Appendix II).

3.1.2. ESTIMATION OF THE CONCENTRATION OF DNA BEFORE THE PREPARATION OF SINGLE STRANDS

The approximate concentration of plasmid DNA was determined by comparing the fluorescence of ethidium bromide-stained DNA with that of known amounts of caesium chloride purified DNA. An agarose gel (0.8%) was prepared in 1X Tris-acetate buffer (see Appendix II), containing ethidium bromide at a final concentration of 0.5 µg/ml.

Redissolved plasmid DNA (1 µl) prepared as described in section 3.1.1. was combined with 4 µl of sample buffer. Aliquots (2, 4 and 8 µl) of the pBR322 DNA standard containing 2,4, and 8 µg of DNA respectively were loaded into three wells of the agarose gel; duplicate samples of the plasmid DNA and a molecular size marker were loaded into adjacent wells.

The DNA samples were electrophoresed and the amount of DNA in the plasmid sample was estimated by comparing the intensity of its fluorescence under ultra-violet light with that of the three pBR322 DNA samples. The size of the plasmid was estimated from the molecular size marker.

3.1.3. PREPARATION OF SINGLE STRANDED DNA FOR SEQUENCING

Plasmid DNA (20 µl at a concentration of 1.5 µg/µl) prepared as described in section 3.1.1. was treated with 5 µl of 2 M NaOH at room temperature for 5 min. The denatured DNA was purified immediately on Sephacryl S-400 columns (prepared as described in Appendix III).

The plasmid solution was slowly applied to the centre of the upper flat surface of the compacted bed of Sephacryl S-400. In this procedure careful application of the sample to the centre of the bed was essential for good separation of the DNA. The column was then centrifuged for 2 min at approximately 400 x g in a swing out bucket rotor. The effluent collected in a microfuge tube contained the purified single stranded plasmid DNA in approximately 25 µl of column buffer.

4.0. SEQUENCING OF DNA

4.1. PREPARATION OF OLIGONUCLEOTIDE PRIMERS

Oligonucleotide primers were synthesised as described in Appendix IV using the phosphoramidite chemistry and prepared by the procedure described below. This procedure gives high yields of de-salted oligonucleotides which are 20 bases or longer.

The ammonia solution (1 ml) containing the oligonucleotides was divided into 3 X 330 µl aliquots in flip-top Eppendorf tubes. The oligonucleotides were precipitated by adding 33 µl of 3M sodium acetate solution pH 6.0, followed by 915 µl of absolute ethanol (maintained at -20° C) to each tube. After overnight incubation at -20° C the tubes were microfuged at 13 krpm for 15 min at 4° C. The ethanol was removed by aspiration and the pellets resuspended in 500 µl of 80% ethanol (stored at -20° C).

After pooling the contents of the three tubes into one conical Eppendorf tube and microfuging at 13 krpm for 5 mins at 4° C the ethanol was removed and the open tube covered with parafilm. The contents of the tube were lyophilised on a freeze-dryer once the parafilm had been perforated with a hypodermic needle.

The oligonucleotide primers could be stored indefinitely in this lyophilised form, or could be resuspended in 600 µl of buffer (for example T.E. pH 7.5 see Appendix II).

4.2. DNA SEQUENCING USING THE SEQUENASE KIT ™

All sequencing reactions were carried out according to the manufacturers instructions in small plastic centrifuge tubes typically 1.5 ml Eppendorfs; these were kept capped between manipulations to minimize evaporation of the small volumes used.

4.2.1. ANNEALING THE TEMPLATE AND THE PRIMER

For each of four sequencing lanes, a single annealing and labelling reaction was used.

Primer (1 μl at a concentration of 0.5 pmol/μl), reaction buffer (2 μl) and 7 μl (2.0 μg) of template DNA were combined in a centrifuge tube. For the control, 5 μl (0.2μg) of DNA supplied in the kit and 2 μl of water were used, this was equivalent to a 1:1 (primer : template) molar stoichiometry.

The capped tubes were held at 65° C for 2 min and then allowed to cool slowly to room temperature over a period of about 30 min. Once the temperature was below 30° C annealing was complete and the tube kept on ice until required.

4.2.2. THE LABELLING REACTION

For reading sequences up to 500 bases from the primer the labelling mix (dGTP) was diluted 5-fold with sterile deionised water. The Sequenase Version 2.0 enzyme (13 U/μl) was diluted 1:8 in ice-cold enzyme dilution buffer and then stored on ice until required. To the annealed template-primer (10 μl) the following additions were made on ice, 1 μl DTT (0.1 M), 2 μl diluted labelling mix, 0.5 μl (10μCi/μl) [α-^{35} S] dATP and 2 μl (1.6 U/μl) of diluted Sequenase version 2.0. After mixing the contents thoroughly the tube was incubated at room temperature for 5 min.

4.2.3. TERMINATION REACTIONS

The ddNTP termination mixes (2.5 μl) were pipetted into their corresponding centrifuge tube, capped and pre-warmed to 37° C for at least 1 min. When the labelling reaction was complete 3.5 μl of the labelling mix were removed, added to each of the pre-warmed ddNTP tubes, mixed and then returned to the 37° C water-bath for a further 5 min. Stop solution (4 μl) was added to each of the termination reactions, mixed thoroughly and stored on ice until required.

Just before loading the gel the samples were heated to 75° C for at least 2 min and immediately loaded onto the gel. Denaturing gel electrophoresis was carried out as outlined in section 4.4.

4.3 DNA SEQUENCING USING THE T7 DNA POLYMERASE KIT

Single-stranded template DNA was isolated as described in section 3.1.1.

4.3.1. ANNEALING THE TEMPLATE AND THE PRIMER

Template (or control) DNA (10 µl containing 2.0 µg of single strand DNA), 2 µl primer (at a concentration of 0.5 pmol/µl) and 2 µl annealing buffer were pipetted into a sterile Eppendorf tube mixed and incubated at 60° C for 10 min. Once the annealing reaction was complete the tube was kept at room temperature until needed.

4.3.2. THE LABELLING REACTION

Labelling mix (3 µl diluted to a final concentration of 1.5 U/µl), 1 µl (10µCi/µl) of [α-^{35}S] dATP and 2 µl (1.5 U/µl) of diluted T7 DNA polymerase were added to the annealed template and primer. The components were mixed by gentle agitation and the contents incubated at room temperature for 5 min.

Whilst the incubation was in progress the "read-short" sequence

mixes were pre-warmed to 37° C for at least 1 min.

4.3.3. TERMINATION REACTIONS

Once the labelling reaction was complete 4.5 μl of this reaction

mixture was transferred into each of the four pre-warmed sequencing mixes.

The contents of each tube were incubated at 37° C for 5 min after mixing by

gentle agitation. Stop solution (5 μl) was added to each tube, mixed by

gentle agitation and then briefly microfuged to collect the contents at the

bottom of the tube. An aliquot (3 μl) of each stopped reaction was

transferred to a separate microfuge tube, heated to 75° C for at least 2 min,

and 2 μl of this solution was then loaded immediately into an appropriate

well on a sequencing gel. Denaturing gel electrophoresis was carried out as

outlined in section 4.4.

4.4. ELECTROPHORESIS OF THE SEQUENCING REACTIONS

4.4.1. PREPARATION OF POLYACRYLAMIDE SEQUENCING GELS

A 40% acrylamide solution was prepared in distilled water, and de-

ionised with a small quantity of mixed bed resin (e.g. Amberlite MB1); the

resin was removed by filtration and the solution stored at 4° C. The sequencing plates were prepared as described below.

The notched plate was treated with "Repel" silane solution whilst the other plate was treated with diluted "Bind" silane. After inserting the side spacers and clamping the two matched plates together the system was ready to receive the gel solution, which was prepared as follows:

To 70 ml of 1X TBE gel mix (see Appendix II), 140 µl of 25 % ammonium persulphate and 90 µl of TEMED were added. It should be noted that 25 % ammonium persulphate solutions were freshly prepared just before use. If necessary the gel could be stored overnight at room temperature.

The electrophoresis tank reservoirs were filled with the appropriate running buffers; the upper chamber contained 0.5 X TBE and the lower buffer chamber contained 1 X TBE. Following the pre-run and before loading the samples approximately 0.5 volume of 3 M sodium acetate was added to the lower chamber.

The sequencing gel was pre-run at 1200 to 1400 V for 2.5 h before loading any samples.

4.4.2. SEQUENCING GELS PREPARED WITH HYDRO-LINK ™ DNA SEQUENCING SYSTEM

Both the glass plates were silanized with "Repel" silane and then pre-assembled in the usual way.

Hydro-link gel solution (75 ml) and Hydro-link denaturing gel buffer (25 ml) were poured directly into a beaker followed immediately by 220 μl of TEMED and 500 μl of freshly prepared 20 % ammonium persulphate. After gentle mixing the gel solution was poured into the pre-assembled glass plates laid on a flat surface, the sample comb was inserted and the gel left to polymerise.

Hydro-link SEQ running buffer (50 ml of 20 X concentrate) was made up to 1 litre and then both the anode and cathode reaction chambers were filled according to the manufacturers recommendations.

Normally for 30 to 300 base extensions the pre-run conditions were as follows: constant power (25 to 30 W) for 15 min.

Alternatively, if up to 300 base extensions and the bases close to the primer were to be read then the conditions outlined above could still be used but the pre-run was extended to 30 min.

The actual sequencing electrophoretic run was carried out at 25 to 30 W constant power for a period of 2 to 2.5 h. The temperature of the water jacket was thermostatically maintained at 45° C.

4.5. AUTORADIOGRAPHY OF SEQUENCING GELS

Upon completion of the electrophoretic run and following the separation of the glass plates the gel was fixed in 10 % acetic acid plus 12 % methanol, washed in water and then dried on a gel-dryer at 80° C.

73

X-ray film (Amersham Hyperfilm MP) was placed in direct contact with the gel and autoradiographed overnight at -70° C.

Standard X-ray developers (Kodak Industrex developer) and fixers (Ilford Hypam fixer) were used to develop the autoradiographs.

However, when using the hydro-link sequencing system the following adaptations were made. After separating the glass plates carefully the gel was allowed to air-dry for 5 min before transferring it to the support paper. The paper and gel were carefully peeled off the glass plate, and then the gel was covered with Saran wrap or ordinary cling film and the resulting "sandwich" dried on a gel-dryer at 60° C.

Once the gel was dry it was autoradiographed and developed as outlined above.

5.0. ELECTROPHORETIC MOBILITY SHIFT ASSAY

5.1. PREPARATION OF OLIGONUCLEOTIDES CORRESPONDING TO THE PUTATIVE BINDING SITE

Oligonucleotides corresponding to a putative DNA - protein interaction site were synthesised as described in Appendix IV.

Following cleavage and deprotection the oligonucleotides were ethanol precipitated, dried and annealed as described in the following section.

5.1.1. ANNEALING OF SYNTHETIC OLIGONUCLEOTIDES

The complementary oligonucleotides were diluted to give a final concentration of 200 μg per ml.

To 4 μl (800 ng) of each oligonucleotide, 2 μl of 5X annealing buffer were added, mixed thoroughly and incubated at 90^0 C for 10 min. Then the water bath was adjusted to 37^0 C and the oligonucleotide solution was allowed to cool, after 30 min incubation at 37° C, the annealed mixture was then allowed to stand at room temperature for 90 min.

5.1.2 5'-END LABELLING OF OLIGONUCLEOTIDES WITH γ-^{32}P ATP AND POLYNUCLEOTIDE KINASE

To 4 μl of annealed oligonucleotide (section 5.1.1), 2 μl of 10X kinase buffer, 10 μl (100μCi) of γ-^{32}P ATP, 2μl (10 U/μl) of Polynucleotide kinase and 2μl of water were added, mixed and incubated at 37^0 C for 30 min, before adding 40 μl of water, 30 μl of 7.5 M ammonium acetate, 1 μl of 0.25% linear polyacrylamide as a carrier and 400 μl of ethanol and incubating at -20^0 C for 30 min.

The resulting mixture was microfuged at 14 krpm for 20 min at 4^0 C, and the supernatant removed. The pellet was washed with 70 % ethanol, dried on a vacuum dessicator before being resuspended in 100 μl of water

and frozen until required for the bandshift or electrophoretic mobility shift assay.

5.2. PREPARATION OF NUCLEAR EXTRACTS

The cells for the preparation of the nuclear extracts were grown on 90 mm plates, the medium was removed and the cells washed with cold phosphate buffered saline (PBS).

The cells were harvested in PBS by scraping with a cell scraper, followed by centrifugation in 50 ml Falcon tubes at 2 krpm for 5 min at 4 °C using the SL50T rotor, resuspension in 4 ml of cold buffer A (see Appendix II), and a 10 min incubation on ice.

Finally the cells were disrupted by 10 strokes with a tight-fitting Dounce homogeniser and the nuclei pelleted by centrifugation at 16 krpm for 20 min at 4 °C using a Sorvall SL50T rotor. The pellet was resuspended in 4 ml of cold buffer C (see Appendix II) and the nuclei disrupted by 10 strokes of a tight fitting Dounce homogeniser.

The nuclear extract was stirred gently on ice for 45 minutes before pelleting the chromatin and nuclear membranes by centrifugation at 4 °C for 20 min at 25 krpm using the Sorvall T865 rotor. Ammonium sulphate was then used to precipitate the proteins present in the supernatant. Solid ammonium sulphate (0.33 g/ml) was gently stirred into the supernatant followed by continual gentle stirring on ice for a further 30 min.

76

The nuclear proteins were pelleted by centrifugation at 4 °C for 20

min at 25 krpm as in the previous step. Finally the protein pellet was

resuspended in 200 μl of cold buffer D (see Appendix II) and dialysed

against the same buffer overnight at 4 °C. The following day the nuclear

extract was snap frozen in liquid nitrogen in aliquots (50 μl) and stored at

–70 °C.

5.3. ESTIMATION OF THE CONCENTRATION OF PROTEIN IN THE NUCLEAR EXTRACTS

A range of protein solutions from 0 to 10 μg/μl were freshly prepared

from a stock protein solution of 10 gl^{-1}, at the same time an aliquot (15 μl) of

the nuclear extract was also thawed and dispensed into a sterile Eppendorf

tube and kept on ice until required. The working solution of Bicinchoninic

acid (BCA) reagent was then prepared by mixing 200 μl of a 4% copper

sulphate solution with 9.8 ml of BCA reagent.

The working BCA reagent (200 μl) was pipetted into an appropriate

number of wells in a microtitre plate, followed by 5 μl of the diluted protein or

nuclear extract sample. The plate contents were gently mixed, incubated at

room temperature for 30 min before being read on a microplate reader at

562 nm

The protein determination experiment was repeated with a standard

and sample volume of 10 μl because the calibration graph obtained from the

5 μl sample was a curve and not linear as expected.

5.4. BINDING ASSAY

The annealed oligonucleotide duplex (^{32}P-labelled at the 5' end to a specific activity of approximately 10 x 10^4 cpm / μg of DNA) was added to binding buffer (see Appendix II) and poly[d(IC).d(IC)] (at a final concentration of 1 μg per 10 μg of protein) in a reaction volume of 20 μl.

Nuclear extract (10 μg in 10 μl) was added, mixed and incubated at room temperature for 30 min. Following incubation, 5 μl of 20 % Ficoll was added to the reaction mixture and 5 μl aliquots loaded onto a 7 % non-denaturing polyacrylamide gel.

Electrophoresis was carried out at 7 Vcm^{-1} until the bromophenol blue marker loaded in an adjacent well had migrated about 5 cm. The gels were dried and autoradiographed as previously described in section 4.5.

5.5. COMPETITION - BINDING ASSAYS

The competition - binding experiments were carried out essentially as described in section 5.4, except that each reaction mixture contained a 100-fold excess of the unlabelled oligonucleotide duplex, in addition to the end-labelled duplex.

6.0. CLONING OF THE PUTATIVE BINDING SITE IN VECTOR pHISI-1 FOR USE IN THE YEAST ONE-HYBRID SYSTEM

6.1 PREPARATION OF OLIGONUCLEOTIDES CORRESPONDING TO THE PUTATIVE BINDING SITE FOR CLONING IN pHISi-1

Complementary strands of the oligonucleotide with four tandem copies of the putative binding site corresponding to nt sequence –609 to –577 (5'-CACCTTTTGTCAAACGACCGCTTCACATCTGGG-3') were synthesised as described in Appendix IV, except that the 5' terminal trityl group was left on the oligonucleotide to aid purification (see Appendix V). An *EcoR1* and an *Mlu1* overhang were also incorporated to facilitate cloning in the reporter vector.

Each oligonucleotide (75 µl at a final concentration of 400 µg/ml) were added to 50 µl of 5X Reaction Buffer (from any sequencing kit, see Appendix II), mixed thoroughly and heated to 80° C for 10 min before slowly tempering down to room temperature before being used in the ligation reaction.

6.2. LIGATION OF THE OLIGONUCLEOTIDES CONTAINING THE BINDING SITE TO pHISi-I PLASMID

The plasmid pHISi-1 (10 µg) was digested with *EcoR1* (20 U/µl) and *Mlu1* (10 U/µl) as previously described, however, following each digestion the restriction endonucleases were inactivated with TE- saturated phenol (2 extractions) followed by ether extraction (6 extractions) to remove any residual phenol, and subsequent ethanol precipitation and drying of the nucleic acid pellet.

The digested reporter vector (10 µg in 10 µl) and 10 µl (4.0 µg) of the annealed oligonucleotide with multiple copies of the binding site were added to 3µl 10X ligation buffer, 3µl of ATP (final concentration 2mM) and 4 µl of T4 DNA Ligase (400 U/µl), and incubated overnight at room temperature.

Following incubation the ligation mixture was electrophoresed on a low melting temperature agarose gel; in order to separate the ligation product from unligated vector and excess annealed oligonucleotides. This gel slice containing the ligation product was melted in a water bath at 65°C, diluted one part to two parts in TE buffer (pH 8.0) (see Appendix II) before being used in the transformation protocol already outlined in Section 2.6.4.2.

6.3. SECREENING OF TRANSFORMED BACTERIAL COLONIES BY PCR

A master mixture for 15 X 50 µl PCR reactions was prepared as described in Table 1:

HIS - RP Primer	15 µl
HIS - 5372 Primer	15 µl
10 X PCR Buffer	75 µl
50 mM Magnesium Chloride	22.5 µl
2 mM dNTPs	75 µl
Water	547.4 µl

Table 1. Components of master mixture

After mixing the components of the master mixture thoroughly 48 μl aliquots were dispensed into labelled Eppendorf tubes and to each was added a bacterial colony, followed by 1 μl of Taq Polymerase (5 U/μl). Finally 50 μl of mineral oil was added to each tube prior to thermal cycling according to the following programme:

STEP 1	94 °C	for 7 min
STEP 2	95 °C	for 1 min
STEP 3	52 °C	for 1 min
STEP 4	72 °C	for 1 min
STEP 5	72 °C	for 10 min
STEP 6	4 °C	for ∞

Cycle Back to step 2 for 34 Cycles

Any 'positive' clones (i.e. those indicating the presence of the binding site ligated to the reporter vector) were cultured in fresh L-broth, the plasmid DNA isolated by previously described small-scale procedures, followed by digestion with *EcoR1* and *Mlu1* according to the protocols outlined earlier. Further confirmation of the presence of the binding site in the vector was obtained from chain termination sequencing using the protocol outlined in section 4.2.

7.0. LARGE SCALE PREPARATION OF YEAST VECTOR CONTAINING THE BINDING SITE (REPORTER VECTOR)

Representative clones identified by PCR were inoculated into L broth (5 ml) containing 5μl of ampicillin (50 mg/ml), and incubated at 37 °C for 18 h with constant shaking prior to the inoculation of 200 ml of Terrific broth (see Appendix II) containing 800 μl of ampicillin (50 mg/ml) and incubation at 37 °C for a further 18 - 24 h.

The overnight cell suspension was transferred into 500 ml centrifuge bottles and centrifuged at 8 krpm for 15 min at 4 °C using a GS3 rotor, after decanting the supernatant the bottles were inverted and allowed to drain before resuspending the pellets in 10 ml of solution A (see Appendix II), this was followed by 20 ml of freshly prepared solution B (see Appendix II). The bottles were swirled for about 1 min until the solution became clear, and then incubated on ice for 10 min.

Chilled solution C (10 ml) (see Appendix II) was added and the bottles shaken vigorously to mix the contents, following a 5 min incubation on ice the bottles were centrifuged at 8 krpm for 15 min at 4 °C using a GS3 rotor. After filtering the supernatant, 25 ml of isopropanol was added to each bottle followed immediately by centrifugation at 8 krpm for 10 min at 4 °C as above, the supernatant was discarded, the bottles inverted and the pellets air dried.

The pellets were resuspended in 2 ml of T.E. buffer (pH 8.0) (see Appendix II), 50 μl of RNase A (10 mg/ml) was added to each sample before incubation at 37 °C for 30 min. The resulting mixture was then extracted with phenol-chloroform twice and chloroform- isoamyl alcohol twice. 500 μl of

PEG 8000 (20 %) and 25μl magnesium chloride (1 M) were added followed

by immediate centrifugation in a microfuge at room temperature for 10 min.

The supernatant was discarded and the pellets washed twice in 70 %

ethanol prior to air drying and resuspension in 500 μl of T.E. buffer (pH8.0).

Aliquots (10 μl) were removed from each sample diluted to 1 ml with

T.E buffer pH 8.0 and the concentration of DNA estimated using the

Pharmacia Gene-Quant. An aliquot (1μl) was also analysed by agarose gel

electrophoresis to check the purity and integrity of the DNA isolated.

8.0. PREPARATION OF THE REPORTER STRAIN FOR USE IN THE YEAST ONE-HYBRID SYSTEM

The one-hybrid system is an *in vivo* yeast genetic assay used for

isolating cDNAs encoding proteins that bind to a short target DNA sequence

e.g. *cis*-acting transcription regulatory elements. It is based on the

interaction between a specific protein DNA-binding domain and the target

DNA sequence. A cDNA that may encode a DNA-binding domain is

expressed as a fusion protein with a target independent transcription

activation domain e.g. GAL4 AD. To carryout a one-hybrid assay at least

three copies of a putative binding sequence are inserted upstream of the

HIS3 gene minimal promoter in a specially designed plasmid e.g. pHISi-1, as

in this case (thus giving a reporter construct). The reporter construct is

integrated *in vivo* into a mutated *HIS3* gene in the yeast genome to create a

yeast reporter strain with a normal copy of the *HIS3* gene. A library of cDNAs

encoding GAL4 AD fusions is expressed in the reporter strain. If an

AD/library protein interacts with the target DNA sequence, *HIS3* promoter

activity is increased resulting in increased amounts of histidine produced.

Clones expressing histidine at a level higher than the reporter strain are

selected by growth on medium lacking histidine in the presence of 3-

aminotriazole (3-AT). In this thesis a rat brain cDNA-GAL4 library was

screened for cDNA(s) encoding protein(s) binding to the –609nt to –577nt

region in the mouse MBP promoter.

Diagrams illustrating the procedure are represented in figures 6a to 6d on

the following pages.
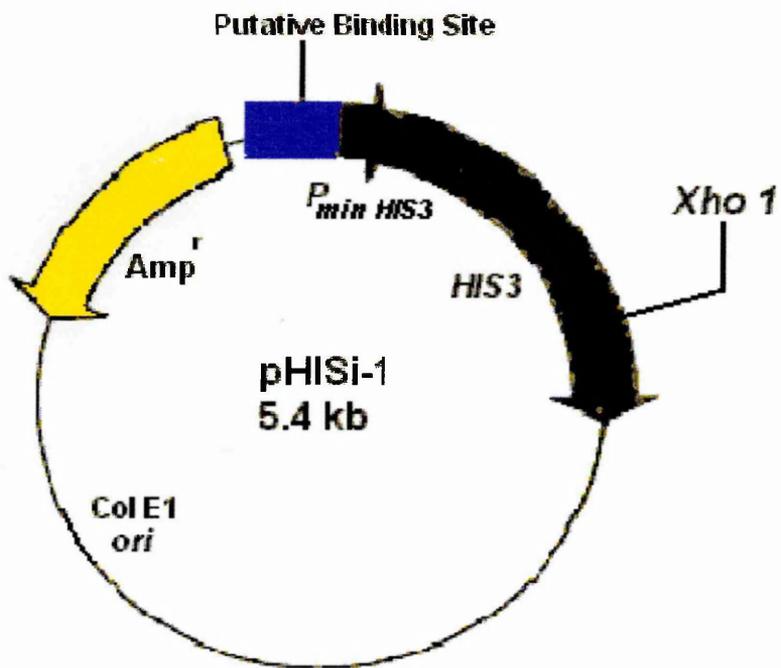
FIGURE 6A. The reporter plasmid pHISi-I, carrying four tandemly repeated copies of the putative binding site, was linearised with *Xho I* restriction endonuclease at a unique site located within the *HIS3* gene before integration by homologous recombination into the mutated *HIS3* gene locus of yeast strain YM4271 (used as the host).
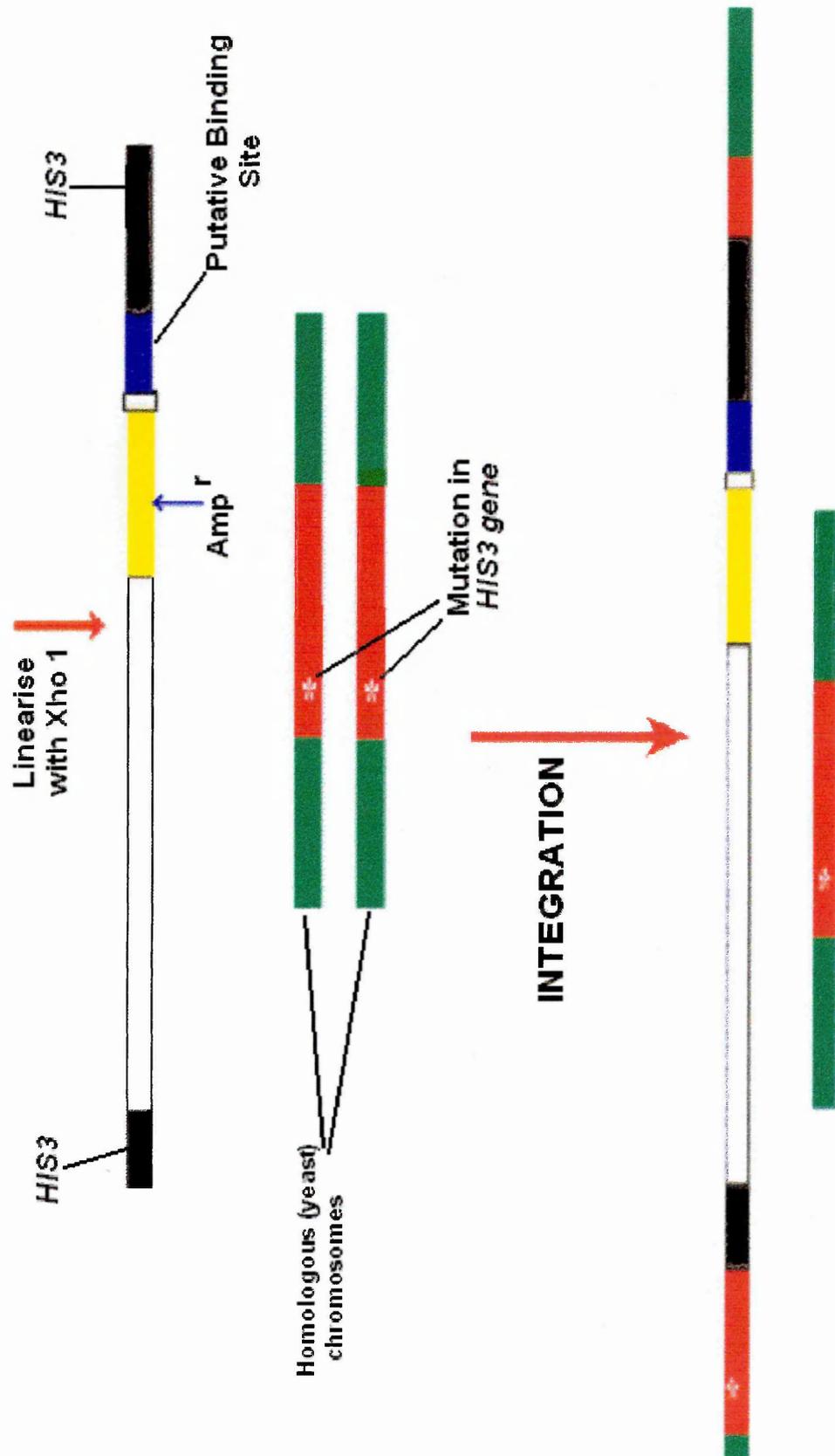
FIGURE 6B. A diagrammatic representation of the integration of the reporter plasmid into the yeast genome to produce a reporter strain carrying a normal copy of the HIS3 gene.

86

CCA AAA AAA GAG ATC TCT CGA GGA TCC GAA TTC
      Bgl II Xho I   EcoR I
           BamH I

CAG ATC TAT GAA TCG TAG ATA CTG AAA AAC
Bgl II stop stop stop
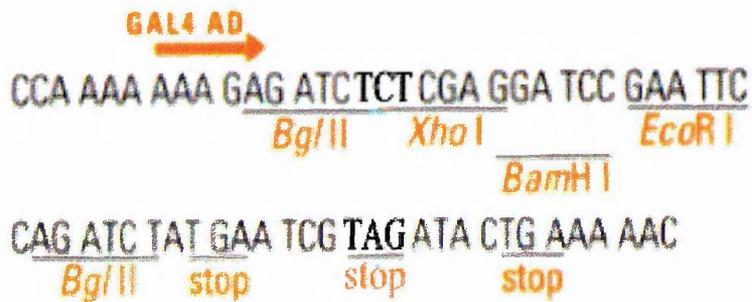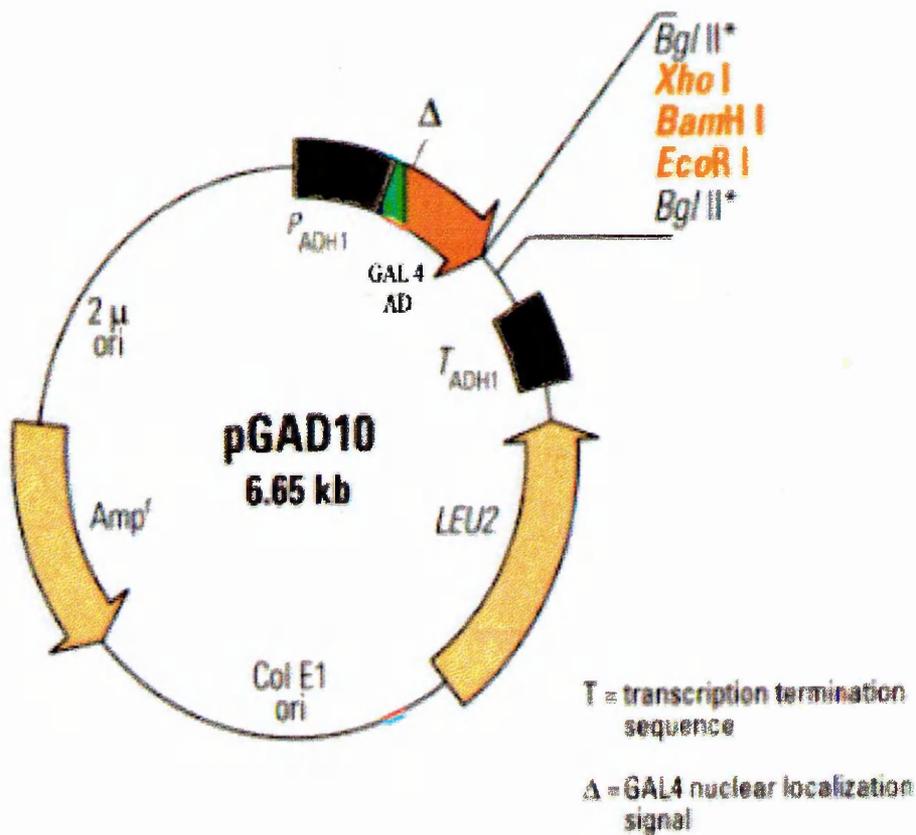
Rat brain cDNA sequences are cloned in the unique EcoRI restriction endonuclease site of the multiple cloning site of pGAD10.

FIGURE 6C. A diagrammatic representation of the yeast AD cloning vector.

(Adapted from the Clontech One-hybrid system protocol)
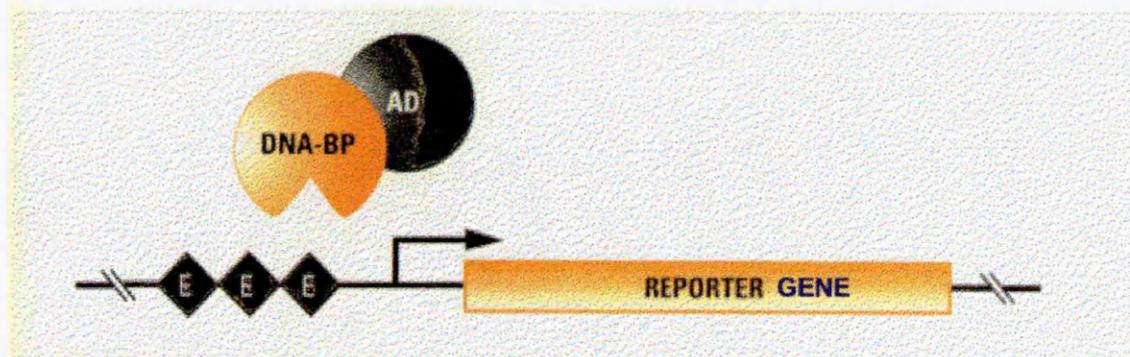
A sequence consisting of at least three copies of a putative protein-binding site (E) is inserted upstream of a reporter gene and integrated into the yeast genome (reporter strain). This reporter strain is transformed with a cDNA library that comprises a fusion of target specific binding proteins and the target-independent GAL4 activation domain. Binding of a cDNA/AD hybrid to one of the target elements results in the activation of reporter gene transcription.

FIGURE 6D. Diagrammatic representation of the activation of transcription following transformation of a cDNA/AD fusion library in the yeast reporter strain.

(Adapted from the Clontech One-hybrid system protocol)

8.1.   LINEARISATION OF THE TARGET-REPORTER-VECTOR

Target-reporter plasmid containing the binding site was prepared as described in 7.0 and digested with *Xho1* enzyme. The reaction mixture consisted of 20 µl DNA (20 µg), 3 µl of *Xho1* buffer (10X), 3 µl (20 U/µl) of *Xho1* restriction enzyme, and 4 µl of water. After digestion for 2.5 h at 37 °C the restriction endonuclease was inactivated with TE - saturated phenol (2 extractions). This was followed by ether extraction (6 times) to remove residual phenol. DNA was precipitated by a standard procedure (as described in 2.3).

The DNA pellet was resuspended in 25 µl of sterile water and the completeness of the digestion was checked by agarose gel electrophoresis using undigested plasmid as a control.

8.2.   PREPARATION OF COMPETENT YEAST CELLS

A single colony of *Saccharomyces cerevisiae* (strain YM4271) was inoculated into 1 ml of YPD medium (see Appendix II) and vortexed vigorously to disperse the cells. The cell suspension was transferred to 50 ml of YPD medium and incubated at 30 °C for 18 to 24 h with shaking until the culture was saturated.

An aliquot of this culture (10 – 15 ml) was transferred to 100 ml of fresh YPD medium to give an initial absorbance at 600 nm of between 0.2 - 0.3 and incubated at 30 °C with shaking for 3 - 4 h.

The cells were collected by centrifugation in 50 ml tubes at 2.2 krpm

for 5 min at room temperature using the SL50T rotor, the pellet rinsed in

25 ml of TE buffer pH 7.5 (see Appendix II) and the centrifugation repeated.

The cell pellet was resuspended in 1 ml of freshly prepared TE/lithium

acetate solution.

8.3. TRANSFORMATION OF COMPETENT YEAST CELLS

Linearised reporter vector carrying four copies of the binding site

(approximately 1μg in 20 μl of TE buffer) was mixed with carrier DNA

(100 μg in10 μl), and added to 100 μl of yeast competent cells.

As a negative control 100 μg of carrier DNA (sheared and denatured)

was added to 20 μl (1 μg) of undigested reporter vector carrying the binding

site, mixed and added to 100 μl of yeast competent cells. The positive

control consisted of 100 μg of carrier DNA (sheared and denatured) and

20 μl (1μg) of Xho1digested p53HIS plasmid containing three tandem

repeats of the consensus p53 binding site.

Freshly prepared PEG/LiAc solution (600 μl) was added to each

transformation mixture followed by incubation with shaking at 30 °C for

30 min to 1 h. 70 μl of DMSO were added to each tube, and a heat shock at

42 °C for 15 min followed by an incubation on ice for 2 min was performed.

The cells were spun down at 14 krpm for 5 – 10 sec in a microfuge, the

supernatant removed and the pellet resuspended in 150 μl of TE buffer.

The pHISi-1 and the p53HIS transformation mixtures were then plated onto SD/-HIS plates, incubated for 4 to 6 days at 30 °C to select for colonies with an integrated reporter gene. Colonies containing a functional integrated reporter gene were distinguishable by their larger size.

Master stock cultures of the strain containing the integrated target – reporter constructs were grown in SD/-HIS medium containing 25 % glycerol and kept at –70 °C until required.

8.4. ANALYSIS OF PUTATIVE REPORTER STRAIN CONSTRUCTS (THE RAPID ISOLATION OF YEAST CHROMOSOMAL DNA)

A culture (10 ml) of the yeast colony to be analysed was grown overnight to stationary phase in YPD medium at 30 °C, with shaking.

Following centrifugation at 3 krpm at room temperature the cell pellet was resuspended in 500 μl water, and transferred to a sterile Eppendorf tube to be pelleted again. The cell pellet was disrupted in 200 μl of breaking buffer (see Appendix II), and 200 μl (approximately 0.3 g) of acid-washed glass beads and 200 μl of phenol/chloroform/isoamyl alcohol were added and vortexed at high speed for 5 min. Following the addition of 200 μl of TE buffer pH 7.5 (see Appendix II) and brief vortexing the mixture was microfuged at full speed for 5 min at room temperature before and the aqueous layer was transferred to a fresh Eppendorf tube.

Absolute ethanol (1 ml) was added to the aqueous phase, mixed by inversion before microfuging at full speed for 5 min at room temperature. The

91

pellet was resuspended in 400 μl of TE buffer, RNase A (3 μl of a 10 mg/ml stock) was added, mixed and incubated at 37 °C for 5 min.

The DNA was recovered by precipitation with 10 μl of 4M ammonium acetate and 1 ml of ethanol and microcentrifugation at full speed for 5 min at room temperature. The DNA pellet was dried *in vacuo*, resuspended in 25 μl of TE buffer and used in the PCR screening protocol as described previously.

8.5. TESTING NEW REPORTER STRAINS FOR BACKGROUND *HIS3* EXPRESSION

The colonies identified as containing integrated target – pHISi-1 were screened for background *HIS3* expression.

A single colony was picked from the SD/-HIS plate and suspended in 1 ml of TE buffer in a sterile Eppendorf tube. The cell suspension (5 μl) was pipetted onto the surface of an SD/-HIS plate containing either 0, 5, 10, 15, 30, 45, or 60 mM 3 – Aminotriazole (3 – AT). The plates were incubated for 4 to 6 days at 30 °C and the results analysed.

Yeast colonies growing on the SD/-His + >45 mM 3 - AT plates were considered to have high (very leaky) background *HIS3* expression. Similarly if the yeast colony growth was inhibited on the SD/-HIS + < 45 mM 3 – AT plates then the background *HIS3* expression was low.

Colonies of *HIS3* reporter strain with the lowest background level of

HIS3 expression were selected for use in the one-hybrid library screening

protocol.

Several SD/-HIS/-LEU plates containing the optimal concentration of

30 mM 3 – AT were prepared. The 3 – AT was present to maintain a positive

selection pressure on the reporter clones. The plates were dried before

spreading the cells as condensation could cause localised variations in the

3 – AT concentration, that could lead to false positive results

9.0.   THE SEARCH FOR TRANSCRIPTION FACTORS

9.1.   SCREENING OF A cDNA / AD FUSION LIBRARY

The preparation of competent reporter strain cells was carried out

essentially as described in section 8.2, the only modification being the

inoculation of reporter strain colonies into liquid SD/-His medium containing

30 mM 3 – AT, instead of YPD medium.

cDNA/AD fusion library (approximately 20 µg in 5 µl of water) was

mixed with carrier DNA (2 mg in 200 µl) and added to 1 ml of competent

reporter strain cells. Freshly prepared LiAc/PEG solution (6 ml) was added

to each transformation mix followed by incubation at 30 °C for 30 min to 1 h

with constant shaking. Then 700 µl of DMSO were added to each tube

followed by a heat shock at 42 °C for 15 min and an incubation on ice for

2 min was performed. The cells were pelleted at 2.2 krpm for 5 min, and

resuspended in 2 X 25 ml of SD/-His liquid medium in two separate 50 ml

tubes.

After incubation at 30 °C for 1 h with constant shaking, the cells were

pelleted at 2.2 krpm for 5 min at room temperature and resuspended in 7 ml

of TE buffer and spread on the SD/-His/-Leu + 30 mM 3 – AT plates.

It should be noted that immediately after pipetting the cells onto the

plate they were spread over the surface to prevent localised dilution of the

3 – AT concentration.

The plates were incubated at 30° C for 4 to 6 days. Large colonies

resulting from His3 activation were subcultured into fresh SD/-His/-Leu/

+ 30 mM 3 – AT medium and used to isolate plasmid DNA for further

analysis.

9.2.   ANALYSIS OF PUTATIVE CLONES CARRYING AN AD FUSION LIBRARY PLASMID (RAPID ISOLATION OF YEAST PLASMID DNA)

Representative clones carrying putative cDNA/AD fusion library

plasmid were inoculated into 5ml of SD/-HIS/-Leu liquid medium containing

30 mM 3 – AT and incubated at 30° C for 48 h with constant shaking.

The cells were pelleted by centrifugation at 2.2 krpm for 5 min at room

temperature; the pellets were washed with TE buffer and resuspended in

200 μl of breaking buffer. Acid-washed glass beads 200 μl (approximately

0.3 g) and 200 μl of phenol/chloroform/iso-amyl alcohol were added, before

94

vortexing at high speed for 2 min. After centrifugation at 14 krpm for 10 min

at room temperature in a microfuge, the supernatant was transferred to a

clean Eppendorf tube and the plasmid DNA precipitated with 0.1 volume

sodium acetate (3 M pH 6.0), 1 µl glycogen (at a concentration of 1 mg/ml),

and 2.5 volumes of ethanol.

After incubation at −70° C for 30 min, the plasmid DNA was pelleted

by microcentrifugation at 14 krpm for 10 min at room temperature, washed

with 70 % ethanol, dried *in vacuo* and resuspended in 20 µl TE buffer.

9.3. ANALYSIS OF PUTATIVE CLONES CARRYING A LIBRARY PLASMID BY PCR

An aliquot (2.5 µl) of the plasmid described in section 9.2 was

analysed by PCR using the library primers,

AD10[S] 5'-CTATTCGATGATGAAGATACC-3' (located between

nucleotides 788 to 808 on the plasmid sequence in Appendix VI), and

AD10[R] 5'-CACAGTTGAAGTGAACTTGCG-3' (located between

nucleotides 903 to 883), essentially as described in section 6.3 except that

the annealing temperature of the primers to template during thermal cycling

was reduced to 50° C.

The products of the PCR were analysed by agarose gel

electrophoresis. Plasmid DNA (10 µl) from positive clones carrying a library

plasmid identified by PCR was transformed into *E. coli* strain DH5α using the

procedure described in section 2.6.4.2.

It should be noted that plasmid DNA isolated by the technique described in section 9.2. is not of suitable quality for sequencing due to yeast genomic DNA contamination, therefore, transformation into *E. coli* cells is essential.

9.4. IDENTIFICATION OF DNA – BINDING PROTEINS BY AUTOMATED SEQUENCING AND DATABASE SEARCHES

Plasmid DNA was isolated and purified from the transformed *E. coli* cells using the Modified Clewell and Helinski technique and centrifugation to equilibrium in caesium chloride-ethidium bromide gradient as described in sections 2.2. and 2.3. respectively.

An aliquot of plasmid DNA (20 μl at a concentration of 0.1 μg/μl) was sequenced automatically.

From the sequence data obtained comparisons were made with those of known nucleotide and protein sequences for transcription factors on other databases (e.g. EMBL, Genbank and Swissprot) using the FastA and Blast search tool.

Chapter 3

Results

3.1. Sub – cloning of the myelin basic protein gene promoter in the plasmid pBluescript for the purposes of sequencing and functional characterisation

The 6.02 kb plasmid pEX1 (a gift from Arthur Roach Division of Biology, California Institute of Technology, Pasadena, California) shown schematically in Figure 7 contains the promoter region and a segment (including exon 1 which is shown cross-striped in blue) of the mouse myelin basic protein gene. On the basis of the restriction map of this plasmid a strategy was designed to obtain the fragment of DNA carrying the promoter region by excising it with *Hind III* and *Sac I* restriction endonucleases. This DNA fragment was subsequently used in cloning for further studies.

Analysis of the restriction endonuclease digestion products (Figure 8 lane 3) shows three distinct bands ranging in size from 1.40 to 2.68 kb. The smallest fragment (1.40 kb) represents the *Hind III – Sac I* digestion product containing the promoter region and segment including exon 1 of the myelin basic protein gene. The 2.68 kb digestion product represents the *Hind III – Hind III* fragment of the plasmid pUC13 (shown cross-striped in red in Figure 7), which was used as the cloning vector, and the third band (1.94 kb) represents the *Sac I – Hind III* fragment of the myelin basic protein gene.
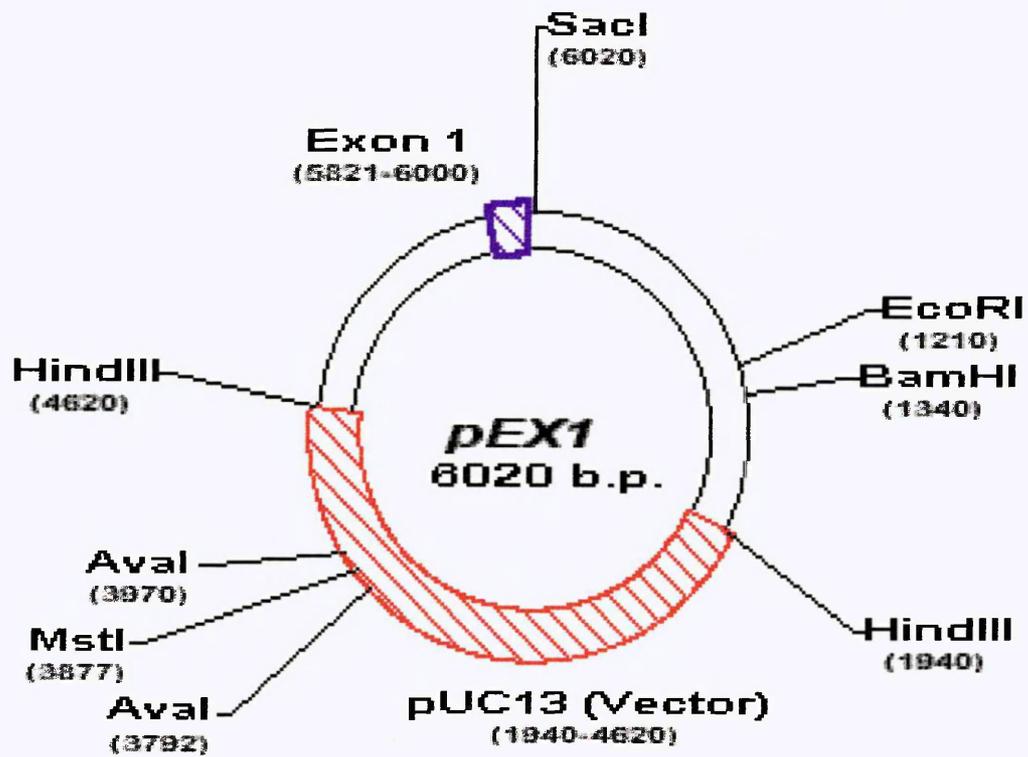
Figure 7. Diagrammatic representation of the recombinant plasmid pEX1 containing the promoter region and segment (including exon 1) of the mouse myelin basic protein gene.

The products of the digestion of pEX1 with *Hind III* form two bands

(Lane 1); the slower band corresponds to the larger 3.34 kb fragment, which

contains the promoter region and segment (including exon 1) of the mouse

myelin basic protein gene, and the smaller fragment (faster band) containing

the 2.68 kb part of vector pUC13.

Only a single band corresponding to the linearised 6.02 kb plasmid

pEX1 after digestion with *Sac I* can be seen in lane 2.

Uncut pEX1 DNA (0.8 µg) used as a control shows more than one

band, these represent the supercoiled form (lower band) and the open

circular form (upper band) of the plasmid (figure 8, Lane 4).
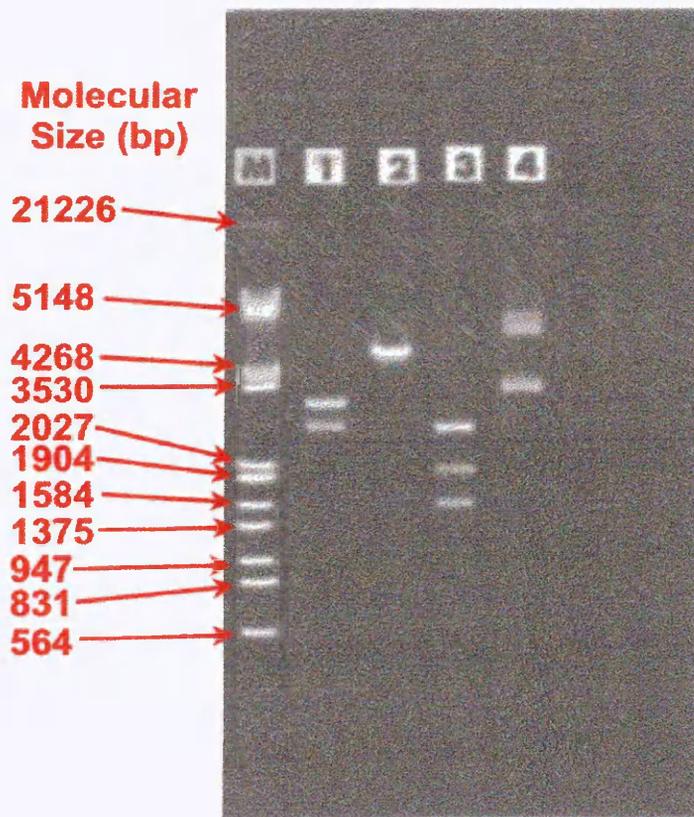
Figure 8. Analysis by agarose gel (1%) electrophoresis of the restriction
endonuclease digestion products of the plasmid pEX1 with
Hind III and Sac I.

Lane M. A λ DNA molecular size marker (*EcoR1 – Hind III* double digest of λ
phage DNA).
The figures in red represent the molecular size in base pairs.

Lane 1. Digestion of the plasmid pEX1 with *Hind III* restriction endonuclease.
Lane 2. Plasmid pEX1 following digestion with *Sac I* restriction
endonuclease.
Lane 3. Digestion of the plasmid pEX1 with a combination of *Hind III* and
Sac I.
Lane 4. Undigested plasmid pEX1 (0.8 μg).

The fragment containing the MBP promoter region was excised from the gel, the DNA electro-eluted, and purified by ion exchange chromatography on DE-52 cellulose. Following ethanol precipitation the DNA was resuspended in sterile deionised water and analysed by agarose gel electrophoresis (Figure 9, Lane 2).

The purified 1.40 kb DNA fragment containing the promoter region of the mouse myelin basic protein gene was then cloned into the plasmid vector pBluescript (a map of this plasmid is given in Appendix VI).

For this purpose, pBluescript was digested with *Sac I* and *Hind III* restriction endonucleases (data not shown). Two different restriction endonucleases were used to prevent religation of the pBluescript vector. The restriction endonucleases were inactivated by treating the digestion mixture with phenol. Water-saturated ether was then added to remove any residual phenol, and the plasmid DNA ethanol precipitated. Following resuspension of the plasmid pBluescript in deionised water the purified 1.40 kb fragment was added and ligation performed as described in 2.6.3.
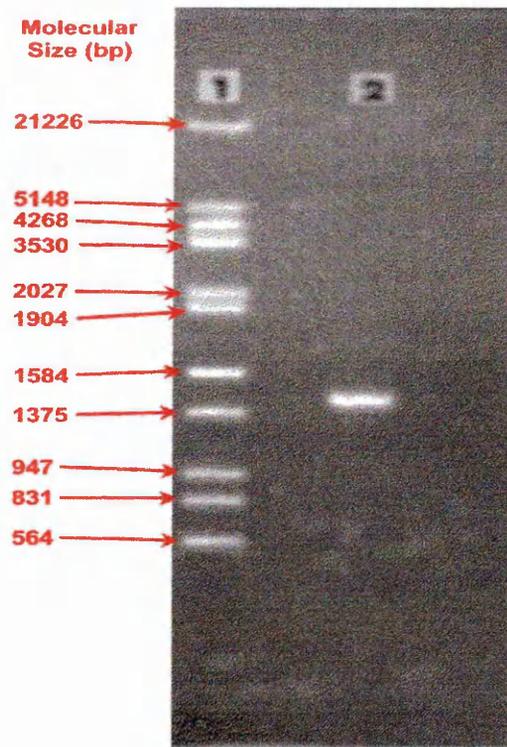
Figure 9. Analysis by agarose gel (1%) electrophoresis of the 1.4 kb
fragment containing the promoter region and segment of the mouse
myelin basic protein gene after isolation and purification from the
plasmid pEX1.

Lane 1. A λ DNA molecular size marker (*EcoR1 – Hind III* double digest of λ
phage DNA).

Lane 2. Purified *Hind III – Sac I* fragment containing the promoter region of
the mouse myelin basic protein gene.

Competent cells were prepared by growing *E coli* (strain DH5α) to an

absorbance of 0.5 (at 650 nm) followed by treatment with ice-cold calcium

chloride as described in 2.6.4.1. The ligation product consisting of

pBluescript, with the promoter region and segment of the mouse myelin

basic protein gene, was mixed with the competent cells. A heat shock at

42° C followed by an incubation on ice was performed. After incubation at

37° C for 1 h, to allow for the expression of the β-lactamase gene carried by

the plasmid pBluescript, the bacteria were spread over the surface of pre-

dried AIX plates (Figure 10).

Many vectors including pBluescript contain the regulatory sequences

and coding region for the first 146 amino acids (called the α-peptide) of the

β-galactosidase gene (*lacZ*). Embedded in this coding region is a

multicloning site. Vectors of this type are used in host cells that are able to

synthesise the remainder of the β-galactosidase gene. Neither the plasmid

nor the host-encoded fragments are active, but they can associate to form an

enzymatically active protein. This process is called α-complementation. The

Lac[+] bacteria that result from α-complementation are easily recognised

because they form blue colonies in the presence of the chromogenic

substrate X-gal. Insertion of a fragment of DNA into the multicloning site of

the plasmid usually results in the production of an NH_2 – terminal fragment

incapable of α-complementation (insertional inactivation), so bacteria

carrying recombinant plasmids will form white colonies.

The chemical analogue of lactose (IPTG), which cannot be cleaved by

the enzyme β-galactosidase, is used as an inducer of the *E.coli lac* operon. It

binds to and inactivates the *lac* repressor; therefore, incubation with IPTG results in the maximum expression of genes cloned into suitable vectors. When used in combination with X-GAL, as in this project, it can be used as a selection system for recombinant plasmids.

Selection by ampicillin resistance was also used to prevent the growth of untransformed host cells.

White colonies were considered putative recombinant clones carrying the promoter region of the myelin basic protein gene and analysed further.

Blue colonies resulting from inefficient digestion were considered negative since they only contained the plasmid vector.
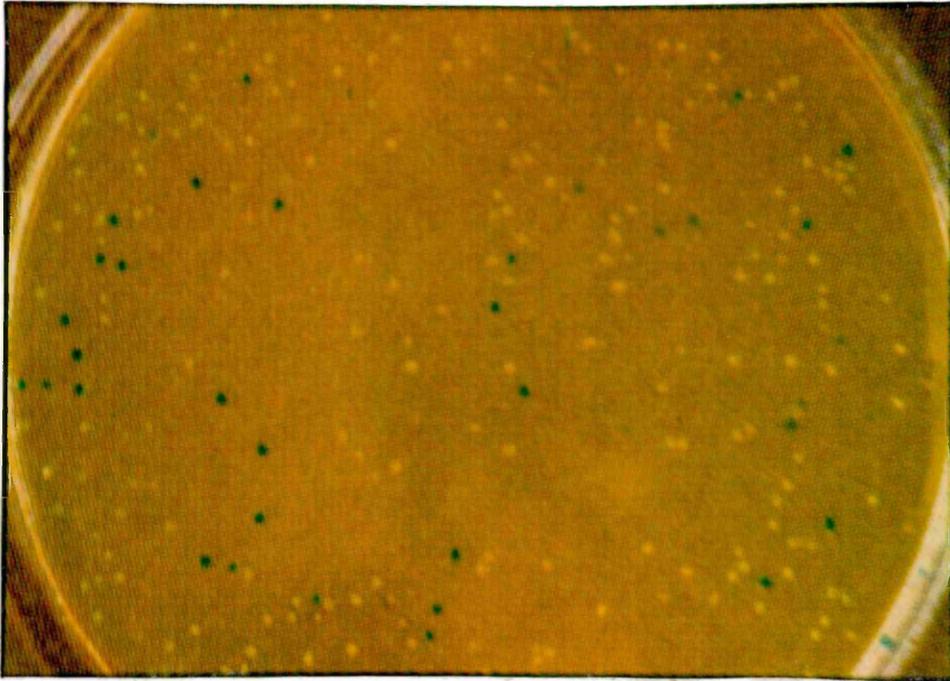
Figure 10. Selection of putative recombinant clones carrying the promoter region and segment (including exon 1) of the *MBP* gene using a selective medium containing the *E. coli lac* operon inducer IPTG and the chromogenic substrate X-gal.

Several (five) white colonies were grown in L-broth containing

ampicillin (50 μg/ml) with vigorous shaking. The plasmid DNA was isolated

using a small-scale procedure based on the alkaline lysis method of

Birnboim and Doly (1979).

Analysis of the plasmid DNA by agarose gel electrophoresis (Figure 11)

showed one band, this represents the supercoiled form of a plasmid of

approximately 4.4 Kb which corresponds to pBluescript (3.0 Kb) containing

the 1.4 Kb myelin basic protein gene promoter insert, (Figure 10, Lanes 1 -

5). pBluescript without the myelin basic protein gene insert was run as a

control sample and is shown in Figure 11, Lane 6 as a 3.0 Kb band.

An *EcoR1 - Hind III* double digest of λ phage DNA was used as a molecular

size marker (lane M) to determine the size of the plasmid DNA band

isolated.
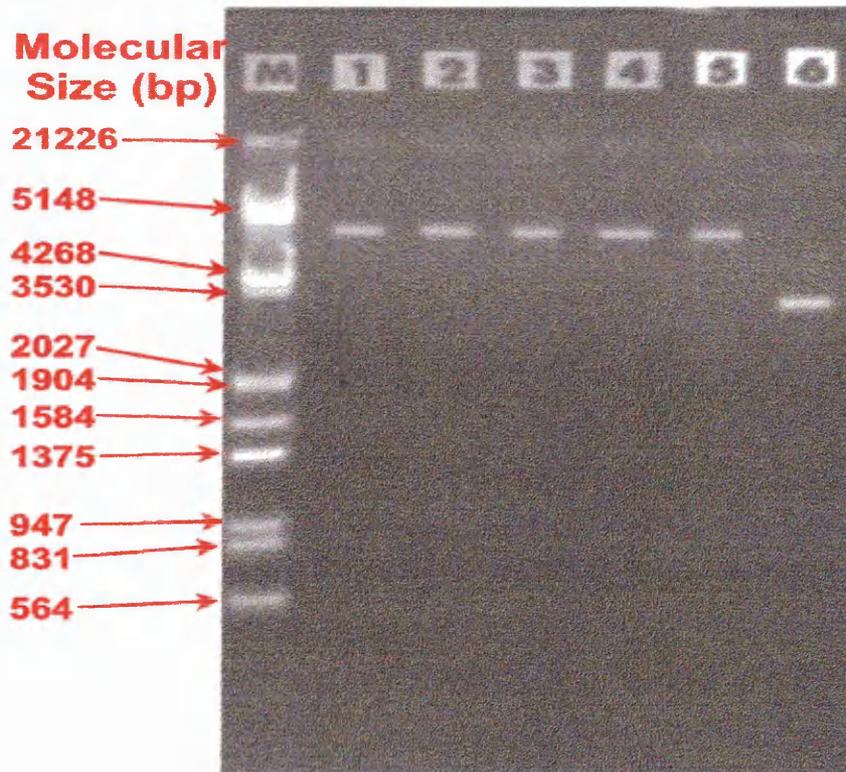
Figure 11. Analysis by agarose gel (1%) electrophoresis of plasmid DNA isolated from putative recombinant clones carrying the promoter region of the mouse myelin basic protein gene.

Lane M. A λ DNA molecular size marker (*EcoR1 – Hind III* double digest of λ phage DNA).

Lanes 1 – 5. Plasmid DNA isolated from the putative clones carrying the myelin basic protein gene insert.

Lane 6. pBluescript (without the *MBP* gene insert) run as control.

107

To confirm the presence of the 1.4 Kb fragment containing the myelin basic protein gene promoter the plasmid DNA was digested with a combination of *Hind III* and *Sac I* restriction endonucleases, and the products analysed by agarose gel electrophoresis (Figure 12). Two distinct bands can be seen:

i.) a 1.4 kb band putatively corresponding to the *HindIII - SacI* fragment of the promoter region and segment including the first exon of the mouse myelin basic protein gene.

ii.) a 3.0 kb band that corresponds to the vector pBluescript.

The weak band at 4.4kb is the remainder of undigested plasmid product from the recombinant plasmid carrying the myelin basic protein gene insert. Figure 12, Lane 6 shows only a single band of linear pBluescript after digestion with Hind III and Sac I restriction endonucleases as expected. A λ DNA molecular size marker (*EcoR1 - Hind III* double digest of λ phage DNA) was used to determine the size of the digestion products (Lane M).

Plasmid DNA was isolated on a large scale from one of the positive clones and used for DNA sequencing studies.
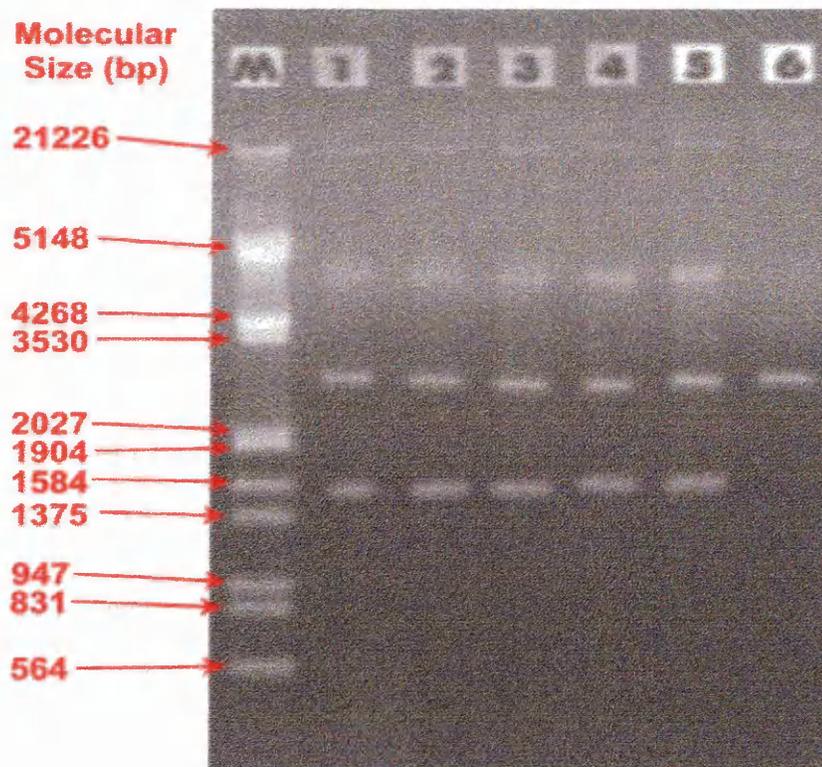
Figure 12.   Analysis of the putative clones carrying the *MBP* gene
promoter following digestion with a combination of *Hind III* and
Sac I restriction endonucleases by agarose gel (1%)
electrophoresis.

Lane M.   A λ DNA molecular size marker (*EcoR1 – Hind III* double digest
of λ phage DNA).

Lanes 1 – 5. Restriction endonuclease digestion products of the plasmid
DNA isolated from the putative positive clones carrying the
myelin basic protein gene insert.

Lane 6.   Linear pBluescript following digestion with *Hind III* and *Sac I*
restriction endonucleases.

3.2. Sequencing of the *Hind III – Sac I* restriction fragment containing the MBP promoter

Single-stranded DNA from a positive clone was prepared using alkaline denaturation followed by chromatography on Sephacryl S-400 columns. DNA obtained by this method is of high enough purity to be used directly for sequencing.

Sequencing was done manually using the chain termination method of Sanger *et al.* (1977), involving the enzymatic synthesis of a new DNA strand complementary to the template using a DNA polymerase enzyme (e.g. Sequenase). Synthesis of the new DNA strand continues in the presence of the deoxynucleoside triphosphates until the incorporation of a dideoxynucleotide, at which point the growing chains are terminated, this occurs due to the lack of a free 3' hydroxyl group on the dideoxynucleoside triphosphate. The ratio of deoxynucleoside triphosphates (dNTPs) to dideoxynucleoside triphosphates (ddNTPs) in the labelling mix is 10 : 1. Incorporation of α-^{35}S – dATP in the growing chains enabled visualisation of the sequencing products by autoradiography.

An oligonucleotide primer corresponding to the universal priming site (located within the multicloning site) of the plasmid pBluescript SK$^+$ (see plasmid map in Appendix VI) was synthesised.

Universal primer sequence	5'- TAAAACGACGGCCAGT -3'

From the sequence determined using the universal primer another

primer (corresponding to the sequence between nt -1168 to −1154 of the

MBP gene) was prepared and further sequence data obtained. The

procedure was repeated until the 1.4 kb Hind III − Sac I fragment carrying

the promoter and segment (including exon 1) of the mouse MBP gene had

been established. The sequences and locations of the primers are shown in

Table 2.

Similarly sequence data for the reverse (complementary) strand was

determined using the T3 promoter primer as the starting primer together with

two other primers, which are represented in Table 2.

T3 promoter primer sequence	5'- AATTAACCCTCACTAAAGGG -3'

An autoradiogram (Figure 13) shows part of the sequence obtained

(approximately 170 bases) of the 1.4 kb fragment containing the promoter

region and segment of the myelin basic protein gene. The 93 bases

identified on the autoradiogram are located between nucleotide positions

− 832 and − 740. Further autoradiogram showing sequencing data can be

seen in figures 14 and 15 and represent the myelin basic protein promoter

region sequence located between nucleotides −1319 and −1244 and −712

and −666. respectively.

111

Sequencing Direction	Nucleotide Position	Oligonucleotide Primer Sequences
Forward	-1168 ⇨ -1154	5'-CATGGGTGCATAGAT-3'
Forward	- 939 ⇨ - 923	5'-CTAAGGTCAGTAATAGA-3'
Forward	- 824 ⇨ - 806	5'-TGTGCTTGGTTGTGGAAGT-3'
Forward	- 590 ⇨ - 575	5'-GCTTCACATCTGGGGC-3'
Forward	- 442 ⇨ - 425	5'-AGTGGTTCTTTTATTGAG-3'
Reverse	- 244 ⇨ - 228	3'-CCGGGTGTAAGTATAG-5'
Reverse	- 66 ⇨ - 51	3'-GGGCTCTACGGGGCCC-5'

Table 2. The sequences of the oligonucleotide primers synthesised in the laboratory for determining the sequence of the *Hind III – Sac I* fragment of the mouse myelin basic protein gene promoter.
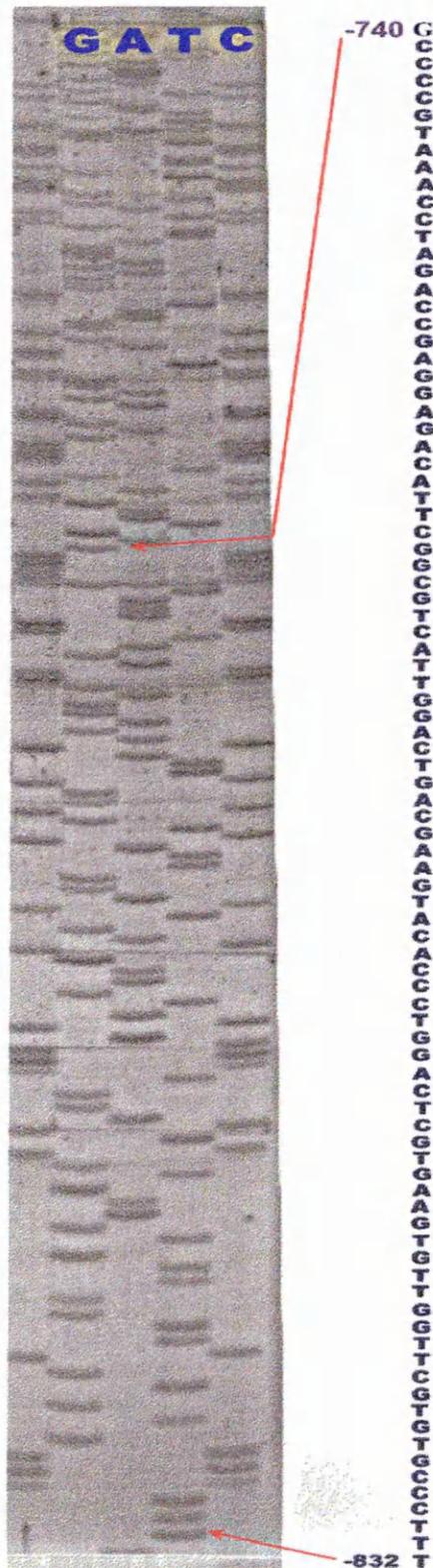
Figure 13.   An autoradiogram showing the sequence of the *MBP* gene promoter located between nucleotides −832 and −740.

113
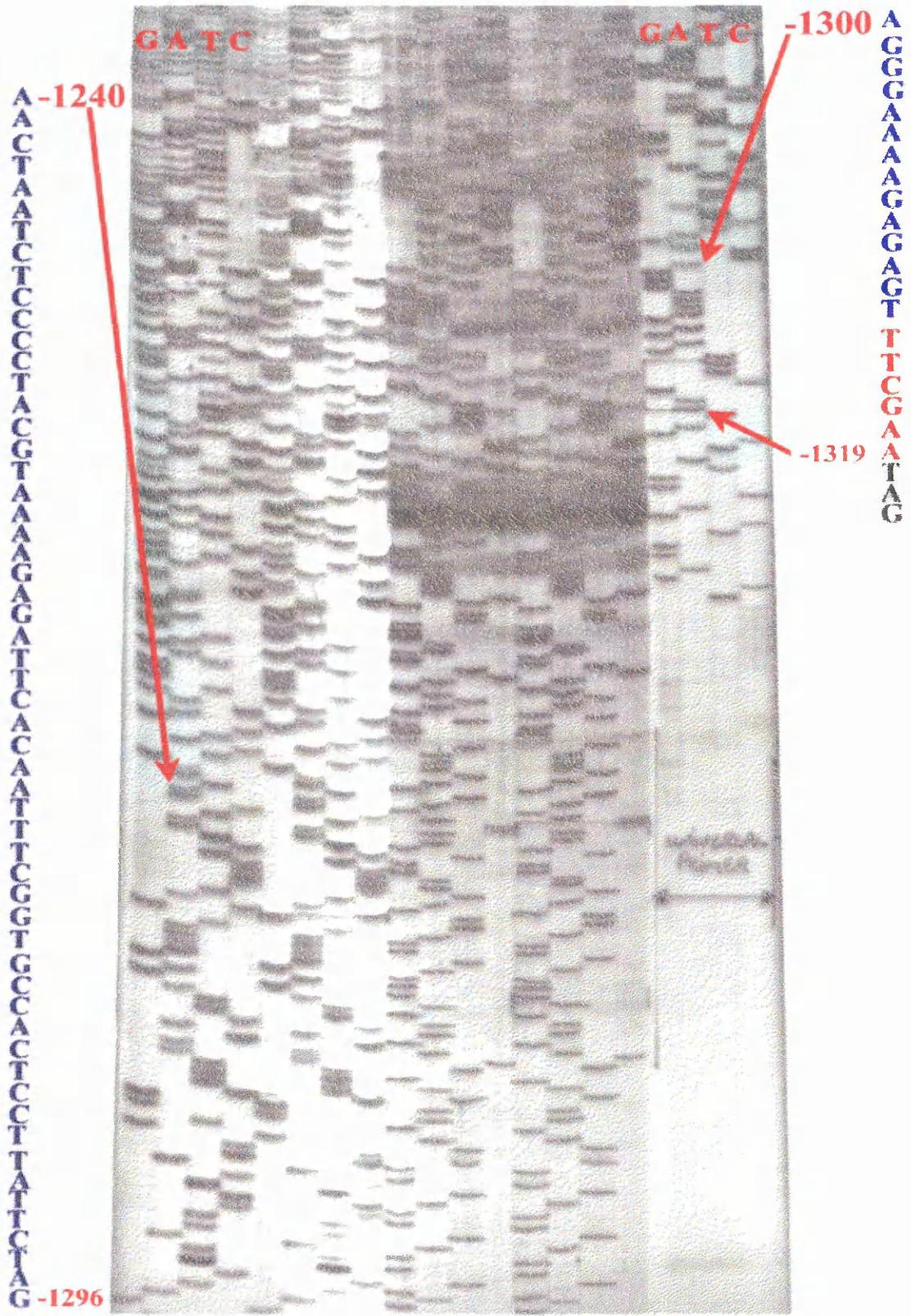
Figure 14. An autoradiogram showing the DNA sequence of the *MBP* gene promoter located between nucleotides −1319 and −1240. Duplicate samples of the sequencing products were loaded onto the gel, with the exception of the sequencing product labelled universal primer.

114
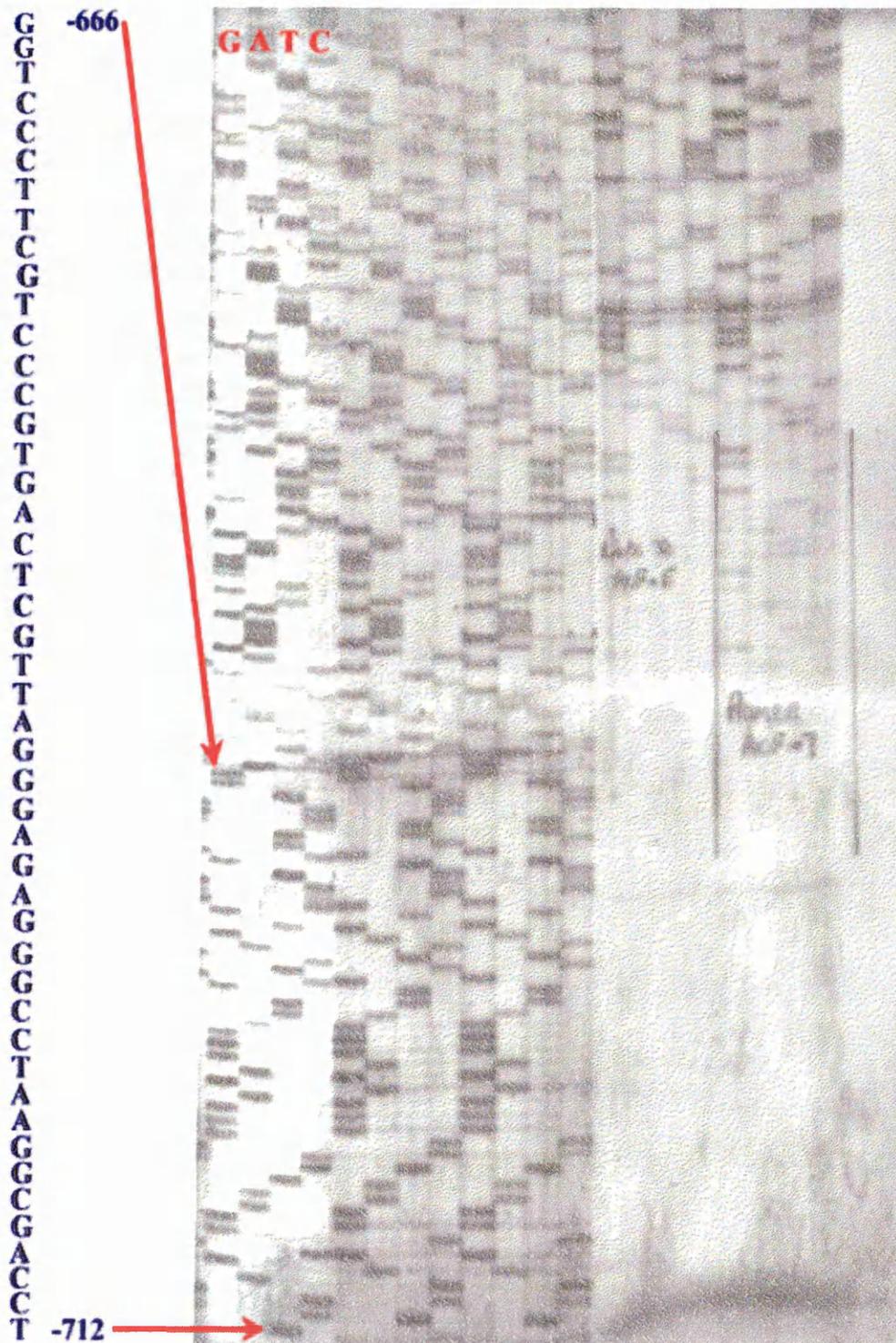
Figure 15. An autoradiogram showing the DNA sequence of the *MBP* gene promoter located between nucleotides –712 and –666. Duplicate samples of the sequencing products were loaded onto the gel.

115

The sequence of the *Hind III – Sac I* fragment (located between nucleotide positions –1319 and +227) containing the promoter region and segment (including exon 1) of the mouse myelin basic protein gene as determined from the sequencing gels is outlined as follows:

DNASIS   ***** DNA SEQUENCE FILE LIST *****

*** INPUT INFORMATION ***

FILE: **MBP_PROMOTER.SEQ**

*** SEQUENCE LIST ***   (DOUBLE STRANDED)


```
-1319       -1310       -1300       -1290
AAGCTTTGA GAGAAAGGG ACCAGATCTT ATTCCTCACC
TTCGAAACT CTCTTTT CCCTGGTCTAGAA TAAGGAGTGG


-1280       -1270       -1260       -1250
 GTGGCTTTAA CACTTAGAGA AAATGCATCC CCTCTAATCA
 CACCGAAATT GTGAATCTCT TTTACGTAGG GGAGATTAGT


-1240       -1230       -1220       -1210
 ATAAGTCATC GACAGTGGGT AGATGGAGGA ACGGCAGTCG
 TATTCAGTAG CTGTCACCCA TCTACCTCCT TGCCGTCAGC


-1200       -1190       -1180       -1170
 GTAGTAGGAT GCGTGCTAAC GATAGTCTCG TGCATGGGTG
 CATCATCCTA CGCACGATTG CTATCAGAGC ACGTACCCAC
```

116

```
-1160        -1150        -1140        -1130
CATAGAT CCG  TGGGCAGGTG  GACAAGGTGG  GGGTGGATAA
GTATCTAGGC   ACCCGTCCAC  CTGTTCCACC  CCCACCTATT


-1120        -1110        -1100        -1090
AGAAGTGGGT   AGATGATTGA  TGTTAGGTAA  ATATCACTGG
TCTTCACCCA   TCTACTAACT  ACAATCCATT  TATAGTGACC


-1080        -1070        -1060        -1050
GTGGACAGAT   GGGTGGTAGG  TGGATGGATG  GTTAGAATAG
CACCTGTCTA   CCCACCATCC  ACCTACCTAC  CAATCTTATC


-1040        -1030        -1020        -1010
TCAGAAGAGG   GATGGATTGA  TAAGGTGAAC  AGATGATAAA
AGTCTTCTCC   CTACCTAACT  ATTCCACTTG  TCTACTATTT


-1000        -990         -980         -970
TGGGTGATAG   ACTGGAAGGG  TTGTCAAAAG  AGGATAAGGG
ACCCACTATC   TGACCTTCCC  AACAGTTTTC  TCCTATTCCC


-960         -950         -940         -930
AAGTGTGAGC   TAGCCGTATT  TCTAAGGTCA  GTAATAGAGT
TTCACACTCG   ATCGGCATAA  AGATTCCAGT  CATTATCTCA


-920         -910         -900         -890
TGGGAGAAGA   GGTTAAGTTA  CATCCATTTA  AACCTCACAC
ACCCTCTTCT   CCAATTCAAT  GTAGGTAAAT  TTGGAGTGTG
```

117

```
-880        -870        -860        -850
 GAAGCTGAGT GGGAATGGAC TTGCTGCCGT TGGTGAGGAA
 CTTCGACTCA CCCTTACCTG AACGACGGCA ACCACTCCTT


-840        -830        -820        -810
 AGCGTTGCAT TTCCCGTGTG CTTGGTTGTG GAAGTGCTCA
 TCGCAACGTA AAGGGCACAC GAACCAACAC CTTCACGAGT


-800        -790        -780        -770
 GGTCCCACAT GAAGCAGTCA GGTTACTGCG GCTTACAGAG
 CCAGGGTGTA CTTCGTCAGT CCAATGACGC CGAATGTCTC


-760        -750        -740        -730
 GAGCCAGATC CAAATGCCCC GAGTAAGCAC GTCCCCGAGC
 CTCGGTCTAG GTTTACGGGG CTCATTCGTG CAGGGGCTCG


-720        -710        -700        -690
 CAGAGGCCTC CAGCGGAATC CGGGAGAGGG ATTGCTCAGT
 GTCTCCGGAG GTCGCCTTAG GCCCTCTCCC TAACGAGTCA


-680        -670        -660        -650
 GCCCTGCTTC CCTGGACTGT AAGCTGCAGA AAGATGTGGG
 CGGGACGAAG GGACCTGACA TTCGACGTCT TTCTACACCC


-640        -630        -620        -610
 AAGTCCTGTT CTCCACTGAG AACACTAAAA GCACCTTTTG
 TTCAGGACAA GAGGTGACTC TTGTGATTTT CGTGGAAAAC
```

```
-600          -590          -580          -570
TCAAACGACC GCTTCACATC TGGGGCTTGT GCACTGGTGG
AGTTTGCTGG CGAAGTGTAG ACCCCGAACA CGTGACCACC


-560          -550          -540          -530
CCTTTTAAAC CAGAGACAAC CCACAAGATA CCTAACCTGC
GGAAAATTTG GTCTCTGTTG GGTGTTCTAT GGATTGGACG


-520          -530          -520          -510
GGGGCTCTCT GGTACAGTGA GCAACTCAGG AAATGCTTTG
CCCCGAGAGA CCATGTCACT CGTTGAGTCC TTTACGAAAC


-480          -470          -460          -450
GCTTGATTGC TGTGGGCTCT CAGGCCATCG CCCTCTGGAG
CGAACTAACG ACACCCGAGA GTCCGGTAGC GGGAGACCTC


-440          -430          -420          -410
TGGTTCTTTT AATGAGAACC TGAAGATTGG CCCCTGAGCC
ACCAAGAAAA TTACTCTTGG ACTTCTAACC GGGGACTCGG


-400          -390          -380          -370
ATGTATACCA AGCAAGCTCA ATCCAGGTTA GCTCCCTCTG
TACATATGGT TCGTTCGAGT TAGGTCCAAT CGAGGGAGAC


-360          -350          -340          -330
GTTGGGGCAA GCTAACGTGC TCCTTGGGCC CCGCGCGTAA
CAACCCCGTT CGATTGCACG AGGAACCCGG GGCGCGCATT
```

119

```
      -320          -310          -300          -290
   CTGTGCGTTT  TATAGGAGAC  AGCTAGTTCA  AGACCCCAGG
   GACACGCAAA  ATATCCTCTG  TCGATCAAGT  TCTGGGGTCC


      -280          -270          -260          -250
   AAGAAAGCGG  CTTTGTCCCC  CTCTAGGCCT  CGTACAGGCC
   TTCTTTCGCC  GAAACAGGGG  GAGATCCGGA  GCATGTCCGG


      -240          -230          -220          -210
   CACATTCATA  TCTCATTGTT  GTTGCAGGGG  AGGCAGATGC
   GTGTAAGTAT AGAGTAACAA  CAACGTCCCC  TCCGTCTACG


      -200          -190          -180          -170
   GATCCAGAAC  AATGGACCT  CGGCTGAGGA  CACGGCGGTG
   CTAGGTCTTG  TTACCCTGGA  GCCGACTCCT  GTGCCGCCAC


      -160          -150          -140          -130
   ACAGACTCCA  AGCACACAGC  AGACCCAAAG  AATAACTGGC
   TGTCTGAGGT  TCGTGTGTCG  TCTGGGTTTC  TTATTGACCG


      -120          -110          -100          -90
   AAGGCGCCCA  CCCAGCTGAC  CCAGGGAACC  GCCCCCACTT
   TTCCGCGGGT  GGGTCGACTG  GGTCCCTTGG  CGGGGGTGAA


      -80           -70           -60           -50
   GATCCGCCTC  TTTTCCCGAG  ATGCCCCGGG  AAGGGAGGAC
   CTAGGCGGAG  AAAAGGGCTC  TACGGGGCCC  TTCCCTCCTG
```

120

```
        -40           -30           -20           -10
     AACACCTTCA AAGACAGGCC CTCAGAGTCC GACGAGCTTC
     TTGTGGAAGT TTCTGTCCGG GAGTCTCAGG CTGCTCGAAG


        +1            +11           +21           +31
     AGACCATCCA AGAAGACCCC ACAGCAGCTT CCGGAGGCCT
     TCTGGTAGGT TCTTCTGGGG TGTCGTCGAA GGCCTCCGGA


        +41           +51           +61           +71
     GGATGTGATG GCATCACAGA AGAGACCCTC ACAGCGATCC
     CCTACACTAC CGTAGTGTCT TCTCTGGGAG TGTCGCTAGG


        +81           +91           +101          +111
     AAGTACCTGG CCACAGCAAG TACCATGGAC CATGCCAGGC
     TTCATGGACC GGTGTCGTTC ATGGTACCTG GTACGGTCCG


        +121          +131          +141          +151
     ATGGCTTCCT CCCAAGGCAC AGAGACACGG GCATCCTTGA
     TACCGAAGGA GGGTTCCGTG TCTCTGTGCC CGTAGGAACT


        +161          +171          +181          +191
     CTCCATCGGG CGCTTCTTTA GCGGTGACAG GGGTGCGCCC
     GAGGTAGCCC GCGAAGAAAT CGCCACTGTC CCCACGCGGG


        +201          +211          +221
     AAGCGGGGCT CTGGCAAGGT GAGCTC
     TTCGCCCCGA GACCGTTCCA CTCGAG
```

121

KEY:

Red   *Hind III* (nt –1319 t0 –1314) and *Sac I* (nt +221 to +226)
     restriction endonuclease sites respectively.

Green   Restriction endonuclease sites.

Acc I	GT/ATAC	-398 to -393
Alw44 I	G/TGCAC	-572 to -567
Bgl II	A/GATCT	-1297 to -1292
Dra I	TTT/AAA	-894 to -889
Hae III	GG/CCA	+89 to +93
Hsp II	CATG/	+102 to +108
Sma I	CCC/GGG	-56 to -51
Stu I	AGG/CCT	-717 to -712
Xma I	C/CCGGG	-56 to -51

Blue   Primer sequences as shown in table 2.

Yellow Identified protein binding sites as shown in table 3.

The data obtained from the sequencing gels representing the promoter region of the mouse myelin basic protein gene, located between nucleotides −1319 and +227, was used as a query or probe sequence to carry out a FastA sequence alignment search against the EMBL rodent database. A significant alignment was found with a sequence published by Miura et al., (1989 and updated March 2000), which showed that the two sequences were essentially identical (99.6%).

However, there are three minor discrepancies with the published database sequence. These discrepancies are located within the 5' untranslated region at nucleotide positions −1202 to −1201, −1181 to −1180 and −1152 to −1151 respectively relative to the transcriptional start site. Miura *et al.*, (1989) shows the sequence at these positions to be GC yet the manual sequencing gels indicate the sequence to be CG. It is possible that these differences could be dinucleotide polymorphisms, whereby the definition of a polymorphism is applied in its loosest sense i.e. any non-pathogenic sequence variant.

Transcription of many genes begins by the binding of RNA

polymerase II, mediated by the TFII group of ubiquitous transcription factors,

to the 'TATA' sequence located at a position upstream of the transcriptional

initiation site. The MBP promoter, however, does not contain typical TATA

(consensus <u>TATA</u>(A/T)A(A/T)), CAAT (consensus GG<u>CCAAT</u>CT) and GC

(consensus GGGCGG) boxes, but contains TATA (TTCAAA) and CAAT

(CACTT) box-like sequences located at positions –34 and –85 respectively.

It is thought that the myelin basic protein promoter may utilise specific TATA

and CAAT box factors (Tamura *et al.* 1989). A GC box-like sequence

(CCGCCC) has also been identified at position –92 to which a specific Sp1-

like transcription factor binds (Tamura *et al.* 1989).

From database searches, published papers by Miura *et al.* (1989)

and Tamura *et al.* (1990) and experimental approaches a number of

important DNA-binding sites for transcription factors have been identified in

the mouse myelin basic protein gene promoter (a summary of the positions

of these binding sites is shown in Table 3).

It has been suggested (Aoyama *et al.* 1990) that the mouse myelin

basic protein gene promoter from nucleotide positions –253 to +62 is

responsible for efficient and tissue-specific transcription, and that this region

can be further divided into the proximal element (–53 to +62 containing the

TATA box-like element at –34) which is required for basal transcription and

the distal element (–253 to –54) which is responsible for brain specificity.

Nucleotide Position	Nucleotide Sequence	Comments	Reference
-34 to –28	5'-TTCAAA-3'	TATA-like sequence	Miura *et al.* (1989)
-85 to –81	5'-CACTT-3'	CAAT box like sequence	Miura *et al.* (1989)
-92 to –87	5'-CCGCCC-3'	GC box	Miura *et al.* (1989)
-130 to -93	5'- AATAACTGGCAAGGCGCCCAC C CAGCTGACCCAGGGAA-3'	Myelin regulation factor-1 (MRF-1)	Haque *et al.* (1994)
-125 to –111	5'-CTGGCAAGGCGCCCA-3'	NFI-like binding site	Miura *et al.* (1989)
-132 to -125	5'-CAAAGAATAAC-3'	MEBA	Taveggia *et al.* (1998)
–160 to –127	5'-ACAGACTCCAAGCACACA GCAGACCCAAAGAATA-3'	Box I and Box II motif	Torron *et al.* (1997)
-183 to -168	5'-ACCTCGGCTGAGGACACG-3'	Thyroid hormone response element	Farsetti *et al.* (1992)
-648 to –635	5'-GATGTGGGAAGTCC-3'	SV40-like enhancer core	Miura *et al.* (1989)
-675 to -647	5' GCTTCCCTGGACTGTAA GCTGCAGAAAGA –3'	Fyn response element	Umemori *et al.* (1999)

Table 3: Important protein-binding sites identified in the sequence of the *Hind III – Sac I* fragment of the mouse myelin basic protein promoter region and segment (including exon 1).

Myelin gene expression factor-2 (MyEF-2) is a protein of 436 amino acids isolated from mouse brain that binds specifically to the noncoding strand of the proximal regulatory sequence known as MB1 (located between nt –50 to –14 relative to the transcription initiation site) and represses transcription of the *MBP* gene in transient transfection assay (Haas *et al.* 1995). MyEF-2 contains the RNA recognition motif (RRM)-type nucleic acid binding motif common to a growing family of proteins that bind single-stranded DNA and RNA in a sequence-specific manner (Kenan *et al.* 1991). MyEF-2 mRNA is developmentally regulated in mouse brain with peak expression occurring at post-natal day 7, before the onset of myelin basic protein expression. MyEF-2 mRNA expression follows that described for SCIP (also called Oct-6 or Tst-1), a POU domain transcription factor that represses several myelin – specific genes notably myelin basic protein and P_0 glycoprotein expression. The MyEF-2 gene maps to mouse chromosome 2 (Haas *et al.*, 1995). Muralidharan *et al.* (1997) using a bandshift assay with purified MEF-1/Purα and MyEF-2 demonstrated that the binding of MyEF-2 to its target sequence is inhibited by MEF-1/Purα. Under similar conditions, MyEF-2 enhances the association of MEF-1/Purα with MB1 DNA. MEF-1/Purα binds to MB1 in mono- and dimeric forms, and inclusion of MyEF-2 in the binding reaction increases the dimeric association of MEF-1/Purα with the MB1 sequence. The use of MEF-1/Purα variants in the bandshift assay suggested that two distinct regions of this protein may be involved in its binding to the MB1 regulatory sequence, and its ability to block MyEF-2 interaction with the MB1 sequence.

Steplewski *et al.* (1998) reported the isolation of a recombinant cDNA

clone, termed myelin expression factor-3 (MyEF-3) from a mouse brain

expression library encoding a novel protein interacting with the MBP

regulatory domain (MB1) located between nt –50 to –14 relative to the

transcription initiation site. The MyEF-3 sequence revealed several

interesting features including four sites for phosphorylation by casein kinase

II, a transmembrane domain at the N-terminus, a nuclear localisation signal

and a zinc finger domain at the carboxyl terminus (Steplewski *et al.* 1998).

The group also showed that MyEF-3 binds as efficiently to the double-

stranded regulatory region as it does to the single-stranded non-coding

strand. Northern blot analysis studies demonstrate that expression of MyEF-

3 is restricted to brain and is developmentally regulated during brain

maturation.

Functional dissection of the regulatory sequence has led to the

identification of a regulatory motif spanning nucleotides –14 to –50 (named

MB1) with respect to the transcription start site (Tretiakova *et al.* 1998).

Analysis of nuclear proteins derived from mouse brain identified a 39 kDa

protein that forms a nucleoprotein complex with a DNA fragment containing

the same sequence. Subsequent studies (Tretiakova *et al.*, 1998) showed

the 39 kDa protein, named Purα, recognised the GGC/GGA – rich

sequences within the MB1 DNA in a single-stranded configuration and has

the ability both *in vitro* and *in vivo* to increase transcription of the myelin

basic protein promoter. Purα binding activity to the MB1 regulatory sequence

occurs in a developmental stage-specific manner coinciding with the pattern

of MBP transcription. At the early stage of brain development (days P3 to

P7) Purα levels associated with MB1 are extremely low, however, during the phase of myelination (days P18 to P20) and in adults (day 30) Purα binding activity to MB1 dramatically rises. Because a significant amount of Purα is detected at an early stage and the association with the MB1 DNA sequence is low, then the implication(s) of a cofactor(s) may determine Purα binding activity to the DNA molecule. The ability of Purα to interact with a GGN repeat in a single-stranded form prompted the examination of its binding ability to RNA molecules. Tretiakova *et al.* (1998) demonstrated that Purα obtained from brain extract is associated with RNA molecules that have the ability to inhibit its binding activity to the MB1 DNA. This RNA named PU-RNA, has significant homology to the 7SL RNA, is expressed during various stages of brain development, and is found in association with Purα in 5 day-old mouse brain. Therefore, PU-RNA functions as a cofactor by determining the binding activity of Purα to the MBP promoter sequence and indirectly participate in the developmental activation of the MBP promoter in mouse brain.

From the earlier study outlined above, Tretiakova *et al.* (1999) have now found evidence for the binding of the ubiquitous transcription factor (Sp1) to the MB1 DNA motif (spanning nucleotides −14 to −50) that partially overlaps the Puralpha binding site. They demonstrated that the binding of Puralpha to its target sequence is enhanced by the inclusion of Sp1 in the binding reaction. Under this condition binding of Sp1 to the MB1 regulatory sequence remained unchanged and there was no evidence for the formation of a Puralpha:MB1:Sp1 complex. This observation suggested that transient

interaction of Puralpha and Sp1 might result in the stable association of Purα and the MB1 element. Results from immunoprecipitation/Western blot studies have established an association of Purα and Sp1 in nuclear extracts from mouse brain. Purα appears to bind to the phosphorylated form of Sp1, which is developmentally regulated and coincides with the periods when *MBP* gene expression is at its maximum level. Results from cotransfection studies revealed that ectopic expression of Purα and Sp1 synergistically stimulates MBP promoter activity in CNS cells (Tretiakova *et al.* 1999).

Haque *et al.* (1994) have reported the isolation of a recombinant clone, which they named myelin regulatory factor-1 (MRF-1), from a mouse brain expression library that encodes a novel protein interacting with the distal regulatory domain (MB3) located between nucleotide positions −130 to −93. Analysis of MRF-1 by Haque *et al.* revealed substantial sequence homology in the central and the COOH-terminal regions of the protein with the splicing factor SC35. Cotransfection studies carried out by the group indicated that MRF-1 increases transcription of the MBP promoter in glial cells and that the activation requires an intact MRF-1-binding site. The gene encoding MRF-1 is located on the distal half of mouse chromosome 11 in a region where the human homologue would be predicted to reside on human chromosome 17.

Taveggia *et al.*, (1998) describe a DNA sequence flanking and partially overlapping the NF1 site, which activates the MBP promoter in oligodendrocytes. This sequence is bound by a unique set of nuclear proteins enriched in the brain, oligodendrocytes, and Schwann cells, designated MEBA (*m*yelinating glia-enriched *b*inding *a*ctivity). The

129

appearance of MEBA parallels both myelination in the developing brain and myelin basic protein mRNA expression in differentiating oligodendrocytes. Mutations within the region of MEBA and NF1 binding distinguish MEBA from NF1 and suggest that MEBA consists of at least two proteins. Because these binding sites with opposing functions overlap, Taveggia *et al.*, (1998) suggested that MEBA might either compete with or alter the NF1 binding of the proximal MBP promoter and derepress the MBP promoter in oligodendrocytes.

The box 1 and box 2 motif (Torron *et al.* 1997) of the myelin basic protein promoter is composed of two DNA sequences 13 and 12 bp long respectively, separated from each other by nine poorly conserved base pairs, and are located at −160 to −127 bp from the transcription initiation site in mouse. The box 1 and 2 motif is flanked at its 5'-end by a thyroid hormone receptor element (at nucleotides −183 to −168, Farsetti *et al.* 1992) and at its 3'-end by a nuclear factor 1 (NF1) related site (at nucleotides −130 to −111, Aoyama *et al.* 1990 and Zhang and Miskimins, 1993). The evolutionary conservation, physical location, and structural features of the box 1 and 2 motif suggest that this is a regulatory sequence of the MBP transcription unit (Fors *et al.* 1993). The primary structure of box 1 appears to be related to that of box 2, since both begin with the sequence GAC and contain the sequence AA near their 3'-ends. The nine non-conserved base pairs separating box 1 and 2 comprise almost one full-turn of the double helix in the B-DNA conformation, and could therefore orientate the last base pair of box 1 and the first base pair of box 2 towards the same face of the molecule. This molecular configuration may serve as a specific recognition site for the

binding of transcription regulatory proteins. Torron *et al.* (1997) used a DNA fragment containing the sequence of the motif to identify DNA binding proteins. Two proteins MBP32 and MBP38 (of minimum $M_r \sim 32$ and ~ 38 kDa) that are found in nuclear extracts from mouse brain and oligodendrocyte-like glioma cells bind to the box 1 and box 2 motif. MBP32 and MBP38 are also found in other cell types and tissues expressing the *golli* transcription unit. Therefore, Torron *et al.*, (1997) propose that these proteins may participate in the modulation of the MBP transcription unit in the context of an active *golli* unit since exon 5a of the *golli*-MBP complex also contains several other regulatory sequences of the MBP promoter, including the TRE, the NF1-related, M1 and MB1 sites, and the GC, CAAT, and TATA-like boxes.

Umemori *et al.*, (1999) have shown that Fyn protein tyrosine kinase signalling is involved in the initial events of myelination. Fyn, which is present in myelin-forming cells, is a member of the Src-family nonreceptor protein tyrosine kinases (PTKs) and is activated by signal transduction through the stimulation of cell surface receptors such as the large myelin-associated glycoprotein (L-MAG) to Fyn protein tyrosine kinase (Fyn PTK). Fyn protein tyrosine kinase then phosphorylates its substrates, which in turn stimulates transcription factors, resulting in the stimulation of *MBP* gene transcription necessary for myelination. The group showed that the sequence between nucleotides –675 and –647 relative to the transcription start site included the Fyn response element to which transcription factors stimulated by Fyn interact. Proteins binding to the region between nucleotides –675 and –647 were found by EMSA, and the binding activity

131

correlated with Fyn activity during myelination thereby suggesting the stage-specific regulation of transcription. In a study using 4-week-old Fyn-deficient mice Umemori *et al.*, (1999) found that the amount of MBP was significantly reduced (by at least 55%), and subsequent electron microscopic analysis showed that the myelin was approximately 70% thinner and more irregular than in wild-type mice. Umemori *et al.* (1999) also searched for previously known consensus transcription factor response elements and showed that the Fyn response region contains the interferon-γ response element core sequence (consensus C(A/T)(G/T)(G/T)ANN(C/T)) and the NF-IL6 core sequence (consensus T(G/T)NNGNAA).

The differentiation of oligodendrocyte and Schwann cell progenitors into mature myelin producing cells is affected by different growth factors and hormones. To date not all these factors have been defined, however, it is evident that increased cyclic AMP levels accelerate the expression of a differentiated phenotype including the increased expression of myelin components such as MBP (Mirsky *et al.* 1980 and Raible *et al.* 1989). It still remains to be elucidated what the endogenous factor is that increases cyclic AMP during differentiation and what signalling pathway(s) lead to differentiation (Clark *et al.* 1998). Using the D6P2T cell line the group investigated the involvement of mitogen-activated protein kinase (MAP kinase) in the signal transduction cascade initiated by cyclic AMP, which led to the increased expression of the *MBP* gene. MAP kinases can be grouped into three families, which include the extracellular signal-regulated kinases, Jun N-terminal kinase/stress-activated protein kinases and p38. Activation of MAP kinases is a rapid cellular response to differentiation and growth factors

(Marais *et al.* 1996 and Moriguchi *et al.* 1996). Phosphorylation of threonine and tyrosine residues by dual specificity enzymes like MAP kinase kinase (MEK) is the means by which MAP kinases are activated and function within the kinase cascade. The D6P2T cell line can be induced to display a differentiated phenotype characterised by *MBP* gene expression in response to increased cyclic AMP, similar to the response of oligodendrocytes and Schwann cells. Clark *et al.* (1998) have demonstrated that blocking MAP kinase activity with inhibitors of the activating kinase (MEK) blocked the activation of the MBP promoter in D6P2T cells. In addition, the group showed that blocking MAP kinase activation during differentiation of an oligodendrocyte-like cell line (CG4) also led to the inhibition of *MBP* gene expression. Clark *et al.* (1998) suggest that these findings imply a role for MAP kinase in the signal transduction cascade initiated by cyclic AMP and leading to the expression of the *MBP* gene during differentiation.

3.3. Analysis of DNA – protein complexes by electrophoretic mobility shift assay (EMSA) and competition binding assay

Nuclear extracts from rat brains (RBX), glial cell lines (C_6 glioma and 33B oligodendrocytoma) and nonglial cell line (HeLa) were prepared by Dounce homogenisation as described in section 5.2. The protein concentration in the nuclear extracts was determined before their application in the electrophoretic mobility shift assay and competition binding assay.

A microtitre plate method utilising BCA reagent was used and albumin served as the protein standard. Data relating to the determination of the protein concentration in each of the nuclear extracts can be found in Appendix VII.

From a culture typically containing 1.0×10^6 cells, 200 µl of nuclear extract was prepared. This produced a protein yield, from the C_6 glioma, HeLa and 33B oligodendrocytoma cell nuclear extracts, ranging from 200 to 260 µg (1.0 to 1.3 µg/µl), whereas, the rat brain nuclear extract (RBX) gave a much higher protein yield of 600 µg (3.0 µg/µl) as expected.

For both the binding assay and the competition binding assay 10 µg of nuclear extract per reaction was used.

Two oligonucleotides corresponding to the putative protein-binding site located between nucleotide positions –609 to –577 were synthesised, the first corresponding to the sense strand of the binding site,

5'-CACCTTTTGTCAAACGACCGCTTCACATCTGGG-3'

and the second, to the complementary strand

3'-GTGGAAAACAGTTTGCTGGCGAAGTGTAGACCC-5'.

This region (located between nucleotides –609 and –577) was chosen for further investigation because work published by other groups (Asipu and Blair 1994) suggested that it contains a cell-type specific transcription activation site for oligodendrocytes.

The oligonucleotides were annealed as described in section 5.1.1, and an aliquot (10 µl containing approximately 1.5 µg of annealed oligonucleotide) analysed by non-denaturing polyacrylamide gel electrophoresis. The difference in the mobility between the annealed (Figure 16 lane 1) and single stranded oligonucleotide (lane 2) can clearly be seen, the annealed species being double-stranded does not migrate as far as the control sample in lane 2.

The double stranded oligonucleotide was 5' end-labelled with (100µCi) of γ-^{32}P ATP to produce a probe with a specific activity of 10 X 10^4 counts per minute (cpm) / µg of DNA for use in the electrophoretic mobility shift assay and competition binding assay.

For the competition binding assay a 100 - fold excess of unlabelled annealed oligonucleotide was used as a specific competitor.
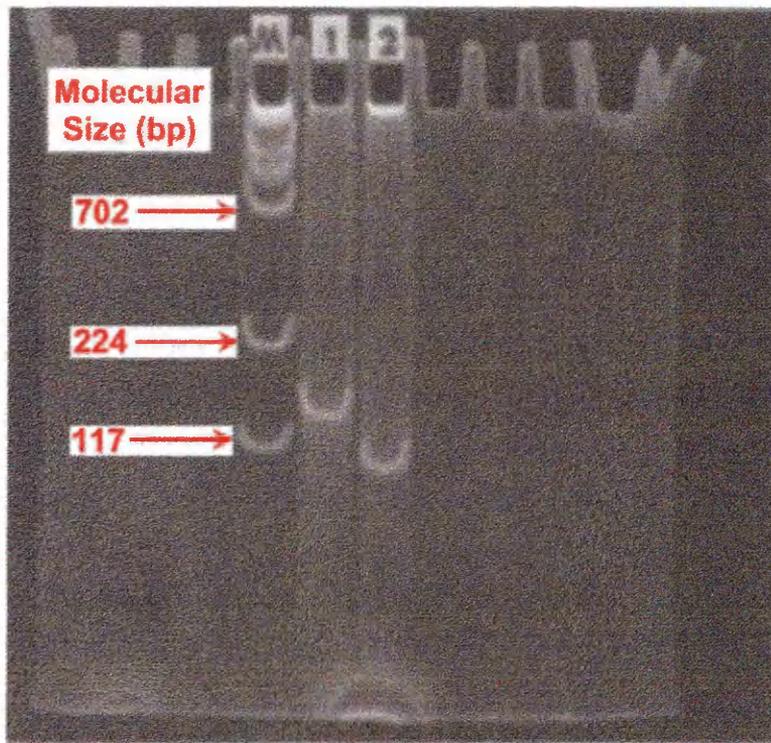
Figure 16. Analysis by non-denaturing polyacrylamide gel (15%)
electrophoresis of the products of annealing of the strands
containing a putative transcription factor binding site.

Lane M. A λ DNA molecular size marker (*BstE II* digest of λ phage DNA).

Lane 1. Annealed oligonucleotide (1.5 μg) corresponding to the putative

protein-binding site.

Lane 2. Single stranded (unannealed) oligonucleotide used as a control

sample (5'-CACCTTTTGTCAAACGACCGCTTCACATCTGGG-3').

The competition binding assay was carried out essentially as the binding assay except that unlabelled oligonucleotide probe (100 - fold molar excess) was used as a specific competitor. In both cases the binding assay and competition binding assay used 1 μg of poly[d(IC).d(IC)] as a non-specific competitor.

The resulting DNA-protein complexes were separated on 7% native polyacrylamide gels. The positions of the complexes are denoted from C1 to C4. The fastest migrating bands at the bottom of the gel correspond to the free probe (labelled oligonucleotide). Behind the free probe is a group of non-specific complexes.

The C$_6$ glioma cell nuclear extract (Figures 17 to 19 lanes 1a and 1b) show three DNA-protein complexes. C1 the slowest, C2 the intermediate and C4 with the fastest mobility. The complex represented by C2 shows high sequence-specificity since it is sensitive to competition by unlabelled oligonucleotide probe and is not present in the competition binding assay (lanes 1b).

Only two DNA-protein complexes C1 and C3 were detected in the HeLa cell nuclear extract (Figure 17, lanes 2a and 2b and Figures 18 and 19 lanes 3a and 3b). Neither of these complexes was specific for the putative binding site.

RBX nuclear extract (Figure 17, lanes 3a and 3b and Figures 18 and 19 lanes 2a and 2b) revealed three complexes (C1, C2 and C3). Complex C2 appears to be the same mobility and possibly the same protein as that of complex C2 in the C$_6$ glioma cell line nuclear extract. The competition binding assay results show this complex to be a sequence specific binding

137

protein to the mouse myelin basic protein gene promoter. C3 appears to be of the same mobility as the corresponding C3 band in the HeLa cell nuclear extract and its binding is not sequence specific.

Two complexes C1 and C3 were detected in the nuclear extract of the 33B oligodendrocytoma cells (Figure 17, 18 and 19 lanes 4a and 4b). Both complexes were of comparable mobility to the C1 and C3 complexes seen in the HeLa and RBX nuclear extracts and neither was a sequence specific binding protein to the mouse myelin basic protein gene promoter.

The binding assay and competition binding assay were repeated using a binding buffer containing magnesium sulphate (at a final concentration of 10 μM) instead of zinc sulphate (at a final concentration of 10 μM). Four DNA-protein complexes were formed between the proteins of the nuclear extracts and the putative DNA binding site, these complexes were denoted as C1 to C4 on the autoradiograms shown in Figures 20 and 21.

Only complex C2 appears to be a sequence-specific binding protein to the mouse myelin basic protein gene promoter since this protein is readily competed out by a 100 fold molar excess of the unlabelled putative binding site in both the C_6 glioma cell line and rat brain nuclear extracts (Figure 20 lanes 1a, 1b and 3a, 3b and Figure 21 lanes 1a, 1b and 2a, 2b respectively). Since complex C2 has the same mobility in both nuclear extracts it would appear to suggest that the identity of protein in C2 is the same. Protein complexes C1, C3 and C4 were not sequence specific binding proteins to the mouse *MBP* gene promoter.

138
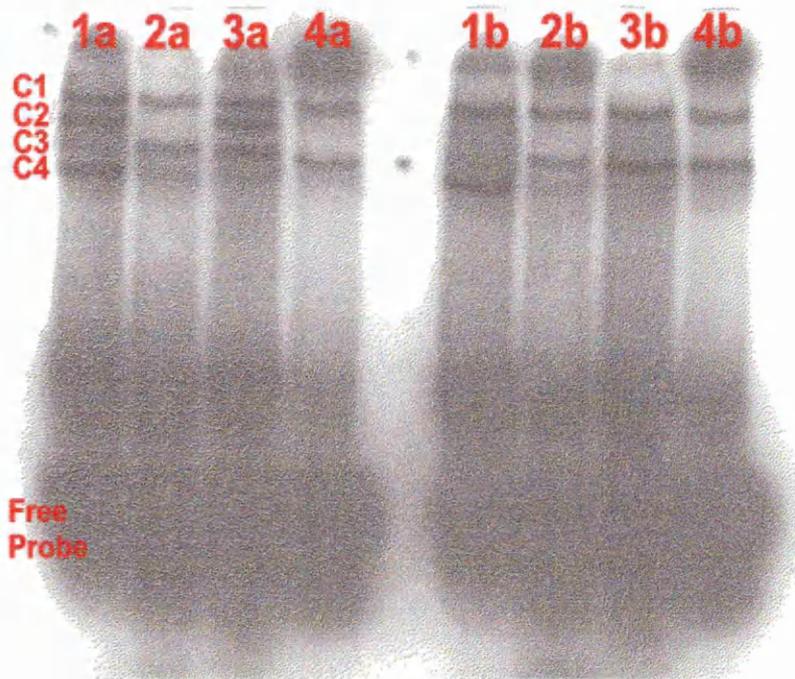
Figure 17.   Gel retardation and competition binding assay analysis of
             protein binding to region located between nucleotides −609 to
             −577 of the mouse *MBP* gene promoter.

Lanes 1a and 1b. C_6 glioma cell nuclear extract.

Lanes 2a and 2b. HeLa cell nuclear extract.

Lanes 3a and 3b. Rat brain extract (RBX) nuclear extract.

Lanes 4a and 4b. 33B oligodendrocytoma cell nuclear extract.

The band labelled free probe represents the labelled oligonucleotide probe
corresponding to the putative binding site.
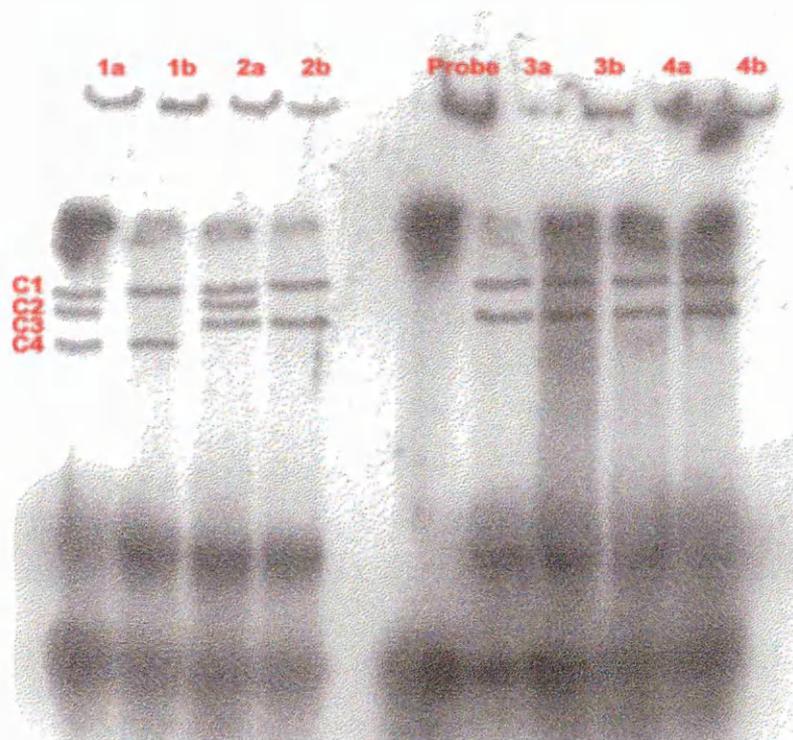
Figure 18.   Analysis of proteins binding to the *MBP* gene promoter
(between nucleotides –609 to –577) by EMSA and competition
binding assay followed by non-denaturing polyacrylamide gel
(7%) electrophoresis.

Lanes 1a and 1b. C $_6$ glioma cell nuclear extract.

Lanes 2a and 2b. Rat brain extract (RBX) nuclear extract.

Lane   Probe.   Labelled oligonucleotide probe.

Lanes 3a and 3b. HeLa cell nuclear extract.

Lanes 4a and 4b. 33B oligodendrocytoma cell nuclear extract.
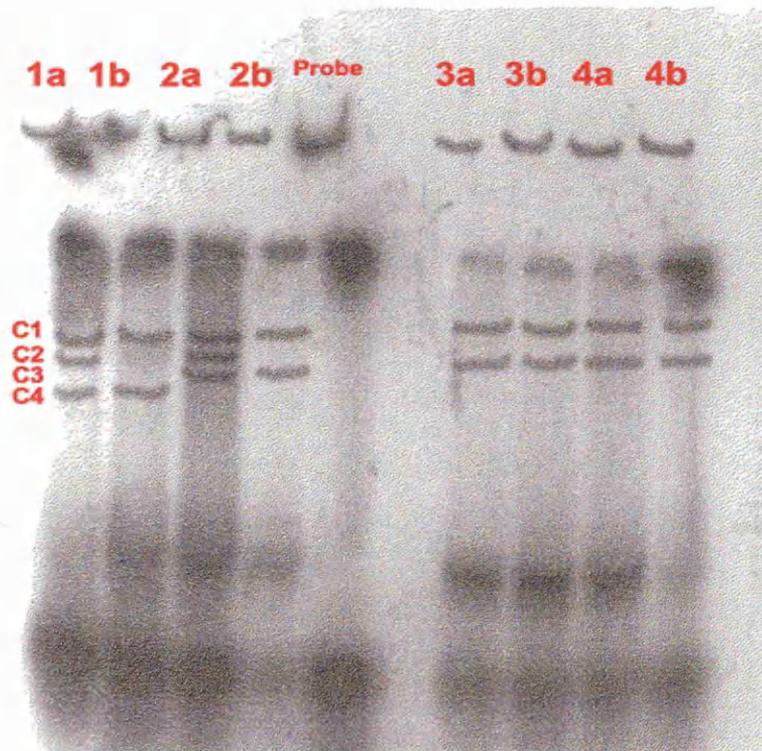
Figure 19. Electrophoretic mobility shift assay and competition binding assay analysis of proteins binding to the region of the *MBP* gene promoter located between nucleotides –609 to -577.

Lanes 1a and 1b. C $_6$ glioma cell nuclear extract.

Lanes 2a and 2b. Rat brain extract (RBX) nuclear extract.

Lane   Probe.   Labelled oligonucleotide probe.

Lanes 3a and 3b. HeLa cell nuclear extract.

Lanes 4a and 4b. 33B oligodendrocytoma cell nuclear extract.
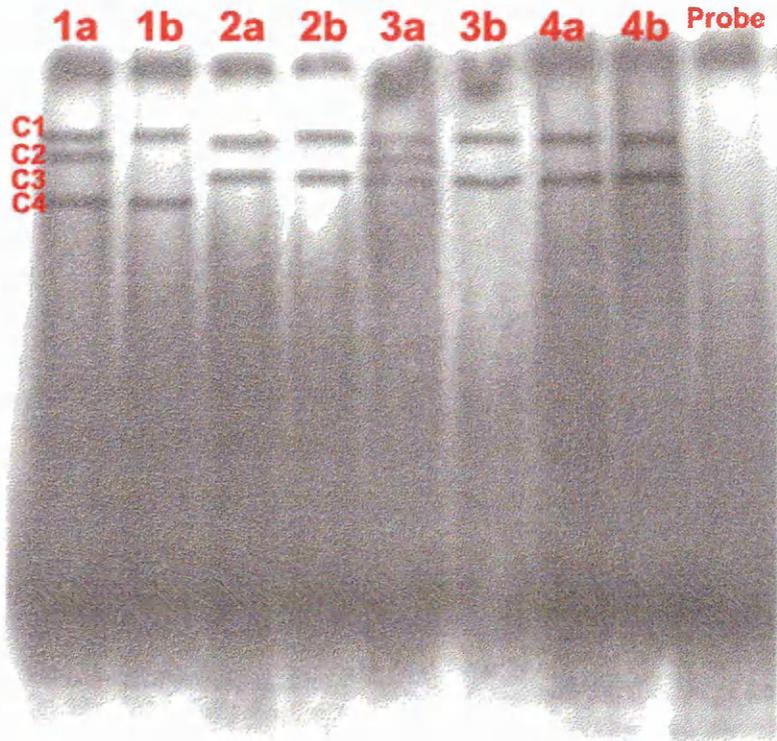
Figure 20.   Analysis of proteins binding to the –609 to –577 region of the
             MBP gene promoter by binding and competition binding assay
             followed by non-denaturing polyacrylamide gel (7%)
             electrophoresis.

Lanes 1a and 1b.   C$_6$ glioma cell nuclear extract.

Lanes 2a and 2b.   HeLa cell nuclear extract.

Lanes 3a and 3b.   Rat brain extract (RBX) nuclear extract.

Lanes 4a and 4b.   33B oligodendrocytoma cell nuclear extract.

Lane   Probe.   Labelled oligonucleotide probe.
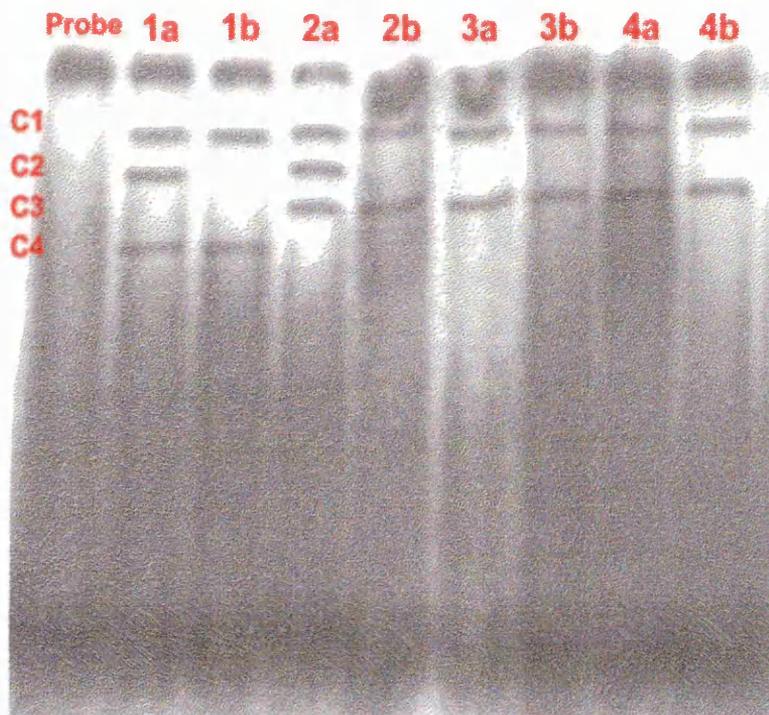
Figure 21.   Gel retardation and competition binding assay analysis of
              protein complexes binding to the region –609 to –577 of the
              MBP gene promoter after non-denaturing polyacrylamide gel
              (7%) electrophoresis.

Lane   Probe.   Labelled oligonucleotide probe.

Lanes 1a and 1b. C $_6$ glioma cell nuclear extract.

Lanes 2a and 2b. Rat brain extract (RBX) nuclear extract.

Lanes 3a and 3b. HeLa cell nuclear extract.

Lanes 4a and 4b. 33B oligodendrocytoma cell nuclear extract.

3.4. Isolation of cDNA(s) encoding proteins binding to the region located between nucleotides - 609 to - 577 of the mouse myelin basic protein gene promoter by a yeast one – hybrid system

3.4.1. Creation of the reporter vector

Complementary oligonucleotides carrying four repeat sequences of the putative binding site (–609nt to –577nt) were synthesised and annealed as described previously. The 'sense' oligonucleotide was synthesised with an *EcoR1* overhang at the 5' end, whereas the complementary oligonucleotide had an *Mlu1* overhang at its 5' end. These restriction endonuclease sites were used to directionally clone the binding site into the multicloning site of the pHISi-1 plasmid (see Appendix VI).

```
5'-AATTCCACCTTTTGTCAAACGACCGCTTCACATCTGGGCACCTTTTG
   3'-GTGGAAAACAGTTTGCTGGCGAAGTGTAGACCCGTGGAAAAC

TCAAACGACCGCTTCACATCTGGGCACCTTTTGTCAAACGACCGCTTC
AGTTTGCTGGCGAAGTGTAGACCCGTGGAAAACAGTTTGCTGGCGAAG

ACATCTGGGCACCTTTTGTCAAACGACCGCTTCACATCTGGG  -3'
TGTAGACCCGTGGAAAACAGTTTGCTGGCGAAGTGTAGACCCACGCG  -5'
```

For this purpose pHISi-1 was digested with *EcoR1* and *Mlu1* restriction endonucleases. After the enzymes were inactivated by treatment with phenol followed by ethanol precipitation, the digested vector and the double stranded oligonucleotide were ligated as described in section 2.6.3. Competent cells of *E.coli* strain DH5α were prepared by treating cells, grown to mid-log phase, with ice cold calcium chloride. The ligation mixture was precipitated with ethanol and used in the transformation of competent cells. A heat shock at 42° C followed by an incubation on ice was performed and the bacterial cells were spread over the surface of L-amp plates. After

overnight incubation at 37° C samples of the bacterial colonies were removed from the plates and analysed by PCR (using the HIS-RP and HIS-5372 primers) to confirm the presence of the binding site in the reporter plasmid.

Three types of PCR product were obtained as the electrophoretogram (figure 22) shows. Some of the bacterial colonies chosen for screening show two PCR products (lanes 6 and 19), this suggests that these colonies are probably mixtures of cells containing the reporter plasmid and the vector pHISi-I. The smaller band corresponds to a size of 117 bp (lanes 2, 3, 5, 7, 9, 11, 14,and 17) and is probably the amplification product corresponding to the particular region in the original pHISi-I. The putative positive clones identified in lanes 1,4,8,10,12, 13,15, 16, 18 and 20 show an amplification product corresponding to a size of 255 bp which represents the size of the region with the binding site (138 bp) and the flanking region in the original pHISi-I plasmid (117 bp). Two representative clones (lanes 4 and 16) were selected for large-scale isolation of the reporter plasmid containing the four tandem copies of the putative binding site for use in the One-hybrid assay system.

A PCR control (negative control) was also run (lane 22); this contains all the components of the reaction mixture except a DNA template and is used to check for reagent contamination.

A size marker was also included on the gel so that the sizes of the PCR products could be determined and compared to the expected products determined from the plasmid sequence data (the sequence of plasmid pHISi-I can be found in Appendix VI).
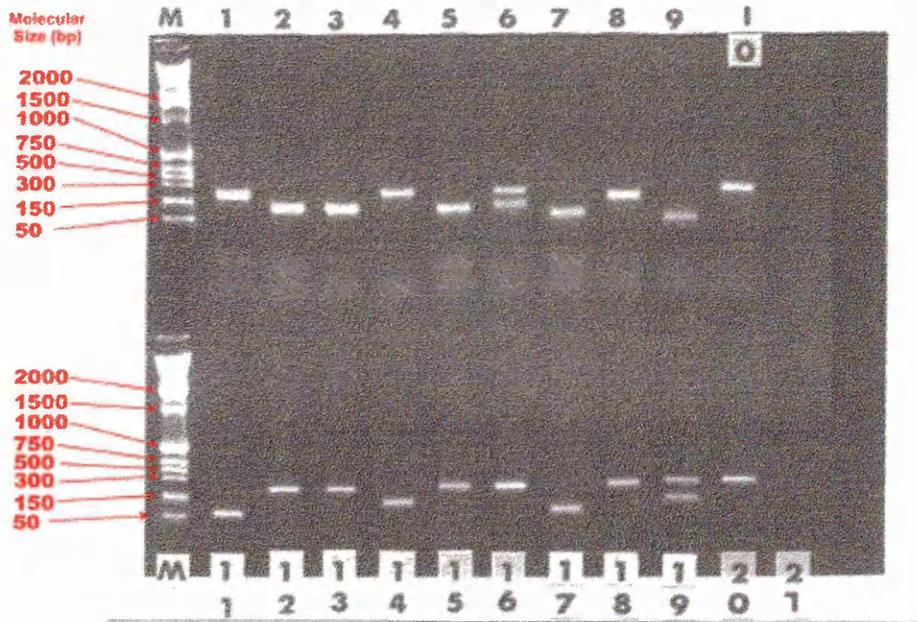
Figure 22. Screening of bacterial colonies by PCR and agarose gel (2%)
electrophoresis following transformation into *E.coli* strain DH5α of
the ligation product pHISi-1 containing the putative binding site.

Lane M. A DNA molecular size marker (1 kb PCR ladder).

Lanes 1,4,8,10,12,13,15,16,18 & 20. show the 255 bp PCR product
corresponding to the four repeat copies of the putative protein
binding site and flanking region of plasmid pHISi-I.

Lane 21 was used to run the PCR control (NEGATIVE CONTROL).

Lanes 2,3,5,7,9,11,14,and 17 show the PCR product (117 bp) corresponding
to the religation of the vector without the binding site insert.

Lanes 6 and 19 show two PCR products.

146

Reporter plasmid DNA was isolated on a large scale from one of the putative positive clones (clone 4) using the method outlined in section 7.0.

An aliquot (1 μg of plasmid DNA) was analysed by agarose gel electrophoresis (Figure 23). In both, the reporter plasmid DNA containing the protein binding site which was isolated from clone 4 (Lane 1), and the plasmid vector pHISi-1 DNA used as a control sample (lane 2) there are two bands, these represent the supercoiled (lower band) and the open circular forms of the plasmid (upper band). The difference in size between the two plasmid DNA samples resulting from the cloning of four copies of the protein binding site (138 bp) into the *EcoR1* and *Mlu I* restriction endonuclease sites located in the multiple cloning site (see Appendix VI), upstream of the *HIS3* gene of the vector pHISi-1 were reproducibly seen.

The plasmid pHISi-1 (5.4 kb) is a yeast integrative plasmid, containing the yeast *HIS3* gene downstream of a multiple cloning site (the map of this plasmid is shown in Appendix VI). $P_{min HIS}$ is the minimal promoter of the *HIS3* gene. pHISi-1 contains a ColE1 origin of replication (*ori*) and the ampicillin resistance gene (Amp^R) for propagation and selection in *E. coli*. The plasmid pHISi-1 cannot replicate autonomously in yeast, however, the presence of the *HIS3* gene allows its integration into the mutated *HIS3* gene of the YM4271 yeast strain, which was used as the host.

The presence of the binding site in the reporter plasmid was further confirmed by digestion of the reporter plasmid with *EcoR1* and *Mlu I* restriction endonucleases and analysis of the products by agarose gel electrophoresis (Figure 24). Two distinct bands present in the digest

147

represent the *EcoR1 – Mlu1* fragment containing the four repeat copies of the putative binding site (138 bp) and the linearised vector pHISi-1 (5.4 kb) respectively.

Lane 4 shows only a single band corresponding to the digested vector pHISi-1, after digestion with *EcoR1* and *Mlu I,* since the second digestion product (23 bp) was too small to detect (see the multiple cloning site of the vector pHISi-I in appendix VI).

It should be noted that the gel had to be overloaded and over exposed to be able to visualise the small 138 bp digestion product (indicated in lane 2) corresponding to the putative binding site, and visible towards the bottom of the gel. The other diffuse and faint band visible in lanes one to four corresponds to a small amount of RNA not removed during the isolation procedure. A size marker was also included on the gel so that the sizes of the digestion products could be determined (Lane M).
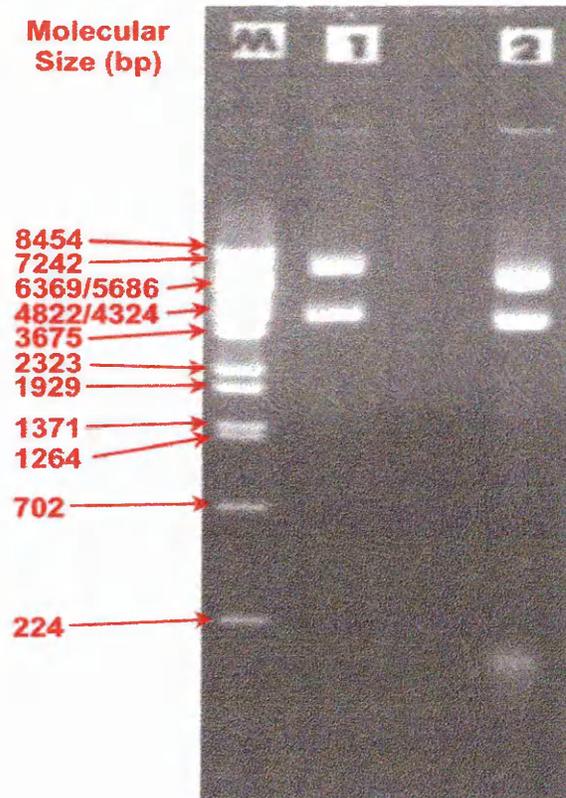
Figure 23. Analysis by agarose gel (1%) electrophoresis of the reporter
plasmid carrying four repeat copies of the putative binding site
after large-scale plasmid DNA isolation.

Lane M. A λ DNA molecular size marker (*BstE II* digest of λ phage DNA).

Lane 1. Reporter plasmid DNA isolated from clone 4.

Lane 2. Vector DNA (pHISi-1) used as a control sample.

Figure 24. Confirmation that the reporter plasmid carries the putative binding
site.

Lane M. A λ DNA molecular size marker (*BstE II* digest of λ phage DNA).

Lane 1. Undigested reporter plasmid DNA.

Lane 2. Double restriction endonuclease digestion of the reporter plasmid
with *EcoR1* and *Mlu I*.

Lane 3. Undigested vector DNA (pHISi-I).

Lane 4. Double restriction endonuclease digestion of the plasmid vector
(pHISi-I) with *EcoR1* and *Mlu I*.

150

The identity of the putative binding site inserted into the pHISi-I

plasmid was further confirmed by manual (Figure 25) and automated

sequencing (Figures 26 and 27).

The primer used for manual and automated sequencing in the forward

direction was HIS – 5372 (which corresponds to the primer located between

nucleotide positions 5372 and 5392 on the pHISi-I plasmid sequence shown

in Appendix VI):

<pre>
        5'- TATTATCATGACATTAACCTA -3'
           5372   5392
</pre>

and the primer used for sequencing the complementary strand (automated

sequencing only) was HIS – RP (where RP means reverse primer located

between nucleotides 99 – 79 on the plasmid sequence):

<pre>
        5'- ATCACATTACTTTATATAATG -3'
           99   99
</pre>

Single stranded template DNA for sequencing was prepared by denaturing

the reporter plasmid DNA with sodium hydroxide (at a final concentration of

0.2 M). Following ethanol precipitation and resuspension in TE buffer the

single stranded DNA was manually sequenced using the method described

in section 4.0. An autoradiogram of the sequencing gel is shown in Figure

25. The sequence GAATTC (highlighted) represents the *EcoR1* restriction

endonuclease site synthesised at the 5' end of the putative binding site and

used for cloning. The binding site sequence is shown in italics.
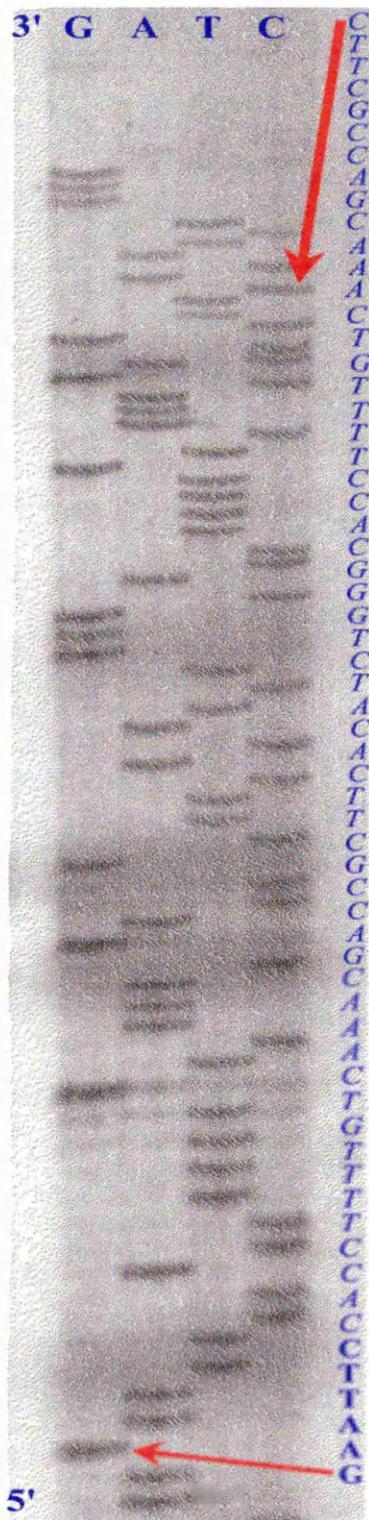
Figure 25. An autoradiogram of the manual sequencing results confirming the presence of the putative binding site in the reporter plasmid.

```
  1   AACCTATAAA AATAGGCGTA TCACGAGGCC CTTTCGTCTT CAAGAATTCC

 51   ACCTTTTGTC AAACGACCGC TTCACATCTG GGCACCTTTT GTCAAACGAC

101   CGCTTCACAT CTGGGCACCT TTTGTCAAAC GACCGCTTCA CATCTGGGCA

151   CCTTTTGTCA AACGACCGCT TCACATCTGG GACGCGTTCG CGAATCGATC

201   CGCGGTCTAG AAATTCCTGG CATTATCACA TAATGAATTA TACATTATAT

251   AAAGTAATGT GATTTCTTCG AAGAGTATAC TAAAAAATGA GCAGGCAAGA

301   TAAA
```

Figure 26. The automated sequencing results corresponding to the 'sense' strand of the reporter plasmid confirming the presence of the putative protein binding site in the reporter plasmid.

The above data show two I.U.B. base-ambiguity symbols represented by the letters V and S at positions 275 and 296 respectively. The symbol V represents the bases G, A, or C, and comparison of the above sequence with that of the vector pHISi-1 (shown in Appendix VI) indicates that the base represented by V should be an A, and that the letter S representing bases G or C should be a C base.

These I.U.B. base-ambiguity symbols lie within the reporter vector sequence and not within the putative binding site sequence identifiable between the restriction endonuclease sites *EcoR1* and *Mlu I*.

153

```
  1  CGAAGAAATC ACATTACTTT ATATAATGTA TAATTCATTA TGTGATAATG
        N                                    Nsi I

 51  CCAGGNATTT CTAGACCGCG GATCGATTCG CGAACGCGTC CCAGATGTGA

101  AGCCGTCGTT TGACAAAAGG TCCCCAGATG TGAAGCGGTC GTTTGACAAA

151  AGGTGCCCAG ATGTGAAGCG GTCGTTTGAC AAAAGGTGCC CAGATGTGAA
                                                        G
                          EcoR I

201  GCGGTCGTTT GACAAAAGGT GGAATTCTTG AAGACGAAAG BCCCTCGTGA

251  TACGCCTATT TTTATAGGTT AATTTCATGA TAATAATGGT TT
                             G
```

Figure 27. Automated sequencing results of the complementary strand of the reporter plasmid confirming the presence of the putative binding site.

The sequencing data above contains two I.U.B. base-ambiguity symbols, the first represented by an N at position 56 corresponding to a G, A, T or C base, and the second a B at position 241 corresponding to either a C, G or T base. Comparison of the above sequence to the vector (pHISi-1) should show an A and a G at the respective positions.
At position 274 a base misincorporation has occurred during sequencing and the T should actually be a G.

154

3.4.2. Construction of the reporter strain

To prepare the reporter strain, the reporter plasmid DNA had to be linearised. This was achieved by digesting the DNA with *Xho I* restriction endonuclease, which cuts the plasmid DNA uniquely within the *HIS3* gene downstream from the multiple cloning site.

The completeness of the digestion was confirmed by analysis of the digestion mixture on a 1 % agarose gel (Data not shown). Isolation of the digestion products was achieved by excising the linearised fragment from the gel, and recovering the DNA by microcentrifugation through siliconised glass wool into a sterile Eppendorf tube. The ethidium bromide was then removed and the DNA precipitated with ethanol, dried *in vacuo* and resuspended in TE buffer for application in the yeast transformation. An aliquot of the DNA sample was analysed by agarose gel electrophoresis (Figure 28).

The undigested DNA samples of the plasmid pHISi-I used as a positive control (lane 3) and the reporter plasmid (lane 1) both show more than one band, these represent the different mobilities of the different forms of the plasmid. The digested plasmid DNA samples (lanes 2 and 4) show only single bands due to the unique *Xho 1* restriction site located within the *HIS3* gene locus.

A *BstE II* digest of λ phage DNA was used as a molecular size marker to determine the size of the digestion products.
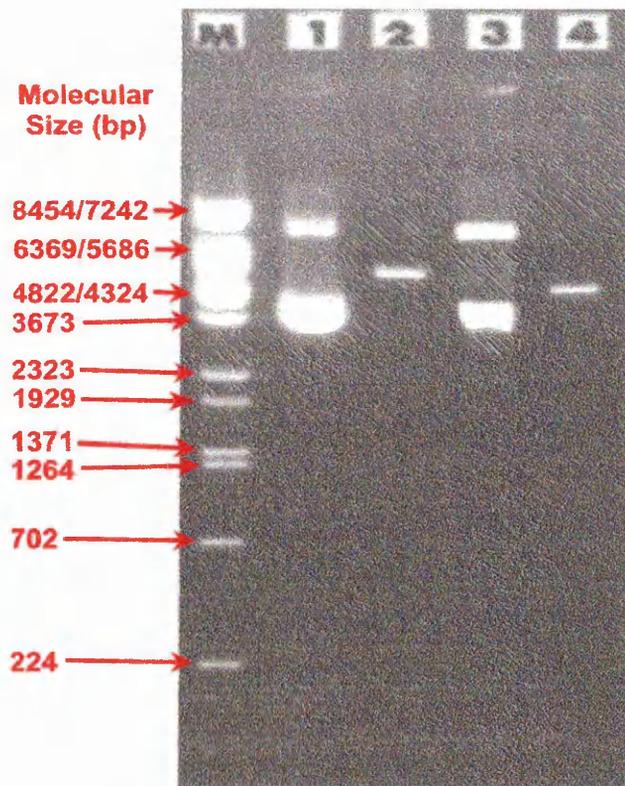
Figure 28.   Analysis by agarose gel (1%) electrophoresis of the linearised
reporter plasmid and positive control DNA before
transformation into *Saccharomyces cerevisiae* YM4271 host
cells.

Lane M.   λ DNA molecular size marker (*BstE II* digest of λ phage DNA).

Lane 1.   Undigested sample of reporter plasmid DNA.

Lane 2.   Linearised reporter plasmid following restriction endonuclease
digestion and purification.

Lane 3.   Undigested sample of the positive control DNA (pHISi-1).

Lane 4.   Positive control DNA (pHISi-I) following linearisation with *Xho 1*.

156

Positive control experiments for various stages of the transformation and integration processes were also carried out. The replicating plasmid YE*pmyc* 181 was applied in an experiment, which acted as a control for the transformation process. YE*pmyc* 181 is able to replicate autonomously because it contains the origin of replication from a naturally occurring yeast plasmid termed a $2\mu m$ circle.

The integrating plasmid p53HIS, which carries three tandem copies of the consensus p53 binding site, was used in a control experiment for the integration process. This plasmid was linearised with *Xho 1* restriction endonuclease and introduced into competent YM4271 yeast cells in the same way as YE*pmyc* 181.

In order to perform the transformations, competent cells were prepared. *Saccharomyces cerevisiae* strain YM4271 cells were grown to mid-log phase represented by an OD_{600} of between 0.5 and 0.7, collected by centrifugation, washed in TE buffer and treated with fresh TE/LiAc. The plasmid DNA was then mixed with the competent cells and after a heat shock at 42°C for 15 min followed by an incubation on ice the transformation mixtures were spread on appropriate plates (SD/+HIS/-LEU for cells carrying YE*pmyc* 181 and SD/-HIS for cells with p53HIS) and incubated for 4 to 6 days at 30° C (Figure 29 A and B).

Figure 29A. Putative clones growing on selective medium following
transformation of yeast strain YM4271 with the (5.76 kb)
replicating plasmid YE*pmyc* 181.
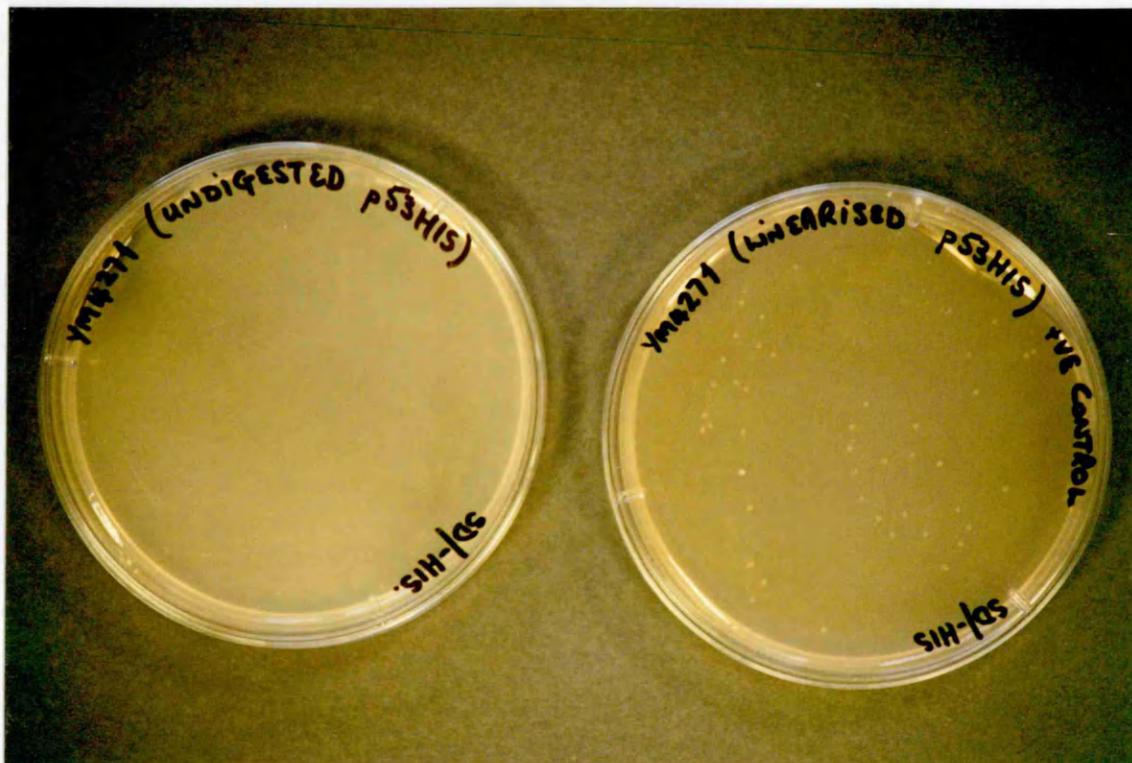
Figure 29B. Putative recombinant clones growing on selective medium after
transformation of yeast strain YM4271 with *Xho 1* linearised
p53HIS DNA.

Several of the larger colonies carrying the putatively integrated p53HIS DNA were isolated and grown in fresh SD/-HIS at 30° C for 4 to 6 days with shaking until the cultures were saturated. Chromosomal DNA was isolated from each clone using the technique described in section 8.4, and an aliquot (1 μl) used to confirm the presence of the consensus p53 binding site by PCR. The PCR was carried out essentially as described in section 6.3 using the HIS-RP primer, (5'-ATCACATTACTTTATATAATG-3', which is located between nucleotide positions 79 – 99 of the plasmid pHISi-I sequence shown in Appendix VI), and the HIS-5372 primer, (5'-TATTAT CATGACATTA ACCTA-3' located between nucleotides 5372 – 5392). The results were analysed by agarose (2%) gel electrophoresis.

Genomic DNA isolated from the wild type yeast strain YM4271 (Figure 30A Lane 2) shows a single high molecular weight band. The faint diffused band at the bottom of this lane is RNA, which was not removed during the isolation. Amplification of the wild type chromosomal DNA did not yield a PCR product, as expected (Figure 30A Lane 1).

The PCR negative control (Figure 30A Lane 3), containing all the PCR components, except a DNA template, and used to check for reagent contamination, shows no PCR product(s) as expected.

Genomic DNA isolated from the putative clones carrying the integrated p53 consensus binding site (Figure 30A Lanes 4 and 6 and Figure 30B Lanes 1 and 3) show high molecular weight DNA bands near the sample wells.

Amplification of the genomic DNA from the putative clones described above Lanes 5 and 7 (Figure 30A) and 2 and 4 (Figure 30B) yielded a major PCR product of approximately 1000 bp, this corresponds exactly with the size of the product obtained in the positive control, (and a minor product of approximately 700 bp).

A positive PCR control corresponding to the amplification product of the p53HIS plasmid DNA (120 ng) Lane 8 (Figure 30A) and Lane 5 (Figure 30B) show one amplification product at approximately 1000 bp.

Lambda phage DNA digested with the restriction endonuclease *BstE II* (Lane M) was also run as a size marker for the identification of the amplification products resulting from the PCR.
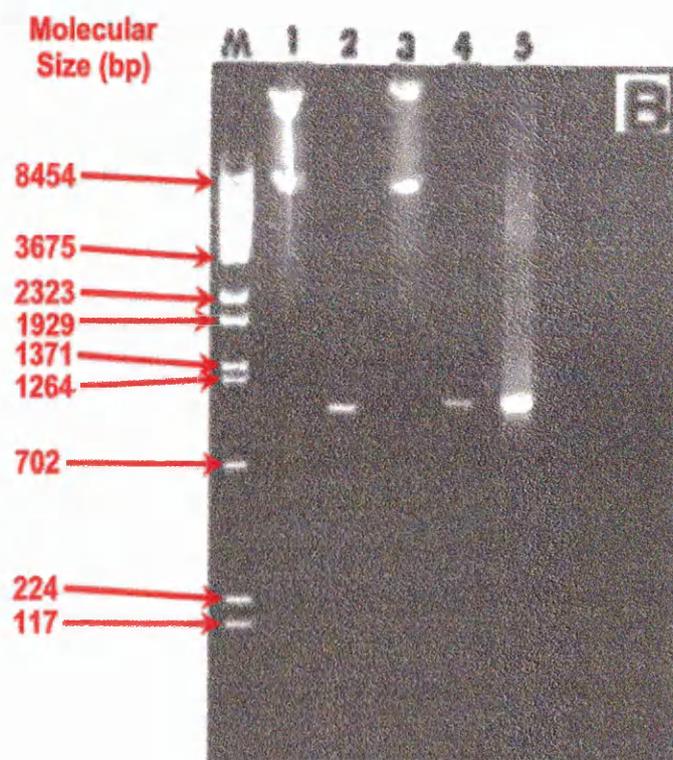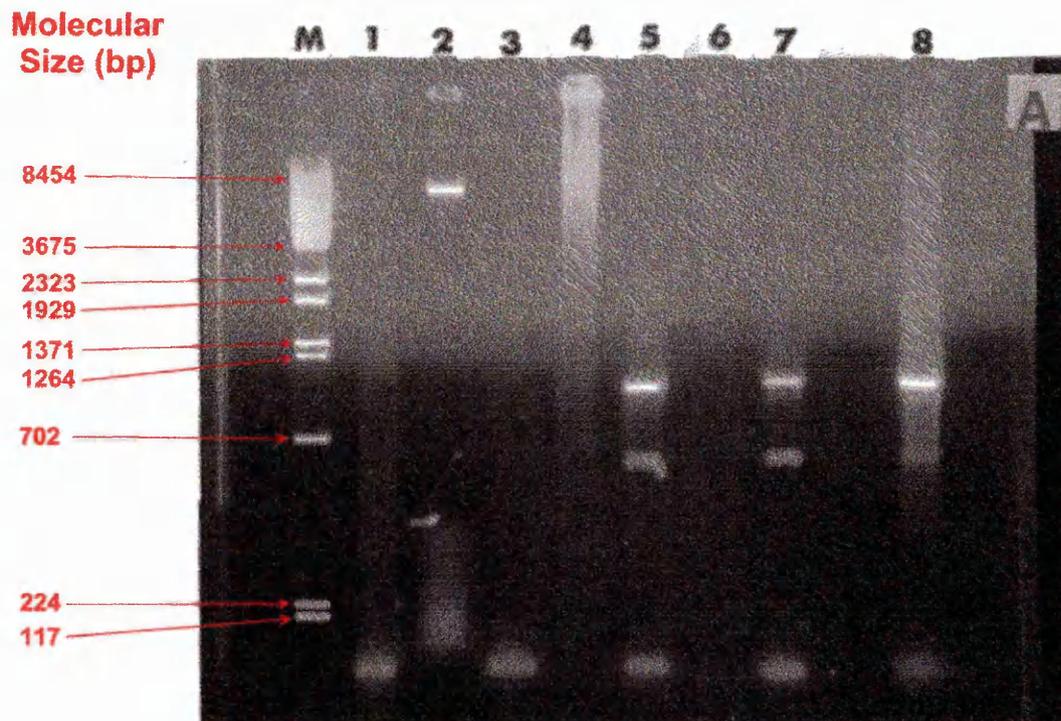
Figure 30. Analysis by agarose gel electrophoresis of genomic and PCR amplified DNA from the putative clones carrying the integrated consensus p53-binding site.

Lane M. A λ DNA molecular size marker (*BstE II* digest of λ phage DNA).

Figure 30A Lane 1. PCR product from the genomic DNA isolated from the

wild type strain YM4271.

Figure 30A Lane 2. Genomic DNA from wild type strain YM4271.

Figure 30A Lane 3. Negative PCR control.

Figure 30A Lane 4 and 6 Genomic DNA isolated from recombinant yeast

clones.

Figure 30A Lane 5 and 7. PCR product from the genomic DNA isolated from

the putative yeast clones.

Figure 30A Lane 8. Positive PCR control (amplified p53HIS template).

Figure 30B Lane 1 and 3. Genomic DNA isolated from recombinant yeast

clones.

Figure 30B Lane 2 and 4. PCR product from the genomic DNA isolated from

the putative yeast clones

Figure 30B Lane 5. Positive PCR control (amplified p53HIS template).

162

The reporter plasmid, which carries four tandem copies of the putative

binding site, was linearised with *Xho1* restriction endonuclease and

introduced into competent YM4271 yeast cells in the same way as the

p53HIS DNA. Cells transformed with the reporter plasmid were spread on

SD/-HIS plates and incubated at 30° C for 4 to 6 days (Figure 31).

Genomic DNA was isolated from nine of the larger putative reporter

strain colonies as well as from the wild type yeast host and its integrity

examined by agarose gel electrophoresis, (Figure 32A). Lanes 1 to 9

respectively show the DNA isolated from these putative clones and Lane 10

from the wild type yeast host cells.

An aliquot (1 µl) of this DNA was used to confirm by PCR the presence of

the putative binding site in the reporter strain (Figure 32B). The PCR was

carried out essentially as described in section 6.4 using HIS-RP and HIS-

5372 primers. The location of the two primers used are shown highlighted on

the pHISi-I plasmid sequence in Appendix VI

The results of the PCR were analysed by agarose (2%) gel

electrophoresis. A PCR negative control containing all the components

necessary for PCR except the DNA template was also run (Figure 32B Lane

1), and was used to monitor any contamination in the reagents.

Chromosomal DNA from the wild type yeast host (YM4271) was amplified by

PCR (Figure 32B Lane 2), as expected no PCR product was obtained.

A molecular size marker (*BstEII* digest of lambda phage DNA) and a positive

PCR control (the reporter plasmid containing the four repeat copies of the

putative binding site) were used to determine the size of any amplification product obtained from the putative reporter strain clones (Figure 32B, Lanes M and 12 respectively). Clones were considered positive if a PCR product corresponding to a size of 255 bp was obtained as can be seen in Lanes 3 to 11.

Figure 31. Putative clones growing on SD/-HIS medium following the transformation of yeast host strain YM4271 with the *Xho 1* linearised 5.4 kb plasmid pHISi-1 (carrying four repeat copies of the putative binding site). Undigested reporter plasmid served as a negative control.
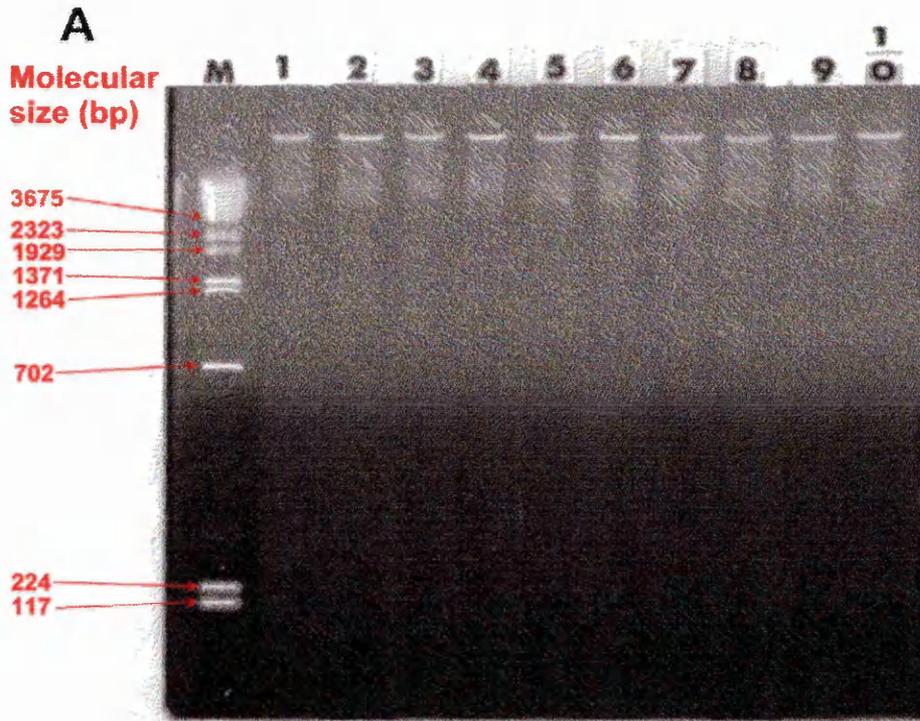
Figure 32A. Analysis by agarose gel electrophoresis of chromosomal DNA isolated from the reporter strain clones carrying the putative binding site.
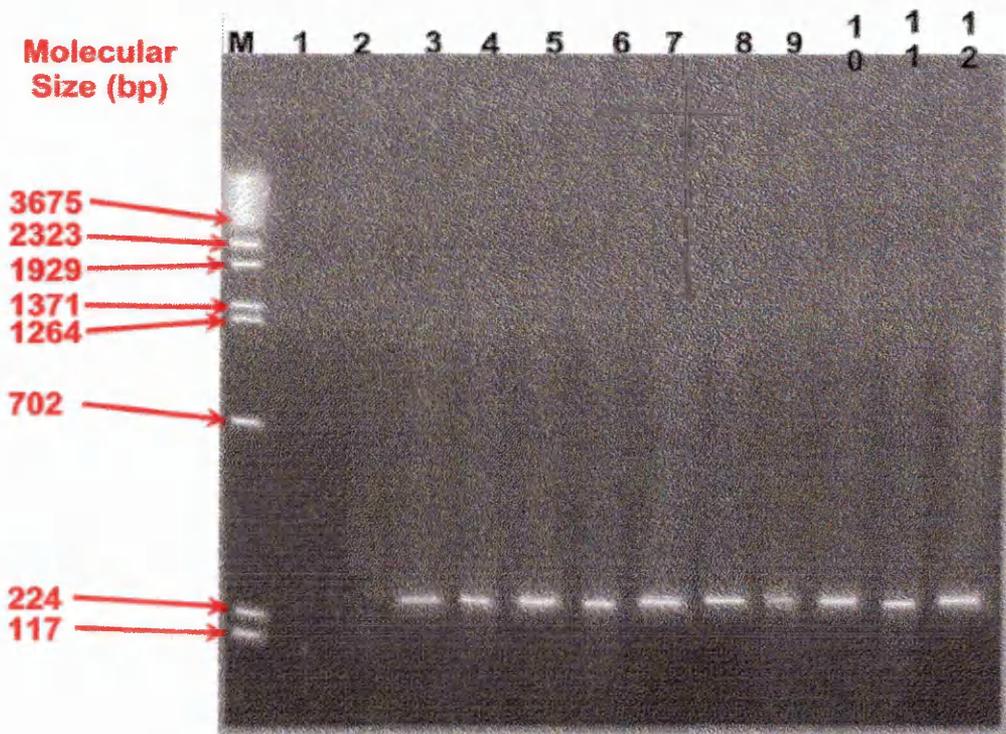
Figure 32B. Analysis by PCR followed by agarose gel (2%) electrophoresis of the reporter strain clones carrying the putative binding site.

166

Figure 32A.

Lane M.   DNA molecular size marker (*BstE II* digest of λ phage DNA).

Lane 1 – 9.   Genomic DNA isolated from putative reporter strain clones.

Lane 10.   Genomic DNA from wild type strain YM4271.

Figure 32B.

Lane M.   DNA molecular size marker (*BstE II* digest of λ phage DNA).

Lane 1.   PCR negative control.

Lane 2.   PCR product from the genomic DNA isolated from the wild type strain YM4271.

Lane 3 – 11.   PCR products from the genomic DNA isolated from the yeast clones

Lane 12.   Positive PCR control (amplified putative binding site using reporter vector DNA).

3.4.3. Testing the reporter strain clones for background *HIS3* expression

A small sample of each of the yeast colonies analysed by PCR was suspended in TE buffer, and an aliquot (5 µl) pipetted onto the surface of pre-dried SD/-HIS plates containing 3 – AT at the following concentrations, 0, 5, 10, 15, 30, 45, and 60 mM.

It should be noted that the plates were pre-dried to prevent condensation causing localised dilution of the 3 – AT concentration, leading to false positive results. The plates were incubated at 30° C for 4 to 6 days, and the growth analysed (Figure 33).

All the putative recombinant clones grew on the SD/-His plates in the presence of 15 mM 3 – AT which shows that the background *HIS3* expression is low, and these clones were used for transformation with the cDNA/AD Fusion library and further study. However, the only exception to this was clone 1 (Figure 33E), which grew in the presence of 45 mM 3 - AT indicating that the background *HIS3* expression is high, as a result this clone was not used for further experimentation.
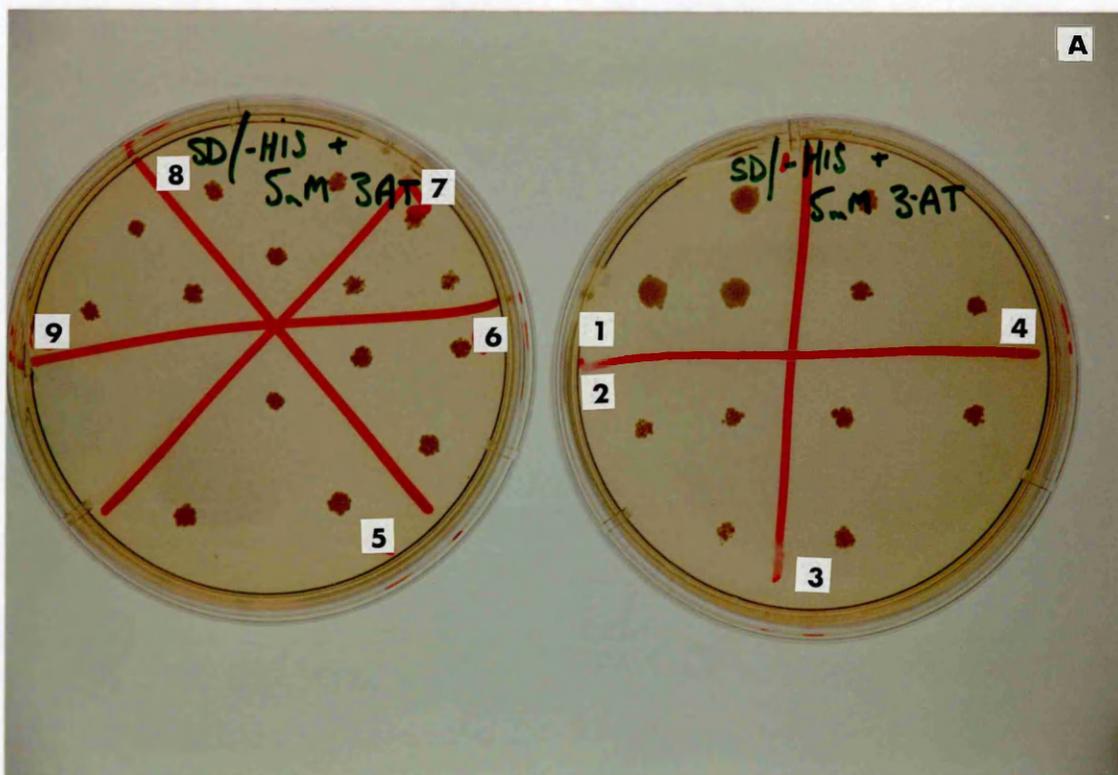
Figure 33A. Testing of reporter strain clones for background expression of the *HIS3* gene by growth on SD/-HIS medium containing 5mM 3 – AT.
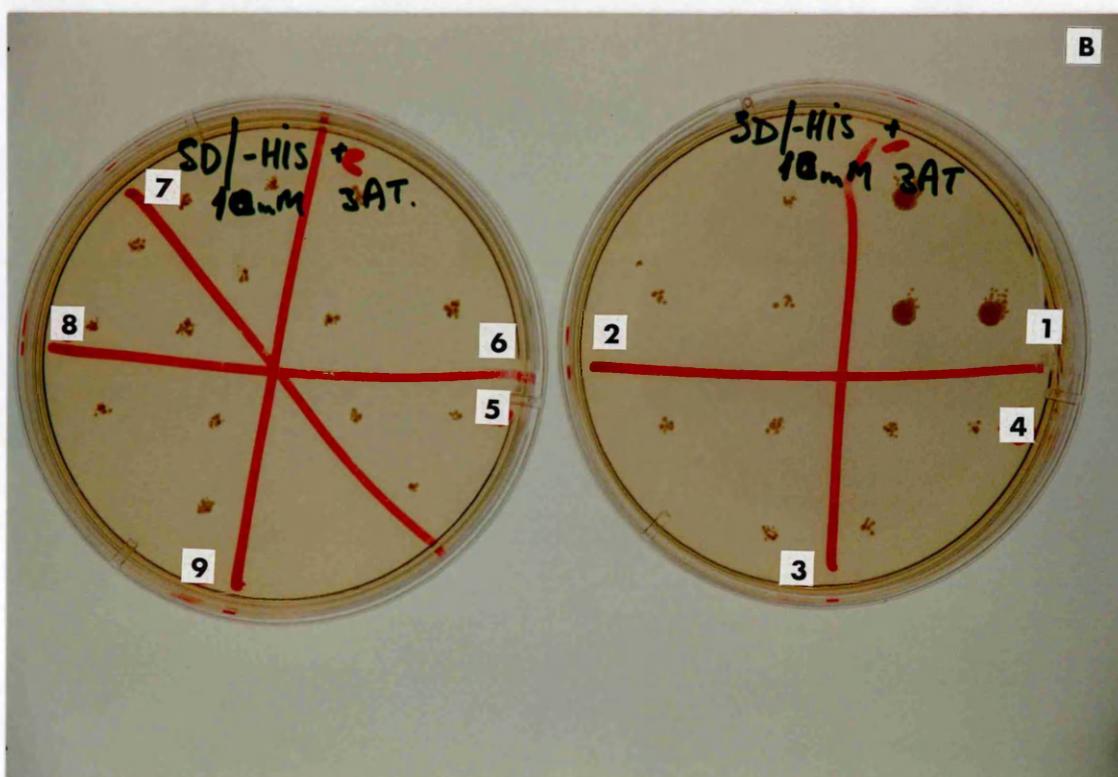
Figure 33B. Testing of reporter strain clones for background expression of the *HIS3* gene by growth on SD/-HIS medium containing 10mM 3 – AT.
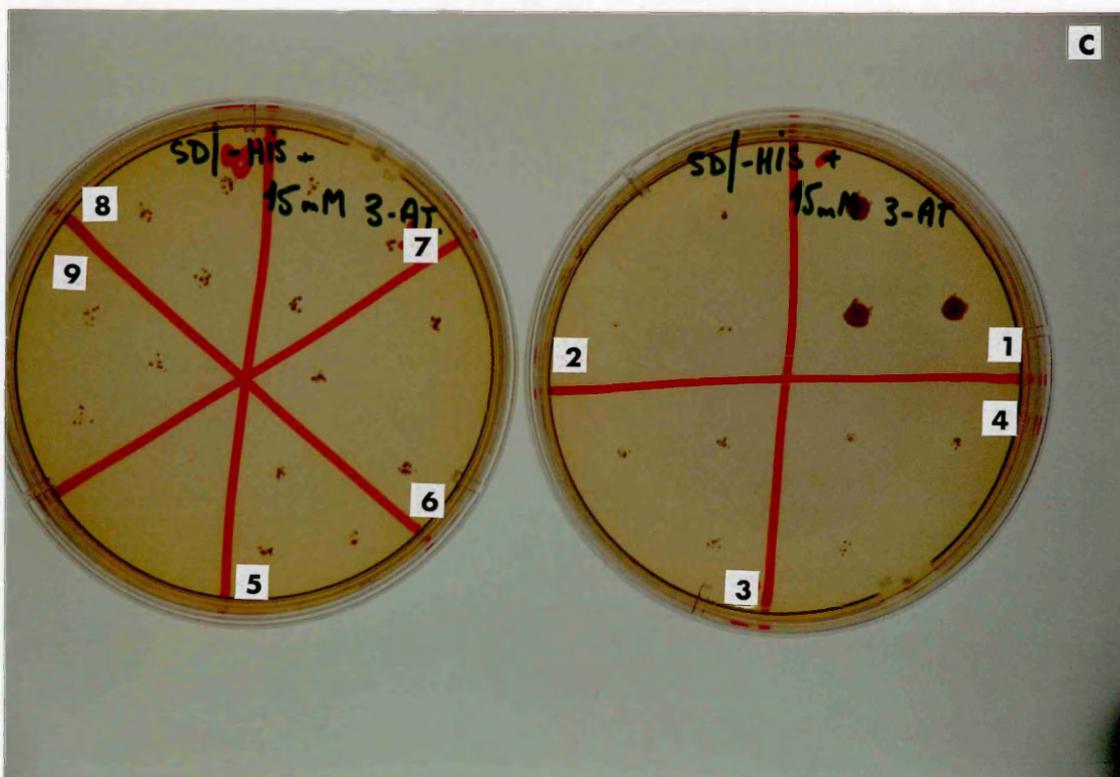
Figure 33C. Testing of reporter strain clones for background expression of the *HIS3* gene by growth on SD/-HIS medium containing 15mM 3 – AT.
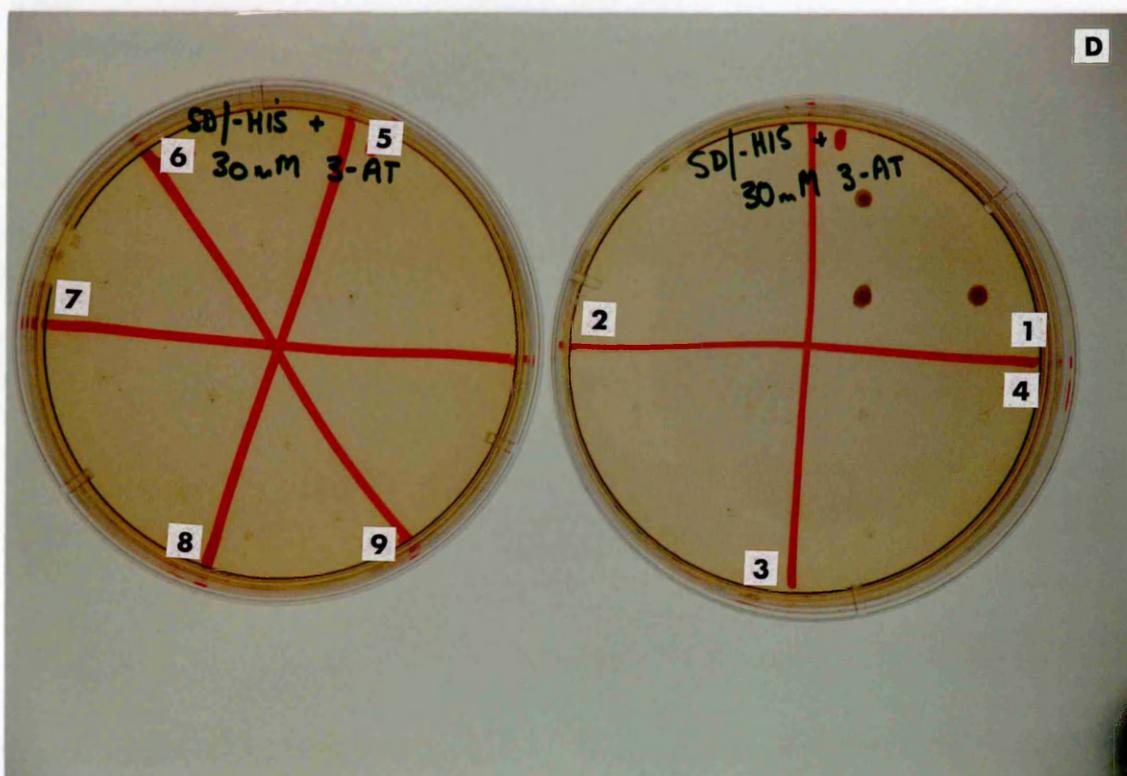
Figure 33D. Testing of reporter strain clones for background expression of the *HIS3* gene by growth on SD/-HIS medium containing 30mM 3 – AT.
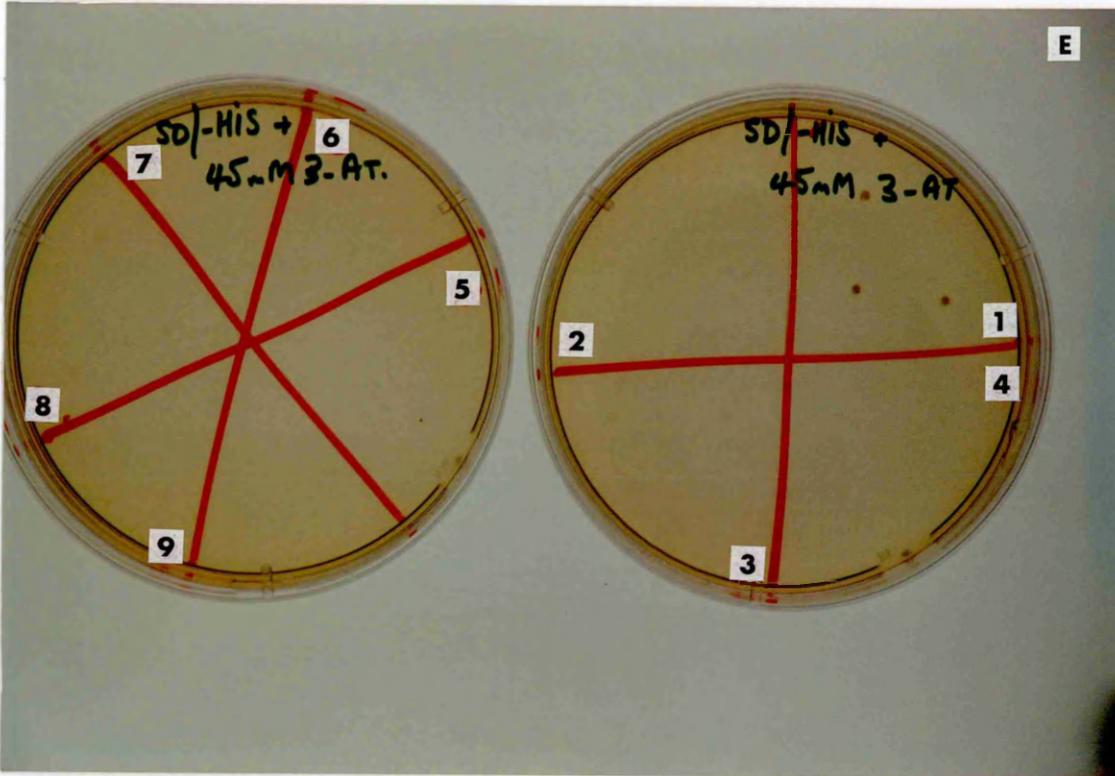
Figure 33E. Testing of reporter strain clones for background expression of the *HIS3* gene by growth on SD/-HIS medium containing 45mM 3 – AT.
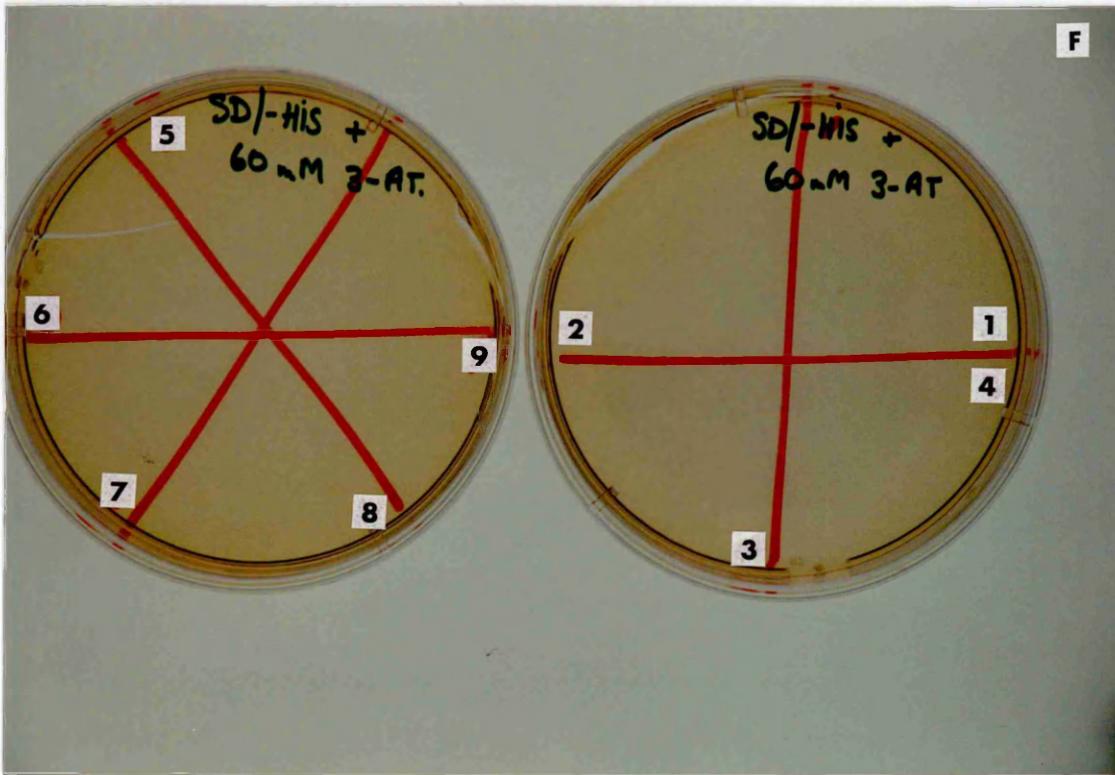
Figure 33F. Testing of reporter strain clones for background expression of the *HIS3* gene by growth on SD/-HIS medium containing 60mM 3– AT.

171

The search for putative transcription factors was carried out by the transformation of yeast reporter strain clones with a rat brain cDNA/AD fusion library. To achieve this a culture of the reporter strain (clone 3) was grown to mid-log phase in SD/-HIS + 30 mM 3 – AT and competent cells prepared. cDNA/AD fusion library and carrier DNA were mixed with competent reporter strain cells, freshly prepared PEG/LiAc solution and incubated for 1 h with shaking. A heat shock for 15 min, followed by an incubation on ice was performed. After pelleting the cells and resuspension in TE buffer (pH 7.5) an aliquot of the transformation mixture (50 µl) was spread over the surface of pre-dried SD/-HIS/-LEU plates containing 30 mM 3 – AT and incubated for 4 to 6 days at 30° C. Colonies resulting from *HIS3* activation were larger than the colonies resulting from background *HIS3* expression (Figure 34). These larger colonies were cultured in fresh SD/-HIS/-LEU medium containing 30 mM 3 – AT for 4 to 6 days and the plasmid DNA isolated as described in section 9.2 (Figure 35).

DNA was isolated from both the wild type host and reporter strain clone 3 (Lanes 1 and 2 respectively). No plasmid DNA was isolated from these strains of yeast as expected, however, the high molecular weight DNA visible in both lanes is chromosomal DNA, a common and unavoidable contaminant when isolating plasmid DNA from yeast cells by this method. Plasmid DNA was isolated from 7 recombinant yeast clones and varied in size from approximately 7.1 to 8.8 kb.

The yield from the clones shown in lanes 7 and 9 was slightly lower than from the other clones. Chromosomal DNA contamination is visible in all lanes near the sample wells.

A λ DNA molecular size marker (Lane M) was used to determine the size of the plasmid DNA isolated.

Figure 34. Putative clones resulting from the transformation of the reporter strain with a cDNA/AD library plasmid growing on selective SD/-HIS/-LEU medium containing 30 mM 3-aminotriazole.
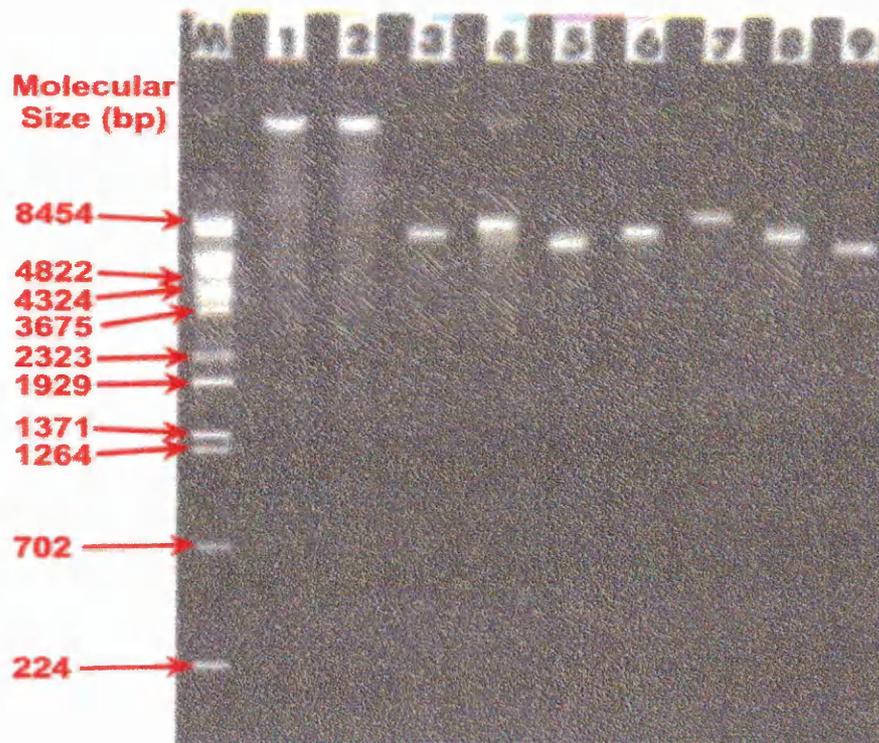
Figure 35. Analysis of the plasmid DNA isolation from the putative positive clones transformed with an cDNA/AD fusion library from rat brain by agarose gel (1%) electrophoresis.

Lane M. Molecular size marker.
Lane 1. DNA isolated from wild type host.
Lane 2. Reporter strain DNA isolated from the reporter strain clone.
Lanes 3 to 9. Isolates of plasmid DNA from the putative clones following transformation with cDNA/AD fusion library.

An aliquot (5 μl containing 0.5μg of plasmid DNA) isolated from each putative recombinant clone was digested with *EcoR1* restriction endonuclease and the products analysed by agarose (1 %) gel electrophoresis (Data not shown).

The plasmid DNA isolated from the seven recombinant clones was further analysed by PCR using the library specific primers AD10[S] 5'-CTAT TCGATGATGAAGATACC-3' and AD10[R] 5'-CACAGTTGAAGTGAACTTGCG-3' (located between nucleotide positions 788 to 807, and 903 to 893 respectively on the pGAD10 sequence shown in Appendix VI). Clones 1,5 and 6 (Figure 36 lanes 1,5 and 6 respectively) show a PCR amplification product of approximately 0.75 kb. Clone 2 (lane 2) showed the smallest PCR product with a molecular size of approximately 0.4 kb. Clones 3 and 7 (lanes 3 and 7) show an intermediate sized PCR product estimated to be 0.6 kb, whereas, recombinant clone 4 showed the largest cDNA insert following amplification by PCR of about 1.2 kb.

Plasmid DNA from the positive cDNA/AD fusion library clones was purified by agarose gel (1%) electrophoresis and recovered by centrifugation through glass wool. Following ethanol precipitation and resuspension in deionised water the plasmid DNA was automatically sequenced using the AD10[S] primer (Figures 37A to G).
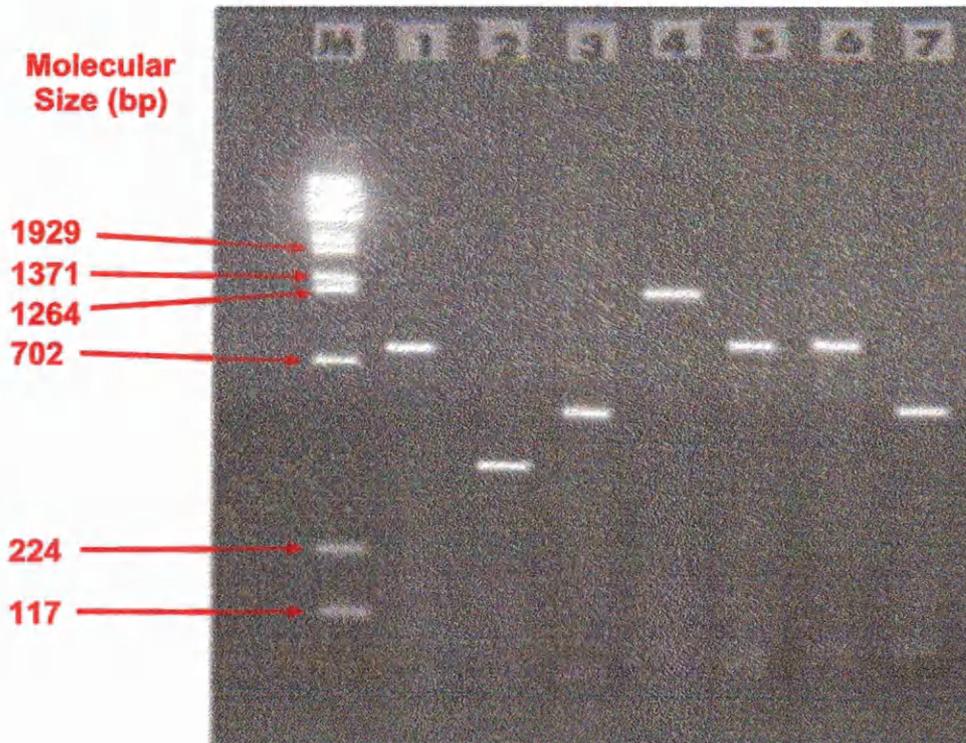
Figure 36. Analysis by agarose gel (2%) electrophoresis of the library clones
following amplification by PCR.

Lane M. Molecular size marker (Lambda phage DNA digested with *BstE II*).

Lanes 1 to 7. PCR amplification products obtained from the plasmid DNA
isolated from the recombinant library clones.

```
1      accaaaccca aaaaaagaga tctctcgagg cggcggtggg
41     atccgaattc atggcgcgac cgggctggcc tttccatcgc
81     agccccagca gcagcagcag ccgcctcagc cgccgcacca
121    cttgccccaa cagatcccgg tgcagcaccc gcggctgcag
161    gcgctgaagg aagagccgca gacggtgccg gagatgccgg
201    gagagacgcc gcccttgtcc cccatcgaca tggagtctca
241    ggagcggatc aaggcggaga ggaagcgcat gagaaaccgc
281    atcgctgcct ccaagtgccg gaaaaggaag ctggagcgga
321    tcgcccggct agaggaaaaa gtgaaaacct tgaaagcgca
361    aaactccgag ctggcgtcca cggccaacat gctcagggaa
401    caggtggcac agcttaaaca gaaagtcatg aaccacgtta
441    acagtgggtg ccaactcatg ctaacgcagc agttgcaaac
481    gttttgagga cagactgtca gggctgaggg gcagtggaag
521    aaaaaaataa cagagataaa cttgagaact tgactggttg
561    cgacagagaa aaaaaagtgt ccgagtactg aagccaaggg
601    tacacaagat ggactgggtt gcgacctgac ggcgccccca
641    atgtgctgga gtgggaagga cgtggcgcgc ctggctttgg
681    cgaggagcca gagagcagcg gcctgttcgc gacgctttgc
721    gaacgggctg tgccgcgacc agaacgatgg actttttcgtt
761    aacattgacc aagaactgca tggacctaac attagaattc
801    cagatctatg aatcgtagat actgaaaaac
```

Figure 37 A. Sequencing results of the plasmid DNA from the positive
     cDNA/AD fusion library clone 1.

```
  1      accaaaccca aaaaaagaga tctctcgagg tttttttttt
 41      atccgaattc gcgaacagaa ggcgcacctg gaagactact
 81      actggatgac cggctacccg cagcagctca acccggaggc
121      gctgggcttc agcccggagg acgcagtcga ggcgcncatc
161      agcaatagcc accagctcca gggtggctnc gvtggctatg
201      cgcggggcgc gcagcagctg gccgcggcag cggggggccgg
241      cgccggcgcc tccctgggcg gcagcggcga ggagatgggc
281      cccgccgccg ccgtggtgtc cgccgtgatc gccgcggcsg
321      ccgcgcagag cggcggggca cctcactacc atcaccacca
361      ccaccacgcc acggggcacc accaccaccc gacggccggc
401      gccccgggag ccgcgggtag cgcgtccgcc tctgcgagcg
441      gcgcgggtgg cgcgggcggc ggtggaattc cagatctatg
481      actgaaaaac
```

Figure 37 B. Sequencing results of the plasmid DNA from the positive cDNA/AD fusion library clone 2.

```
  1      accaaaccca aaaaaagaga tctctcgagg gtgcctcaaa
 41      atccgaattc cggagccgag cggctcgccc gtgcacgtcc
 81      agcttcccca gcaggcggcc ccggtgacag ccgcggcggc
121      ggctccggcg gccgcgacat cagcaccggc cccggccccg
161      gccccggcgg cgcccgcagc cccggccccg gctccagctg
201      cggctccagc cccggccccg gcagctcagg cggtcggctg
241      gcccatctgc agggacgcgt acgagctcca ggaggttatc
281      ggcagtggag cgaccgccgt ggttcaggca gccctgtgca
321      aacccaggca agaacgcgta gccataaagc ggatcaactt
361      ggaaaagtgc cagacgagta tggatgaact cttaaaagaa
401      attcaagcca tgagccagtg cagccatccc aacgtagtga
441      cttattatac ttcctttgtg gtcaaagatg aactttggct
481      ggtcatgaaa ttactaagtg gaggttccat gttggatatc
521      atcaaataca tcgtcaatcg gggagaaatt cagatctatg
561      aatcgtagat actgaaaaac
```

Figure 37 C. Sequencing results of the plasmid DNA from the positive cDNA/AD fusion library clone 3.

```
1        accaaaccca aaaaaagaga tctctcgagg gacgactggt
41       atccgaattc ggagaacctc tgtcgctggg gctggtccgc
81       gggctccgag gaaccgctgc ttcctgagag cgctccgtga
121      gtgaccgcga cttttcaaag ctccggatcg cgcgggacca
161      accaacgtga gtgcaagcgg tgtcttaacc ctgcgctccc
201      tggagcgaac tggggaggag ggcgcagggg ggagcactgc
241      ggtctggagc gcacgctcct aaacaaactt tgttactgaa
281      gcggggacgc gcgggtatcc cccgcttccc ggcgcgctgt
321      tgcggccccg aaacttctgc gcacagccca ggctaacccc
361      gcgtgaagtg accgactgtt ctatgactgc aaagatggaa
401      acgaccttct acgacgatgc cctcaacgcc tcgttcctcc
441      agtccgagag tggcgcctac ggctacagta accctaagat
481      tctgaagcag agcatgacct tgaacctggc cgacccggtg
521      ggcaatctga agccgcacct ccgagccaag aactcggacc
561      ttctcacgtc gcccgacgtc gggctgctca agctggcgtc
601      gccggagctg gagcgcctga tcatccagtc cagcaatggg
641      cacatcacca ctacaccgac ccccactcag ttcttgtgcc
681      ccaagaacgt gaccgacgag caggagggct tcgccgaagg
721      cttcgtgcgc gccctagctg aactgcatag ccagaatacg
761      ctgcccagtg tcacctccgc ggcacaacct gtcagtgggg
801      cgggcatggt cgctcccgct gtggcctcag tagctggcgc
841      tggcggcggc ggcggctaca gcgccagcct gcacagtgag
881      cctccggtct acgccaacct cagcaacttc aacccgggtg
921      cgctgagcag cggcggtggg gcgccctcct atggcgcgac
961      cgggctggcc tttccatcgc agccccagca gcagcagcag
1001     ccgcctcagc cgccgcacca cttgccccaa cagatcccgg
1041     tgcagcaccc gcggctgcag gcgctgaagg aagagccgca
1081     gacggtgccg gagatgccgg gagagacgcc gcccttgtcc
1121     cccatcgaca tggagtctca ggagcggatc aaggcggaga
1161     ggaagcgcat gagaaaccgc atcgctgcct ccaagtgccg
1201     gaaaaggaag ctggagcgga tcgcccggct agaggaaaaa
1241     gtgaaaacct tgaaagcgca attagaattc cagatctatg
1281     aatcgtagat actgaaaaac
```

Figure 37 D. Sequencing results of the plasmid DNA from the positive
cDNA/AD fusion library clone 4.

```
1     accaaaccca aaaaaagaga tctctcgagg tgggagagca
41    atccgaattc ttctgctgcc ctgcctctga gaatgcacca
81    cagtgccgct gagggtctcc cagcctccaa ccacgccacc
121   aacgtgatgt ccacagcgac aggacttcat tattctgtgc
161   cttcctgtca ttatggaaac cagccatcca cctacggagt
201   gatggcaggc actttaaccc cttgtcttta caagtttcca
241   gaccacaccc tgagtcatgg gtttcctccc ctgcaccaac
281   ctctgctggc ggaggaccca acagcctctg aatttaagca
321   ggaactcagg cggaaaagta aattggtgga agagccaata
361   gacatggact ccccggaaat ccgagaactg gagcagtttg
401   ccaacgaatt taaagtgaga agaattaagt taggatacac
441   ccagacaaac gtgggcgaag ctctggccgc tgtccacggc
481   tccgaattca gtcaaacaac catctgccga tttgaaaact
521   tgcagctcag tttcaaaaat gcttgcaaac tgaaagcaat
561   tttatccaag tggctggagg aagctgagca ggtcggagct
601   ttgtacaatg aaaaagtggg agcaaacgaa aggaagagga
641   aacggaggac aactatcagt atcgccgcta aggatgcttt
681   ggagagacac tttggagagc acagcaaacc ttcttcgcag
721   gagatcatgc ggatggctga agaattgaat ctcgagaaag
761   aagtagtaag agtgtggttt tgcaagaatt cagatctatg
801   aatcgtagat actgaaaaac
```

Figure 37 E. Sequencing results of the plasmid DNA from the positive cDNA/AD fusion library clone 5.

```
1     accaaaccca aaaaaagaga tctctcgagg tcgttattgt
41    atccgaattc tgctgcagga ggcgcaggac ggcgtcctgg
81    atctcaaagc ggctgcagat actttggctg tgaggcaaaa
121   gcgaagaatt tatgatatca ccaacgtctt agaaggaatt
161   gatctaattg aaaaaaaatc aaagaacagt atccagtgga
201   agggtgtagg tgctggctgt aatactaaag aagttataga
241   tagattaaga tgtcttaaag ctgaaattga agatctagaa
281   ttgaaggaaa gagaacttga ccagcagaag ttgtggctac
321   agcaaagcat caaaaacgtg atggaagact ccattaataa
361   cagattttct tatgtaactc acgaagacat ctgcagttgc
401   tttaatggtg atacactgtt ggccattcag gcaccctctg
441   gtacacagct ggaagtacct attccagaaa tgggacagaa
481   tggacaaaag aaataccaga taaatctgaa gagtcactca
521   ggaccaatcc atgtgctact tataaataaa gagtccaact
561   catctaagcc ggtggttttt cctgtccccc cacctgatga
601   cctcacacaa ccttcctccc agtcctcaac ttcagtgact
641   ccaccgaaat ccaccatggc tgctcaaaac ctgcccgagc
681   agcatgtttc tgaaagaagc cagaatttcc agcagacacc
721   agctacagaa atatcttcag gatctattag tggagacatc
761   attgatgaac tgatgtcttc tgacgaaatt cagatctatg
801   aatcgtagat actgaaaaac
```

Figure 37 F. Sequencing results of the plasmid DNA from the positive cDNA/AD fusion library clone 6.

181

```
1     accaaaccca aaaaaagaga tctctcgagg gtccagacct
41    atccgaattc aaatgcctct agatctgtcc cctcttgcaa
81    cacccatcat aagaagcaaa attgaggaac cttctgttgt
121   agaaacaact caccaggata gcccgttacc tcaccccgag
161   tctactacca atgatgaaaa ggaaatacca ttggcacaaa
201   ctgcacagcc cacatcagct atcgttcgtc cagcatcatt
241   acaggttccc aatgtgctgc tcacaagttc tgactsaagt
281   gtaattattc aacaagcagt accttcacca acctcavgta
321   ctgtaatcac ccaggcacca tcctctaaca ggccaattgt
361   tcctgtacca ggcccatntc ctcttctatt acatcttcct
401   aatggacaaa ccatgcccgt tgctattcct gcatcaatta
441   caagttctaa tgtgcatgtt ccagctgcag tcccacttgt
481   tcggccagtc accatggtgc ctagtgttcc aggaatccca
521   ggccnttcct ctcctgaatt cagatctatg aatcgtagat
561   actgaaaaac
```

Figure 37 G. Sequencing results of the plasmid DNA from the positive cDNA/AD fusion library clone 7.

3.5. Analysis of the rat brain cDNA library clones for sequence homology to other transcription factors

To establish whether the cDNA sequences obtained from the recombinant AD/fusion library show homology with known transcription factors, sequence alignment searches were performed. To achieve this the cDNA sequence was first translated (in all reading frames) into an amino acid sequence using the EBI translate programme. The next step was to determine the correct open reading frame (ORF), this was deemed to be the longest frame uninterrupted by a stop codon (TGA, TAA, or TAG). Usually the initiation codon in the coding sequence is that for methionine (ATG), but methionine is also a common residue within the coding sequence so its presence is not an absolute indicator of an ORF. Therefore, recognition of the flanking Kozak sequence (e.g. CCGCCAUGG) may also be helpful in pinpointing the beginning of the coding sequence. Once the open reading frame was established, the amino acid sequence was used as a query (or probe) sequence to search the Swissprot and Swall databases for sequence alignments to other known transcription factors using the BLAST sequence alignment program. The BLAST (Basic Local Alignment Search Tool) algorithm was described by Altschul et al. in 1990. The algorithm itself is straightforward, with the important concept being the segment pair. Given two sequences, a segment pair is defined as a pair of sub-sequences of the same length that form an ungapped alignment. BLAST calculates all the segment pairs between the query and the database sequences above a scoring threshold. The resulting high scoring pairs form the basis of the

ungapped alignments that characterise BLAST outputs as shown for each of the clones below. Sequences clustered at greater than or equal to 62% identity are used to generate the BLOSUM 62 matrix (BLOck SUbstitution Matrix). For each of the searches the 'identities' keyword indicates the number of identical residue matches in relation to the match length, whereas the 'positives' keyword considers both identities and similarities.

The following alignments indicate high levels of sequence similarity and correspondingly low Expected frequency or E-values (e.g. Clone 1 shows an E value of 3e –79) indicating that the probability of this match having occurred by chance is virtually zero.

After the search summary, the output shows the complete pairwise alignment of the high scoring pairs (HSPs) of the query sequence (line 1) with that of the database or subject (line 3). The identities for each aligned high scoring pairs are indicated with the corresponding amino acid symbol shown in line 2.

The translation of clone 1 in all six reading frames:

Reading frame 1.

<u>GATGLAFPSQPQQQQQPPQPPHHLPQQIPVQHPRLQALKE
EPQTVPEMPGETPPLSPIDMESQERIKAERKRMRNRIAAS
KCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQ
LKQKVMNHVNSGCQLMLTQQLQTF*GQTVRAEGQWKKKIT</u>
EINLRT*LVATEKKKCPSTEAKGTQDGLGCDLTAPPMCWS
GKDVARLALARSQRAAACSRRFANGLCRDQNDGLFVNIDQ
ELHGPNIX

Reading frame 2.

MARPGWPFHRSPSSSSSRLSRRTTCPNRSRCSTRGCRR*R
KSRRRCRRCRERRRPCPPSTWSLRSGSRRRGSA*ETASLP
PSAGKGSWSGSPG*RKK*KP*KRKTPSWRPRPTCSGNRWH
SLNRKS*TTLTVGANSC*RSSCKRFEDRLSGLRGSGRKK*
QR*T*ELDWLRQRKKSVRVLKPRVHKMDWVAT*RRPQCAG
VGRTWRAWLWRGAREQRPVRDALRTGCAATRTMDFSLTLT
KNCMDLTLX

Reading frame 3.

WRDRAGLSIAAPAAAAASAAAPLAPTDPGAAPAAAGAEG
RAADGAGDAGRDAALVPHRHGVSGADQGGEEAHEKPHRCL
QVPEKEAGADRPARGKSEN*ESAKLRAGVHGQHAQGTGGT
A*TESHEPR*QWVPTHANAAVANVLRTDCQG*GAVEEKNN
RDKLENLTGCDREKKVSEY*SQGYTRWTGLRPDGAPNVLE
WEGRGAPGFGEEPESSGLFATLCERAVPRPERWTFR*H*P
RTAWT*H*

Reading frame 4.

LMLGPCSSWSMLTKSPSFWSRHSPFAKRREQAAALWLLAK
ARRA*SFPLQHIGGAVRSQPSPSCVPLASVLGHFFFSVAT
SQVLKFISVIFFFHCPSALTVCPQNVCNCCVSMSWHPLLT
WFMTFCLSCATCSLSMLAVDASSEFCAFKVFTFSSSRAIR
SSFLFRHLEAAMRFLMRFLSALIRS*DSMSMGDKGGVSPG
ISGTVCGSSFSACSRGCCTGICWGKWCGG*GGCCCCWGCD
GKASPVAPX

Reading frame 5.

*C*VHAVLGQC*RKVHRSGRGTARSQSVANRPLLSGSSPK
PGAPRPSHSSTLGAPSGRNPVHLVYPWLQYSDTFFSLSQP
VKFSSLSLLFFSSTAPQP*QSVLKTFATAALA*VGTHC*R
GS*LSV*AVPPVP*ACWPWTPARSFALSRFSLFPLAGRSA
PASFSGTWRQRCGFSCASSPP*SAPETPCRWGTRAASLPA
SPAPSAALPSAPAAAGAAPGSVGASGAAAEAAAAAGAAM
ERPARSRH

Reading frame 6.
NVRSMQFLVNVNEKSIV*VAAQPVRKASRTGRCSLAPRQS
QARHVLPTPAHWGRRQVATQSILCTLGFSTRTLFFLCRNQ
SSSQVYLCYFFLPLPLSPDSLSSKRLQLLR*HELAPTVNV
VHDFLFKLCHLFPEHVGRGRQLGVLRFQGFHFFL*PGDPL
QLPFPALGGSDAVSHALPLRLDPLLRLHVDGGQGRRLSRH
LRHRLRLFLQRLQPRVLHRDLLGQVVRRLRRLLLLLGLRW
KGQPGRAX

Reading frame one is deemed to be the open reading frame.

Program: NCBI-blastp
Database: Swissprot
Title: CLONE-1
SeqLen: 184
Matrix: -M blosum62
Filter: -F f

Sequences producing significant alignments: (bits) Value

SW:AP1_RAT P17325 TRANSCRIPTION FACTOR AP-1 (PROTO-ONCOGENE) .
293 3e-79

SW:AP1_MOUSE P05627 TRANSCRIPTION FACTOR AP-1 (PROTO-ONCOGENE).
291 1e-78

SW:AP1_HUMAN P05412 TRANSCRIPTION FACTOR AP-1 (PROTO-ONCOGENE).
275 5e-74

>SW:AP1_RAT P17325 TRANSCRIPTION FACTOR AP-1 (PROTO-
ONCOGENE C-JUN).   Length = 334

Score = 293 bits (741),   Expect = 3e-79
Identities = 145/145 (100%),   Positives = 145/145   (100%)

```
Query:3       GATGLAFPSQPQQQQQPPQPPHHLPQQIPVQHPRLQALK    41
              GATGLAFPSQPQQQQQPPQPPHHLPQQIPVQHPRLQALK
Sbjct:190   YGATGLAFPSQPQQQQQPPQPPHHLPQQIPVQHPRLQALK    229

Query:42      EEPQTVPEMPGETPPLSPIDMESQERIKAERKRMRNRIAA    82
              EEPQTVPEMPGETPPLSPIDMESQERIKAERKRMRNRIAA
Sbjct:230   EEPQTVPEMPGETPPLSPIDMESQERIKAERKRMRNRIAA    270

Query:83      SKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVA    123
              SKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVA
Sbjct:271   SKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVA    310

Query: 124  QLKQKVMNHVNSGCQLMLTQQLQTF    147
              QLKQKVMNHVNSGCQLMLTQQLQTF
Sbjct: 311  QLKQKVMNHVNSGCQLMLTQQLQTF    334
```

The translation of clone 2 in all six reading frames is shown below.

Reading frame 1.

EQKAHLEDYYWMTGYPQQLNPEALGFSPEDAVEAXISNSH
QLQGGXXGYARGAQQLAAAAGAGAGASLGGSGEEMGPAAA
VVSAVIAAXAAQSGGAPGQLFLYQAATGKGNTPTAGAPGA
AGSASASASGAGGAGGG

Reading frame 2.

ANRRRTWKTTTG*PATRSSSTRRRWASARRTQSRRXSAIA
TSSRVAXXAMRGARSSWPRQRGPAPAPPWAAAARRWAPPP
PWCPP*SPRXPRRAAGHLTTITTTTTPRGTTTTRRPAPRE
PRVARPPLRAARVARAAV

Reading frame 3.

RTEGAPGRLLLDDRLPAAAQPGGAGLQPGGRSRGXHQQ*P
PAPGWXXWLCAGRAAAGRGSGGRRRRLPGRQRRGDGPRRR
RGVRRDRRGXRAERRGTSLPSPPPPRHGAPPPPDGRRPGS
RG*RVRLCERRGWRGRRW

Reading frame 4.

PPPPAPPAPLAEADALPAAPGAPAVGWWWCPVAWWWW*W*
*GAPPLCAXAAAITADTTAAAGPISSPLPPREAPAPAPAA
AASCCAPRA*PXXPPWSWWLLLMXASTASSGLKPSASGLS
CCG*PVIQ**SSRCAFCS

Reading frame 5.

HRRPRHPRRSQRRTRYPRLPGRRPSGGGGAPWRGGGGDGS
EVPRRSARXPRRSRRTPRRRRGPSPRRCRPGRRRRRPPLP
RPAAARPAHSXXSHPGAGGYC*XAPRLRPPG*SPAPPG*A
AAGSRSSSSSLPGAPSVR

187

Reading frame 6.

TAARATRAARRGGRATRGSRGAGRR*VVVPRGVVVVVMVV
RCPAALRGXRGDHGGHHGGGGAHLLAAAAQGGAGAGPRCR
GQLLRAPRIAXXATLELVAIADXRLDCVLRAEAQRLRVEL
LRVAGHPVVVFQVRLLFX

Reading frame 1 (underlined) represents the open reading frame and was used as the query sequence to search the database(s) for homology to other known transcription factors.

Title: CLONE-2

The results of the database search show that at present Clone 2 has not been identified and could possibly be a novel transcription factor.

EQKAHLEDYYWMTGYPQQLNPEALGFSPEDAVEAXISNSH

QLQGGXXGYARGAQQLAAAAGAGAGASLGGSGEEMGPAAA

VVSAVIAAXAAQSGGAPGQLFLYQAATGKGNTPTAGAPGA

AGSASASASGAGGAGGG

The translation of clone 3 in all six reading frames:

Reading frame 1.

EPSGSPVHVQLPQQAAPVTAAAAAPAAATSAPAPAPAPAA
PAAPAPAPAAAPAPAPAAQAVGWPICRDAYELQEVIGSGA
TAVVQAALCKPRQERVAIKRINLEKCQTSMDELLKEIQAM
SQCSHPNVVTYYTSFVVKDELWLVMKLLSGGSMLDIIKYI
VNRGX

Reading frame 2.

RSRAARPCTSSFPSRRPR*QPRRRLRRPRHQHRPRPRPRR
RPQPRPRLQLRLQPRPRQLRRSAGPSAGTRTSSRRLSAVE
RPPWFRQPCANPGKNA*P*SGSTWKSARRVWMNS*KKFKP
*ASAAIPT**LIILPLWSKMNFGWS*NY*VEVPCWISSNT
SSIGEX

Reading frame 3.

GAERLARARPASPAGGPGDSRGGGSGGRDISTGPGPGPGG
ARSPGPGSSCGSSPGPGSSGGRLAHLQGRVRAPGGYRQWS
DRRGSGSPVQTQARTRSHKADQLGKVPDEYG*TLKRNSSH
EPVQPSQRSDLLYFLCGQR*TLAGHEITKWRFHVGYHQIH
RQSGR

Reading frame 4.

SPRLTMYLMISNMEPPLSNFMTSQSSSLTTKEV**VTTLG
WLHWLMA*ISFKSSSILVWHFSKLIRFMATRSCLGLHRAA
*TTAVAPLPITSWSSYASLQMGQPTA*AAGAGAGAAAGAG
AGAAGAAGAGAGAGADVAAAGAAAAAVTGAACWGSWTCTG
EPLGSX

Reading frame 5.

LPD*RCI**YPTWNLHLVIS*PAKVHL*PQRKYNKSLRWD

GCTGSWLEFLLRVHPYSSGTFPS*SALWLRVLAWVCTGLP

EPRRSLHCR*PPGARTRPCRWASRPPELPGPGLEPQLEPG

PGLRAPPGPGPVLMSRPPEPPPRLSPGPPAGEAGRARA

SRSAP

Reading frame 6.

SPIDDVFDDIQHGTST**FHDQPKFIFDHKGSIISHYVGM

AALAHGLNFF*EFIHTRLALFQVDPLYGYAFLPGFAQGCL

NHGGRSTADNLLELVRVPADGPADRLSCRGRGWSRSWSRG

RGCGRRRGRGRGRC*CRGRRSRRRGCHRGRLLGKLDVHGR

AARLX

Reading frame 1 represents the open reading frame (underlined).

Program: NCBI-blastp
Database: Swissprot
Title: CLONE-3
SeqLen: 184
Matrix: -M blosum62
Filter: -F f

Sequences producing significant alignments: (bits) Value

SW:SPAK_RAT O88506 STE20/SPS1-RELATED PROLINE-ALA   9.9e-50

SW:SPAK_MOUSE Q9Z1W9 STE20/SPS1-RELATED PROLINE-A   6.1e-47

SW:SPAK_HUMAN Q9UEW8 STE20/SPS1-RELATED PROLINE-A   1.2e-43

SW:PAK4_HUMAN O96013 SERINE/THREONINE-PROTEIN KIN   9.4e-06

>>SW:SPAK_RAT O88506 STE20/SPS1-RELATED PROLINE-ALANINE
Length =553
Smith-Waterman score: 837; 100.000% identity (100.000% ungapped) in
164 aa overlap (1-164:3-166)

```
Query:   6    EPSGSPVHVQLPQQAAPVTAAAAAPAAATSAPAPA        35
              EPSGSPVHVQLPQQAAPVTAAAAAPAAATSAPAPA
Sbjct:   1    SPAMAEPSGSPVHVQLPQQAAPVTAAAAAPAAATSAPAPA    40


Query:  36    PAPAAPAAPAPAPAAAPAPAPAAQAVGWPICRDAYELQEV    75
              PAPAAPAAPAPAPAAAPAPAPAAQAVGWPICRDAYELQEV
Sbjct:  41    PAPAAPAAPAPAPAAAPAPAPAAQAVGWPICRDAYELQEV    80


Query:  76    IGSGATAVVQAALCKPRQERVAIKRINLEKCQTSMDELLK   115
              IGSGATAVVQAALCKPRQERVAIKRINLEKCQTSMDELLK
Sbjct:  81    IGSGATAVVQAALCKPRQERVAIKRINLEKCQTSMDELLK   120


Query: 116    EIQAMSQCSHPNVVTYYTSFVVKDELWLVMKLLSGGSMLD   155
              EIQAMSQCSHPNVVTYYTSFVVKDELWLVMKLLSGGSMLD
Sbjct: 121    EIQAMSQCSHPNVVTYYTSFVVKDELWLVMKLLSGGSMLD   160


Query: 156    IIKYIVNRGX                                165
              IIKYIVNRGX
Sbjct: 161    IIKYIVNRGXEHKNGVLEEAIIATILKEVLEGLDYLHRNG   200
```

The translation of clone 4 in all six reading frames:

Reading frame 1.

TAKMETTFYDDALNASFLQSESGAYGYSNPKILKQSMTLN
LADPVGNLKPHLRAKNSDLLTSPDVGLLKLASPELERLII
QSSNGHITTTPTPTQFLCPKNVTDEQEGFAEGFVRALAEL
HSQNTLPSVTSAAQPVSGAGMVAPAVASVAGAGGGGGYSA
SLHSEPPVYANLSNFNPGALSSGGGAPSYGATGLAFPSQP
QQQQQPPQPPHHLPQQIPVQHPRLQALKEEPQTVPEMPGE
TPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIA
RLEEKVKTLKAQLXNSELASTANMLREQVAQLKQKVMN*P
RLFKAPDRAGPTNVSASGVLTLRSLERTGEEGAGGSTAVW
SARS*TNFVTEAGTRGYPPLPGALLRPRNFCAQPRLTPRE
VTDCSM

Reading frame 2.
GEPLSLGLVRGLRGTAAS*ERSVSDRDFSKLRIARDQPT*
VQAVS*PCAPWSELGRRAQGGALRSGAHAPKQTLLLKRGR
AGIPRFPARCCGPETSAHSPG*PRVK*PTVL*LQRWKRPS
TTMPSTPRSSSPRVAPTATVTLRF*SRA*P*TWPTRWAI*
SRTSEPRTRTFSRRPTSGCSSWRRRSWSA*SSSPAMGTSP
LHRPPLSSCAPRT*PTSRRASPKASCAP*LNCIARIRCPV
SPPRHNLSVGRAWSLPLWPQ*LALAAAAATAPACTVSLRS
TPTSATSTRVR*AAAVGRPPMARPGWPFHRSPSSSSSRLS
RRTTCPNRSRCSTRGCRR*RKSRRRCRRCRERRRPCPPST
WSLRSGSRRRGSA*ETASLPPSAGKGSWSGSPG*RKK*KP
KRN

Reading frame 3.
ENLCRWGWSAGSEEPLLPESAP*VTATFQSSGSRGTNQRE
CKRCLNPALPGANWGGGRRGEHCGLERTLLNKLCY*SGDA
RVSPASRRAVAAPKLLRTAQANPA*SDRLFYDCKDGNDLL
RRCPQRLVPPVREWRLRLQ*P*DSEAEHDLEPGRPGGQSE
AAPPSQELGPSHVARRRAAQAGVAGAGAPDHPVQQWAHHH
YTDPHSVLVPQERDRRAGGLRRRLRARPS*TA*PEYAAQC
HLRGTTCQWGGHGRSRCGLSSWRWRRRLQRQPAQ*ASGL
RQPQQLQPGCAEQRRWGALLWRDRAGLSIAAPAAAAASA
AAPLAPTDPGAAPAAAGAEGRAADGAGDAGRDAALVPHRH
GVSGADQGGEEAHEKPHRCLQVPEKEAGADRPARGKSENL
ESAIX

Reading frame 4.

LIALSRFSLFPLAGRSAPASFSGTWRQRCGFSCASSPP*S
APETPCRWGTRAASLPASPAPSAALPSAPAAAGAAPGSVG
ASGAAAEAAAAAGAAMERPARSRHRRAPHRRCSAHPG*S
C*GWRRPEAHCAGWRCSRRRRQRQLLRPQRERPCPPH*QV
VPRR*HWAAYSGYAVQLGRARSLRRSPPARRSRSWGTRTE
WGSV*W*CAHCWTG*SGAPAPATPA*AARRRAT*EGPSSW
LGGAASDCPPGRPGSRSCSASES*GYCSRRRHSRTGGTRR
*GHRRRRSFPSLQS*NSRSLHAGLAWAVRRSFGAATARRE
AGDTRASPLQ*QSLFRSVRSRPQCSPLRPPPQFAPGSAGL
RHRLHSRWLVPRDPEL*KVAVTHGALSGSSGSSEPADQPQ
RQRFS

Reading frame 5.

*LRFQGFHFFL*PGDPLQLPFPALGGSDAVSHALPLRLDP
LLRLHVDGGQGRRLSRHLRHRLRLFLQRLQPRVLHRDLLG
QVVRRLRRLLLLLGLRWKGQPGRAIGGRPTAAAQRTRVEV
AEVGVDRRLTVQAGAVAAAASASY*GHSGSDHARPTDRL
CRGGDTGQRILAMQFS*GAHEAFGEALLLVGHVLGAQELS
GGRCSGDVPIAGLDDQALQLRRRQLEQPDVGRREKVRVLG
SEVRLQIAHRVGQVQGHALLQNLRVTVAVGATLGLEERGV
EGIVVEGRFHLCSHRTVGHFTRG*PGLCAEVSGPQQRAGK
RGIPARPRFSNKVCLGACAPDRSAPPCALLPSSLQGAQG*
DTACTHVGWSRAIRSFEKSRSLTERSQEAAVPRSPRTSPS
DRGSX

Reading frame 6.

NCAFKVFTFSSSRAIRSSFLFRHLEAAMRFLMRFLSALIR
S*DSMSMGDKGGVSPGISGTVCGSSFSACSRGCCTGICWG
KWCGG*GGCCCCWGCDGKASPVAP*EGAPPPLLSAPGLKL
LRLA*TGGSLCRLAL*PPPPPAPATEATAGATMPAPLTGC
AAEVTLGSVFWLCSSARARTKPSAKPSCSSVTFLGHKN*V
GVGVVVMCPLLDWMIRRSSSGDASLSSPTSGDVRRSEFLA
RRCGFRLPTGSARFKVMLCFRILGLL*P*APLSDWRNEAL
RASS*KVVSIFAVIEQSVTSRGVSLGCAQKFRGRNSAPGS
GGYPRVPASVTKFV*ERALQTAVLPPAPSSPVRSRERRVK
TPLALTLVGPARSGALKSRGHSRSALRKQRFLGARGPAPA
TEVLX

Program: NCBI-blastp
Database: Swissprot
Title: CLONE-4
SeqLen: 430
Matrix: -M blosum62
Filter: -F f

Sequences producing significant alignments: (bits) Value

SW:AP1_RAT P17325 TRANSCRIPTION FACTOR AP-1 (PROTO-ONCOGENE)
591 e-169

SW:AP1_MOUSE P05627 TRANSCRIPTION FACTOR AP-1 (PROTO-ONCOGENE)...
587 e-167

SW:AP1_HUMAN P05412 TRANSCRIPTION FACTOR AP-1 (PROTO-ONCOGENE)...
559 e-159

>SW:AP1_RAT P17325 TRANSCRIPTION FACTOR AP-1 (PROTO-
ONCOGENE C-JUN).   Length =334

 Score = 591 bits (1507), Expect = e-169
 Identities = 293/293 (100%), Positives = 293/293 (100%)

Query: 129   TAKMETTFYDDALNASFLQSESGAYGYSNPKILKQSMTL   167
             TAKMETTFYDDALNASFLQSESGAYGYSNPKILKQSMTL
Sbjct: 1   MTAKMETTFYDDALNASFLQSESGAYGYSNPKILKQSMTL   40

Query: 168   NLADPVGNLKPHLRAKNSDLLTSPDVGLLKLASPELERLI   207
             NLADPVGNLKPHLRAKNSDLLTSPDVGLLKLASPELERLI
Sbjct: 41   NLADPVGNLKPHLRAKNSDLLTSPDVGLLKLASPELERLI   80

Query: 208   IQSSNGHITTTPTPTQFLCPKNVTDEQEGFAEGFVRALAE   247
             IQSSNGHITTTPTPTQFLCPKNVTDEQEGFAEGFVRALAE
Sbjct: 81   IQSSNGHITTTPTPTQFLCPKNVTDEQEGFAEGFVRALAE   120

Query: 248   LHSQNTLPSVTSAAQPVSGAGMVAPAVASVAGAGGGGGYS   287
             LHSQNTLPSVTSAAQPVSGAGMVAPAVASVAGAGGGGGYS
Sbjct: 121   LHSQNTLPSVTSAAQPVSGAGMVAPAVASVAGAGGGGGYS   160

Query: 288   ASLHSEPPVYANLSNFNPGALSSGGGAPSYGATGLAFPSQ   327
             ASLHSEPPVYANLSNFNPGALSSGGGAPSYGATGLAFPSQ
Sbjct: 161   ASLHSEPPVYANLSNFNPGALSSGGGAPSYGATGLAFPSQ   200

```
Query: 328  PQQQQQPPQPPHHLPQQIPVQHPRLQALKEEPQTVPEMPG   367
             PQQQQQPPQPPHHLPQQIPVQHPRLQALKEEPQTVPEMPG
Sbjct: 201  PQQQQQPPQPPHHLPQQIPVQHPRLQALKEEPQTVPEMPG   240


Query: 368  ETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERI   407
             ETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERI
Sbjct: 241  ETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERI   280


Query: 408  ARLEEKVKTLKAQ   420
             ARLEEKVKTLKAQ
Sbjct: 281  ARLEEKVKTLKAQ   293
```

The translation of clone 5:

Reading frame 1.

SAALPLRMHHSAAEGLPASNHATNVMSTATGLHYSVPSCH
YGNQPSTYGVMAGTLTPCLYKFPDHTLSHGFPPLHQPLLA
EDPTASEFKQELRRKSKLVEEPIDMDSPEIRELEQFANEF
KVRRIKLGYTQTNVGEALAAVHGSEFSQTTICRFENLQLS
FKNACKLKAILSKWLEEAEQVGALYNEKVGANERKRKRRT
TISIAAKDALERHFGEHSKPSSQEIMRMAEELNLEKEVVR
VWFC

Reading frame 2.

FCCPASENAPQCR*GSPSLQPRHQRDVHSDRTSLFCAFLS
LWKPAIHLRSDGRHFNPLSLQVSRPHPESWVSSPAPTSAG
GGPNSL*I*AGTQAEK*IGGRANRHGLPGNPRTGAVCQRI
*SEKN*VRIHPDKRGRSSGRCPRLRIQSNNHLPI*KLAAQ
FQKCLQTESNFIQVAGGS*AGRSFVQ*KSGSKRKEEETED
NYQYRR*GCFGETLWRAQQTFFAGDHADG*RIESRERSSK
SVVLQX

Reading frame 3.

LLPCL*ECTTVPLRVSQPPTTPPT*CPQRQDFIILCLPVI
METSHPPTE*WQAL*PLVFTSFQTTP*VMGFLPCTNLCWR
RTQQPLNLSRNSGGKVNWWKSQ*TWTPRKSENWSSLPTNL
K*EELS*DTPRQTWAKLWPLSTAPNSVKQPSADLKTCSSV
SKMLAN*KQFYPSGWRKLSRSELCTMKKWEQTKGRGNGGQ
LSVSPLRMLWRDTLESTANLLRRRSCGWLKN*ISRKK**E
CGFAX

Reading frame 4.

LAKPHSYYFFLEIQFFSHPHDLLRRRFAVLSKVSLQSILS
GDTDSCPPFPLPFVCSHFFIVQSSDLLSFLQPLG*NCFQF
ASIFETELQVFKSADGCLTEFGAVDSGQSFAHVCLGVS*L
NSSHFKFVGKLLQFSDFRGVHVYWLFHQFTFPPEFLLKFR
GCWVLRQQRLVQGRKPMTQGVVWKLVKTRG*SACHHSVGG
WLVSIMTGRHRIMKSCRCGHHVGGVVGGWETLSGTVVHSQ
RQGSRX

Reading frame 5.

LQNHTLTTSFSRFNSSAIRMISCEEGLLCSPKCLSKASLA
AILIVVLRFLFLSFAPTFSLYKAPTCSASSSHLDKIAFSL
QAFLKLSCKFSNRQMVV*LNSEPWTAARASPTFVWVYPNL
ILLTLNSLANCSSSRISGESMSIGSSTNLLFRLSSCLNSE
AVGSSASRGWCRGGNP*LRVWSGNL*RQGVKVPAITP*VD
GWFP**QEGTE**SPVAVDITLVAWLEAGRPSAALWCILR
GRAAE

Reading frame 6.
```
CKTTLLLLLSRDSILQPSA*SPAKKVCCALQSVSPKHP*R
RY**LSSVSSSFRLLPLFHCTKLRPAQLPPATWIKLLSVC
KHF*N*AASFQIGRWLFD*IRSRGQRPELRPRLSGCILT*
FFSL*IRWQTAPVLGFPGSPCLLALPPIYFSA*VPA*IQR
LLGPPPAEVGAGEETHDSGCGLETCKDKGLKCLPSLRRWM
AGFHNDRKAQNNEVLSLWTSRWWRGWRLGDPQRHCGAFSE
AGQQX
```

Reading frame 1 (underlined) is deemed to be the open reading frame.

Program: NCBI-blastp
Database: Swissprot
Title: CLONE-5
SeqLen: 249
Matrix: -M blosum62
Filter: -F f

Sequences producing significant alignments: (bits) Value

SW:HMP1_MOUSE Q00286 PITUITARY-SPECIFIC POSITIVE TRANSCRIPTION
494 e-139

SW:HMP1_RAT P10037 PITUITARY-SPECIFIC POSITIVE TRANSCRIPTION
487 e-137

SW:HMP1_HUMAN P28069 PITUITARY-SPECIFIC POSITIVE TRANSCRIPTION
481 e-136

SW:HMP1_BOVIN P10036 PITUITARY-SPECIFIC POSITIVE TRANSCRIPTION
470 e-132

>SW:HMP1_RAT P10037 PITUITARY-SPECIFIC POSITIVE
TRANSCRIPTION FACTOR 1 (PIT-1)
        (GROWTH HORMONE FACTOR 1) (GHF-1).

 Score = 487 bits (1239), Expect = e-137
 Identities = 244/272 (89%), Positives = 246/272 (89%), Gaps = 26/272 (9%)


```
Query: 1     SSAALPLRMHHSAAEGLPASNHATNVMST----------     30
             SAALPLRMHHSAAEGLPASNHATNVMST
Sbjct: 18    DASAALPLRMHHSAAEGLPASNHATNVMSTVPSILSLIQT     57


Query: 31    ---------------ATGLHYSVPSCHYGNQPSTYGVMA     54
                            ATGLHYSVPSCHYGNQPSTYGVMA
Sbjct: 58    PKCLHTYFSMTTMGNTATGLHYSVPSCHYGNQPSTYGVMA     97
```

```
Query:  55  GTLTPCLYKFPDHTLSHGFPPLHQPLLAEDPTASEFKQEL    94
            GTLTPCLYKFPDHTLSHGFPPLHQPLLAEDPTASEFKQEL
Sbjct:  98  GTLTPCLYKFPDHTLSHGFPPLHQPLLAEDPTASEFKQEL   137


Query:  95  RRKSKLVEEPIDMDSPEIRELEQFANEFKVRRIKLGYTQT   134
            RRKSKLVEEPIDMDSPEIRELEQFANEFKVRRIKLGYTQT
Sbjct: 138  RRKSKLVEEPIDMDSPEIRELEQFANEFKVRRIKLGYTQT   177


Query: 135  NVGEALAAVHGSEFSQTTICRFENLQLSFKNACKLKAILS   154
            NVGEALAAVHGSEFSQTTICRFENLQLSFKNACKLKAILS
Sbjct: 178  NVGEALAAVHGSEFSQTTICRFENLQLSFKNACKLKAILS   217


Query: 155  KWLEEAEQVGALYNEKVGANERKRKRRTTISIAAKDALER   194
            KWLEEAEQVGALYNEKVGANERKRKRRTTISIAAKDALER
Sbjct: 218  KWLEEAEQVGALYNEKVGANERKRKRRTTISIAAKDALER   257

Query: 195  HFGEHSKPSSQEIMRMAEELNLEKEVVRVWFC 246
            HFGEHSKPSSQEIMRMAEELNLEKEVVRVWFC
Sbjct: 258  HFGEHSKPSSQEIMRMAEELNLEKEVVRVWFC 289
```

The translation of clone 6:

Reading frame 1.

LQEAQDGVLDLKAAADTLAVRQKRRIYDITNVLEGIDLIE
KKSKNSIQWKGVGAGCNTKEVIDRLRCLKAEIEDLELKER
ELDQQKLWLQQSIKNVMEDSINNRFSYVTHEDICSCFNGD
TLLAIQAPSGTQLEVPIPEMGQNGQKKYQINLKSHSGPIH
VLLINKESNSSKPVVFPVPPPDDLTQPSSQSSTSVTPPKS
TMAAQNLPEQHVSERSQNFQQTPATEISSGSISGDIIDEL
MSSDX

Reading frame 2.
CCRRRRTASWISKRLQILWL*GKSEEFMISPTS*KELI*L
KKNQRTVSSGRV*VLAVILKKL*ID*DVLKLKLKI*N*RK
ENLTSRSCGYSKASKT*WKTPLITDFLM*LTKTSAVALMV
IHCWPFRHPLVHSWKYLFQKWDRMDKRNTR*I*RVTQDQS
MCYL*IKSPTHLSRWFFLSPHLMTSHNLPPSPQLQ*LHRN
PPWLLKTCPSSMFLKEARISSRHQLQKYLQDLLVETSLMN
*CLLTX

Reading frame 3.
AAGGAGRRPGSQSGCRYFGCEAKAKNL*YHQRLRRN*SN*
KKIKEQYPVEGCRCWL*Y*RSYR*IKMS*S*N*RSRIEGK
RT*PAEVVATAKHQKRDGRLH**QIFLCNSRRHLQLL*W*
YTVGHSGTLWYTAGSTYSRNGTEWTKEIPDKSEESLRTNP
CATYK*RVQLI*AGGFSCPPT**PHTTFLPVLNFSDSTEI
HHGCSKPARAACF*KKPEFPADTSYRNIFRIY*WRHH**T
DVF*R

Reading frame 4.
SSEDISSSMMSPLIDPEDISVAGVCWKFWLLSETCCSGRF
*AAMVDFGGVTEVEDWEEGCVRSSGGGTGKTTGLDELDSL
FISSTWIGPE*LFRFIWYFFCPFCPISGIGTSSCVPEGA*
MANSVSPLKQLQMSS*VT*ENLLLMESSITFLMLCCSHNF
CWSSSLSFNSRSSISALRHLNLSITSLVLQPAPTPFHWIL
FFDFFSIRSIPSKTLVIS*ILRFCLTAKVSAAALRSRTPS
CASCSX

Reading frame 5.
RQKTSVHQ*CLH**ILKIFL*LVSAGNSGFFQKHAARAGF
EQPWWISVESLKLRTGRKVV*GHQVGGQEKPPA*MSWTLY
L*VAHGLVLSDSSDLSGISFVHSVPFLE*VLPAVYQRVPE
WPTVYHH*SNCRCLRELHKKICY*WSLPSRF*CFAVATTS
AGQVLFPSILDLQFQL*DILIYL*LL*YYSQHLHPSTGYC
SLIFFQLDQFLLRRW*YHKFFAFASQPKYLQPL*DPGRRP
APPAA

199

Reading frame 6.
VRRHQFINDVSTNRS*RYFCSWCLLEILASFRNMLLGQVL
SSHGGFRWSH*S*GLGGRLCEVIRWGDRKNHRLR*VGLFI
YK*HMDWS*VTLQIYLVFLLSILSHFWNRYFQLCTRGCLN
GQQCITIKATADVFVSYIRKSVINGVFHHVFDALL*PQLL
LVKFSFLQF*IFNFSFKTS*SIYNFFSITASTYTLPLDTV
L*FFFN*INSF*DVGDIINSSLLPHSQSICSRFEIQDAVL
RLLQX

Reading frame 1 (underlined) was deemed to be the open reading frame.

Program: NCBI-blastp
Database: Swissprot
Title: CLONE-6
SeqLen: 245
Matrix: -M blosum62
Filter: -F f

Sequences producing significant alignments: (bits) Value

SW:E2F5_RAT Q62814 TRANSCRIPTION FACTOR E2F5 (E2F   (300) 1246 401
1.2e-111

SW:E2F5_MOUSE Q61502 TRANSCRIPTION FACTOR E2F5 (E   (335) 1213 391
1.8e-108

SW:E2F5_HUMAN Q15329 TRANSCRIPTION FACTOR E2F5 (E   (346) 1156 373
4.5e-103

>>SW:E2F5_RAT Q62814 TRANSCRIPTION FACTOR E2F5 (E2F-5)
Length=300
Smith-Waterman score: 1246; 100.000% identity (100.000% ungapped) in
244 aa overlap (1-244:21-264)

```
Query:  21                       LQEAQDGVLDLKAAADTLAV         40
                                 LQEAQDGVLDLKAAADTLAV
Sbjct:   1  GGSSRHEKSLGLLTTKFVSLLQEAQDGVLDLKAAADTLAV         40


Query:  41  RQKRRIYDITNVLEGIDLIEKKSKNSIQWKGVGAGCNTKE         80
            RQKRRIYDITNVLEGIDLIEKKSKNSIQWKGVGAGCNTKE
Sbjct:  41  RQKRRIYDITNVLEGIDLIEKKSKNSIQWKGVGAGCNTKE         80


Query:  81  VIDRLRCLKAEIEDLELKERELDQQKLWLQQSIKNVMEDS        120
            VIDRLRCLKAEIEDLELKERELDQQKLWLQQSIKNVMEDS
Sbjct:  81  VIDRLRCLKAEIEDLELKERELDQQKLWLQQSIKNVMEDS        120
```

```
Query:  121  INNRFSYVTHEDICSCFNGDTLLAIQAPSGTQLEVPIPEM     160
             INNRFSYVTHEDICSCFNGDTLLAIQAPSGTQLEVPIPEM
Sbjct:  121  INNRFSYVTHEDICSCFNGDTLLAIQAPSGTQLEVPIPEM     160


Query:  161  GQNGQKKYQINLKSHSGPIHVLLINKESNSSKPVVFPVPP     200
             GQNGQKKYQINLKSHSGPIHVLLINKESNSSKPVVFPVPP
Sbjct:  161  GQNGQKKYQINLKSHSGPIHVLLINKESNSSKPVVFPVPP     200


Query:  201  PDDLTQPSSQSSTSVTPPKSTMAAQNLPEQHVSERSQNFQ     241
             PDDLTQPSSQSSTSVTPPKSTMAAQNLPEQHVSERSQNFQ
Sbjct:  201  PDDLTQPSSQSSTSVTPPKSTMAAQNLPEQHVSERSQNFQ     241


Query:  241  QTPATEISSGSISGDIIDELMSSD                 264
             QTPATEISSGSISGDIIDELMSSD
Sbjct:  241  QTPATEISSGSISGDIIDELMSSDVFPLLRLSPTPADDYN     280
```

The translation of clone 7:

Reading frame 1.

QMPLDLSPLATPIIRSKIEEPSVVETTHQDSPLPHPESTT
NDEKEIPLAQTAQPTSAIVRPASLQVPNVLLTSSDXSVII
QQAVPSPTSXTVITQAPSSNRPIVPVPGPXPLLLHLPNGQ
TMPVAIPASITSSNVHVPAAVPLVRPVTMVPSVPGIPGXS
SP

Reading frame 2:

KCL*ICPLLQHPS*EAKLRNLLL*KQLTRIARYLTPSLLP
MMKRKYHWHKLHSPHQLSFVQHHYRFPMCCSQVLTXV*LF
NKQYLHQPXVL*SPRHHPLTGQLFLYQAHXLFYYIFLMDK
PCPLLFLHQLQVLMCMFQLQSHLFGQSPWCLVFQESQAXP
LL

Reading frame 3.

NASRSVPSCNTHHKKQN*GTFCCRNNSPG*PVTSPRVYYQ
**KGNTIGTNCTAHISYRSSSIITGSQCAAHKF*XKCNYS
TSSTFTNLXYCNHPGTIL*QANCSCTRPXSSSITSS*WTN
HARCYSCINYKF*CACSSCSPTCSASHHGA*CSRNPRXFL
SX

Reading frame 4:

QERXGLGFLEH*APW*LAEQVGLQLEHAH*NL*LMQE*QR
AWFVH*EDVIEEXMGLVQEQLAC*RMVPG*LQYXRLVKVL
LVE*LHXSQNL*AAHWEPVMMLDER*LMWAVQFVPMVFPF
HHW**TRGEVTGYPGELFLQQKVPQFCFL*WVLQEGTDLE
AF

Reading frame 5:

RRGXAWDSWNTRHHGDWPNKWDCSWNMHIRTCN*CRNSNG
HGLSIRKM**KRXWAWYRNNWPVRGWCLGDYSX*GW*RYC
LLNNYTXVRTCEQHIGNL**CWTNDS*CGLCSLCQWYFLF
IIGSRLGVR*RAILVSCFYNRRFLNFASYDGCCKRGQI*R
HX

Reading frame 6:

GEEXPGIPGTLGTMVTGRTSGTAAGTCTLELVIDAGIATG
MVCPLGRCNRRGXGPGTGTIGLLEDGAWVITVXEVGEGTA
C*IITLXSELVSSTLGTCNDAGRTIADVGCAVCANGISFS
SLVVDSG*GNGLSW*VVSTTEGSSILLLMMGVARGDRSRG
IX

Reading frame 1 (underlined) was deemed to be the open reading frame.

202

Program: NCBI-blastp
Database: Swissprot
Title: CLONE-7
SeqLen: 162
Matrix: -M blosum62
Filter: -F f

Sequences producing significant alignments: (bits) Value

SW:ATF2_RAT Q00969 CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR
312 2e-85

SW:ATF2_MOUSE P16951 CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR
310 1e-84

SW:ATF2_HUMAN P15336 CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR
308 5e-84

SW:ATF2_CHICK O93602 CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR
302 3e-82

>SW:ATF2_RAT Q00969 CYCLIC-AMP-DEPENDENT TRANSCRIPTION
FACTOR ATF-2 (ACTIVATING TRANSCRIPTION FACTOR 2)
       (CAMP RESPONSE ELEMENT BINDING PROTEIN CRE- BP1).
       Length = 487

 Score = 312 bits (791), Expect = 2e-85
 Identities = 157/161 (97%), Positives = 157/161 (97%)

```
Query: 1    MPLDLSPLATPIIRSKIEEPSVVETTHQDSPLPHPESTTN      40
            MPLDLSPLATPIIRSKIEEPSVVETTHQDSPLPHPESTTN
Sbjct: 89   MPLDLSPLATPIIRSKIEEPSVVETTHQDSPLPHPESTTN      128


Query: 41   DEKEIPLAQTAQPTSAIVRPASLQVPNVLLTSSDXSVIIQ      80
            DEKEIPLAQTAQPTSAIVRPASLQVPNVLLTSSD SVIIQ
Sbjct: 129  DEKEIPLAQTAQPTSAIVRPASLQVPNVLLTSSDSSVIIQ      168


Query: 81   QAVPSPTSXTVITQAPSSNRPIVPVPGPXPLLLHLPNGQT 120
            QAVPSPTS TVITQAPSSNRPIVPVPGP PLLLHLPNGQT
Sbjct: 169  QAVPSPTSSTVITQAPSSNRPIVPVPGPFPLLLHLPNGQT 208

Query: 121  MPVAIPASITSSNVHVPAAVPLVRPVTMVPSVPGIPGXSSP 161
            MPVAIPASITSSNVHVPAAVPLVRPVTMVPSVPGIPG SSP
Sbjct: 209  MPVAIPASITSSNVHVPAAVPLVRPVTMVPSVPGIPGPSSP 249
```

From the results of the database searches clones 1 and 4 both show sequence homology to the transcription factor AP-1 (proto-oncogene c-jun), therefore, the next step was to search for an overlapping sequence between the two clones. To achieve this it was necessary to use the BLAST2 multiple sequence alignment tool. This tool produces the alignment of two given sequences using the BLAST2 search engine for local alignment (Tatusova T.A. and Madden T.L. 1999).

Interestingly the DNA sequence of the *MBP* promoter region located between nucleotides –609 to –577 resembles the AP-1 binding site.

```
   -609                                    -577
5′  CACCTTTTGTCAAACGACCGCTTCACATCTGGG 3′
          :    : : : :
5′          TGAGTCAG 3′      AP-1 binding site
```

BLAST 2 SEQUENCES: RESULTS [VERSION BLASTP 2.1.2]

Sequence 1 lcl|seq_1   **Length**   252   (1 .. 252)

Sequence 2 lcl|seq_2   **Length**   433   (1 .. 433)
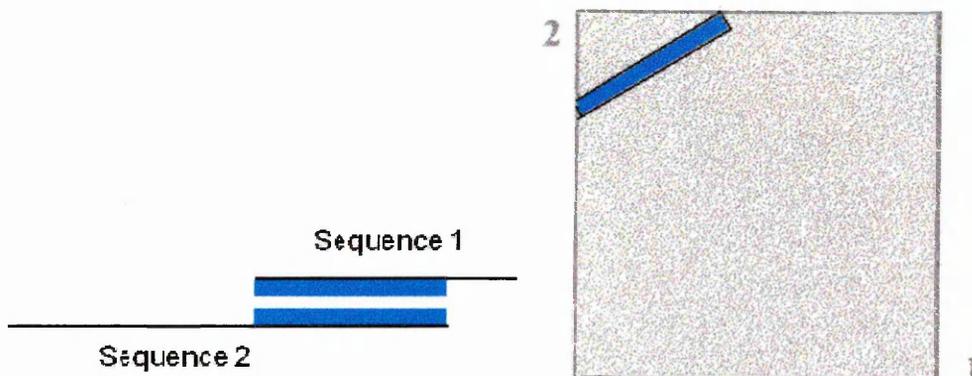
Score = 139 bits (351), Expect = 2e-32
Identities = 104/104 (100%), Positives = 104/104 (100%)

204

```
Query:    2    GATGLAFXXXXXXXXXXXXXXXXXXXXXXXXXXXRLQALK      42
               GATGLAF                          RLQALK
Sbjct:  317    YGATGLAFPSQPQQQQQPPQPPHHLPQQIPVQHPRLQALK    356


Query:   42    EEPQTVPEMPGETPPLSPIDMESQERIKAERKRMRNRIAA      81
               EEPQTVPEMPGETPPLSPIDMESQERIKAERKRMRNRIAA
Sbjct:  357    EEPQTVPEMPGETPPLSPIDMESQERIKAERKRMRNRIAA    396


Query:   82    SKCRKRKLERIARLEEKVKTLKAQ      105
               SKCRKRKLERIARLEEKVKTLKAQ
Sbjct:  397    SKCRKRKLERIARLEEKVKTLKAQ      420
```

```
Gapped
Lambda       K          H
0.318        0.133      0.389

Gapped
Lambda       K          H
0.267        0.0410     0.140
```

Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 711
Number of Sequences: 0
Number of extensions: 47
Number of successful extensions: 1
Number of sequences better than 10.0: 1
Number of HSP's better than 10.0 without gapping: 1
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 0
Number of HSP's gapped (non-prelim): 1
length of query: 252
length of database: 187,523,045
effective HSP length: 56
effective length of query: 196
effective length of database: 145,851,261
effective search space: 28586847156
effective search space used: 28586847156

The results of the BLAST2 search above show that clones 1 and 4

contain an overlapping sequence, and since the alignment indicates a high

level of sequence similarity and a correspondingly low E-value (Expect = 2e-

32) this would imply that the alignment is unlikely to be a random match. In

205

the data shown above the program SEG has been implemented as part of the BLAST routine. The purpose of the SEG program is to mask out low complexity regions (i.e. regions within a sequence which have high densities of particular residues, e.g. `PSQPQQQQQPPQPPHHLPQQIPVQHP`), which would otherwise result in large numbers of spurious high-scoring matches swamping the 'hitlist'. Regions masked by SEG are denoted by strings of Xs in the query sequence.

Clones 1 and 4 have been identified as the transcription factor *c-jun*, which is capable of stimulating transcription via an AP-1 site. The AP-1/*jun* family of transcription factors bind to the consensus nucleotide sequence TGAGTCAG found in the enhancers of many cellular and viral genes. *c-jun* is a proto-oncogene which when mutated becomes an oncogene. Implicit in the growth functions of all oncogenes is the potential to induce abnormal cell growth and cancer as a result of deregulated gene expression (Vogt and Bos, 1990). As well as Jun, purified preparations of AP-1 contain several different proteins e.g. Jun-related proteins, *c-fos* and Fos-related proteins known as *fras* (fos-related antigens) (Karin et al., 1997). Unlike Jun, which can form a DNA-binding homodimer, Fos cannot bind to the AP-1 site as a homodimer. However, Fos can bind (with approximately 30 fold greater affinity than the Jun homodimer alone) to AP-1 sites in the presence of another nuclear protein, p39, which has been identified as Jun (Rauscher et al., 1988). The p39 protein is also identical to yeast protein GCN4 (Hattori *et al.*, 1988). The strong homology of the Jun and GCN4 proteins suggests they recognise very similar DNA-binding sequences (e.g. GCN4 binds to the

consensus sequence 5'-TGA(C/G)TCA-3' and Jun binds to the consensus

sequence 5'-TGAGTCAG-3' respectively). It has been proposed by Vogt and

Bos (1990) that the C-terminus of the Jun protein (located between amino

acid residues 260 to 311) is the DNA-binding domain. Within this region two

features stand out: the basic domain located between amino acid residues

260 to 279) and the leucine zipper (located between amino acid residues

283 to 311). The zipper itself consists of a periodic repetition of leucine

residues at every seventh position, and regions containing them appear to

span 8 turns of alpha-helix. The leucine side chains that extend from one

helix interact with those from a similar helix, thus facilitating dimerisation,

which in turn results in the correct positioning of the two basic DNA-binding

domains in the dimeric molecule for DNA binding to occur. Leucine zippers

are present in many gene regulatory proteins, including the CREB proteins,

Jun, Fos oncogene and Fos-related proteins (Fras), C-myc, L-myc and N-

myc oncogenes. Promoters containing the AP-1 binding site sequence are

inducible by treatment with phorbol esters which appear to increase DNA

binding of Jun by activating protein kinase C. However, it has been shown

that increased DNA-binding of Jun is mediated by a dephosphorylation at

three specific sites (one threonine and two serine) following phorbol ester

treatment (Boyle et al., 1991). This indicates protein kinase C acts by

stimulating a phosphatase enzyme, which in turn dephosphorylates Jun.

Such an inhibitory effect of phosphorylation on the activity of a transcription

factor is not unique to Jun. Interestingly, there are two amino acid residues

(both serine) located at amino acid positions 63 and 73 respectively which

can be modified by phosphorylation by JNK1 which inhibits transcriptional

activity (Boyle *et al.*, 1991).

Clone 3 has been identified as the PASK protein (<u>p</u>roline-<u>a</u>lanine rich

<u>S</u>TE20-related <u>k</u>inase). It is thought that this protein may act as a mediator of

stress-activated signals (Ushiro *et al.*, 1998). The PAPA box (proline-alanine

repeats) may target the kinase to a specific subcellular location by

facilitating the interaction with intracellular proteins such as actin or actin-like

proteins. Sequence homology puts the PASK protein into the serine /

threonine family of protein kinases (STE20 subfamily). According to the

Ushiro group PASK is preferentially expressed in cells rich in Na^+, K^+-

ATPase and active in ion transport (i.e. those cells maintaining an

electrochemical sodium and potassium gradient across the plasma

membrane). Sodium-coupled active transport relies on these gradients within

the cell (e.g. in neurons, Na^+, K^+-ATPase is active to maintain their resting

membrane potential (Albers *et al.*, 1994).

Clone 5 has been identified as the pituitary specific positive

transcription factor 1 (Pit-1), growth hormone factor-1 (GHF-1) or general

transcription factor 2i. The transcription factor Pit-1 plays a crucial role in

pituitary-specific gene expression, and belongs to the homeobox / POU

family of transcription factors. The POU-domain located between amino acid

residues 154 to 224 is a 70 to 75 amino acid region found upstream of a

homeobox domain in some transcription factors. Such proteins bind to

specific DNA sequences to cause spatial and temporal regulation of the

expression of genes, many of which are involved in the regulation of

neuronal development in the central nervous system of mammals (Johnson

208

and Hirsh, 1990). Other genes are also regulated by the POU-domain, including those for the immunoglobulin light and heavy chains (Oct-2) (Petryniak *et al.*, 1990), and trophic hormone genes, such as those for prolactin and growth hormone (Pit-1). The POU-domain may also be involved in protein-protein interactions (Mathis *et al.*, 1992). The 3-dimensional structure of the POU-domain has been determined by multidimensional nuclear magnetic resonance (Assa-Munt *et al.*, 1993) and X-ray crystallography (Klemm *et al.*, 1994).

The homeobox (DNA-binding domain) located between residues 240 to 299 is a 60 amino acid region, first identified in a number of Drosophila segmentation and homeotic proteins, is known to be well conserved in many other species (Gehring, 1992). Proteins containing homeobox domains are likely to play a vital role in development and most are known to be sequence-specific DNA-binding transcription factors. The homeobox domain binds DNA through a helix-turn-helix (HTH) structure. The HTH motif is characterised by two α-helices, which make intimate contacts with the DNA and are joined by a short turn. The first helix helps to stabilise the structure. The second helix binds to the DNA via a number of hydrogen bonds and hydrophobic interactions, which occur between specific side chains and the exposed bases and thymine methyl groups within the major groove of the DNA (Schofield, 1987). Interestingly there is also an alternative splicing site located between amino acid residues 48 to 73 which gives rise to two functionally distinct homeodomain proteins (Theill *et al.*, 1992). This transcription factor is implicated in the specification of the lactotrope, somatotrope and tyrotrope phenotypes in the developing anterior

209

pituitary. It activates the growth hormone and prolactin genes by specifically

binding to the consensus sequence TAAAT (Ingraham et al., 1988).

Clone 6 was identified as the transcription factor E2F5. The EF

transcription factor family are transcriptional activators that bind to E2F sites,

these sites are present in the promoters of many genes whose products are

involved in cell proliferation and therefore play an important role in

regulating the expression of genes that are required for passage through the

cell cycle. It is also thought they may mediate growth factor-initiated signal

transduction, and may participate in differentiation processes (Itoh *et al.*,

1995). Multiple E2F family members have been identified that bind to DNA

as heterodimers, interacting with proteins known as dimerisation partners

(DP).

Clone 7 has been identified as the cyclic-AMP-dependent

transcription factor atf-2 (activating transcription factor 2). This protein factor

belongs to the basic DNA-binding/leucine zipper and zinc finger families of

transcription factors. The zinc finger motif is located between residues 7 to

31 and is of the two cysteine-two histidine type. The motif is referred to as a

zinc finger because of its proposed structure i.e. a loop of twelve amino

acids project from the surface of the molecule which is anchored at its base

by the cysteine and histidine residues which tetrahedrally co-ordinate a zinc

atom (Kadonaga *et al.,* 1987). The zinc finger motif has been shown to

constitute the DNA-binding domain of the protein. The basic DNA-binding

domain is located between amino acid residues 333 and 356, and the

leucine zipper between residues 362 to 390 respectively. Interestingly there

are two tyrosine residues at positions 51 and 53 that are susceptible to

modification (phosphorylation) by MAP kinase. This protein factor is thought to function as a constitutive transcriptional activator binding to the cyclic AMP response element (CRE) a sequence that is present in many cellular and viral promoters (Kageyama *et al.*, 1991).

Comparison of the basic DNA binding and leucine zipper motifs of the clones have revealed some interesting similarities, e.g. comparison of the basic DNA binding domain of the clones encoding the AP-1 transcription factor (clones 1 and 4) with that of the ATF-2 transcription factor clone (clone 7) showed a 10% amino acid sequence similarity.

```
AP-1      RKRMRNRIAASKCRKRKLER
            :                 :              10% Similarity
ATF-2     PDEKRRKFLERNRAAASRCRQKRK
```

However, comparison of the leucine zipper motifs of the respective clones showed an amino acid sequence similarity of 45%.

```
AP-1    LEEKVKTLKAQNSELASTANMLREQVAQL
        :: :   :    :   : :      ::  ::::         45% Similarity
ATF-2   LEKKAEDLSSLNGQLQSEVTLLRNEVAQL
```

The identification of c-jun is very interesting as there is a site in the –609 and –577 region of the MBP promoter which resembles an AP-1 binding site. However further investigations including EMSA with recombinant c-Jun and labelled ds oligonucletide corresponding to this site are required. Cross-linking studies using nuclear extracts from oligodendrocyte cell lines and this oligonucleotide followed by protein micro-sequencing would be also useful. Work on the unidentified clone, possibly

211

representing a novel transcription factor, should be extended by screening a brain cDNA library in order to isolate a full-length copy of the corresponding cDNA. Expression of this protein in heterologous systems combined with established approaches applied in studies of transcription factors would permit the determination of a role, if any, of this protein in transcription of MBPs. Possible investigations which could be based on this thesis are also described in the section entitled Future Work (p 230). Surprisingly, several known transcription factors have been also detected in the one-hybrid system. Their cognate DNA sequences are not present in the studied region therefore they might represent false positives.

It should be noted that one of the limitations of the yeast one-hybrid system is the frequent occurrence of false positive results (e.g. clone 3) which reflect non-specific interactions between the cDNA (encoding the protein) and the target DNA-binding sequence. However, to reduce the frequency of such spurious results, most one-hybrid systems make use of dual reporters, typically *HIS3* and *LacZ*. Since the minimal promoter of the *HIS3* gene ($P_{minHIS3}$) is different from that of the LacZ gene (P_{CYC1}), non-specific interactions fail to activate the expression of both reporter genes. Therefore, the first genetic selection can be achieved using the *HIS3* reporter gene, and the second selection can be performed by measuring *LacZ* expression with a β-galactosidase assay. From published data (Chen and Zack, 2000) it has been shown that up to half the *HIS3* positives from a library screen are *LacZ* negative, suggesting these represent false positive results. False positive results can also be due to condensation caused whilst

the SD medium solidifies resulting in localised dilution of the 3 –

aminotriazole thus enabling colony growth due to the background expression

of the *HIS3* gene.

Chapter 4

DISCUSSION

4.1.Discussion and Conclusions

A study has been undertaken to identify new protein factor(s) that could be involved in the expression of the mouse myelin basic protein gene. To achieve this the promoter region of the mouse myelin basic protein gene was cloned in the vector pBluescript and manually sequenced by the chain termination method of Sanger *et al.* (1977). The *Hind III – Sac I* fragment which corresponds to the region between nucleotides –1319 and +227 relative to the transcriptional start site of the *MBP* gene was used for this purpose (Roach *et al.*, 1985). The data obtained from the sequencing gels was used to carry out a FastA sequence alignment search against the EMBL rodent database. A significant alignment was found with a sequence published by Miura et al., (1989 and updated March 2000) which showed that the two sequences were essentially identical (99.6%). The analysis of the sequence allowed the identification of putative binding sites for many known transcription factors e.g. MyEF2 (Haas *et al.*, 1995), MEBA (Taveggia *et al.*, 1998), Purα (Tretiakova *et al.*, 1998), MRF-1 (Haque *et al.*, 1994) and others. The region – 609 to – 577 was selected for further studies because work published by other groups suggested that it contains cell-type specific transcription activator for oligodendrocytes (Asipu and Blair 1994). To establish if specific protein factor(s) bind to this region, the electrophoretic mobility shift assay was performed using nuclear extracts from several cell lines and rat brain extracts. A protein complex (C2) has been identified in the

214

C$_6$ glioma cell and rat brain nuclear extracts (Figures 17 to 21) which was absent in HeLa cells and 33B oligodendrocytoma cells. The binding of the protein factor(s) to this region of the promoter is highly specific, as the competition binding assay results show. When a 100 - fold molar excess of unlabelled oligonucleotide corresponding to the binding site is added to the reaction mixture, protein complex 2 is competed out in both the C$_6$ glioma cell and rat brain nuclear extracts. Once the specificity of the binding had been established the next step was to identify the protein factor(s), this was achieved using the yeast one-hybrid assay (Clontech Laboratories). The one-hybrid system is an *in vivo* genetic assay used for the isolation of novel genes encoding proteins that bind to a target, *cis*-acting regulatory element (or any other short DNA biding site).

The one-hybrid assay is based on the notion that many eukaryotic transcriptional activators are composed of physically and functionally independent DNA-binding domains and transcription activation domains. This allows the creation of various gene fusions carrying a cDNA that encodes the protein of interest, when expressed in yeast the fusion protein can simultaneously bind to a target sequence and activate transcription. The one-hybrid system is centred on this interaction between a target – independent activation domain and a target – specific DNA binding domain. To create the yeast reporter strain it was first necessary to construct a reporter plasmid by cloning the putative binding site in the multiple cloning site (MCS) of a yeast vector (pHISi-I). The vector pHISi-1 (5.4 kb) is a yeast integrative plasmid, containing the yeast *HIS3* gene downstream of a multiple cloning site (see the map of this vector in Appendix VI). pHISi-1

cannot replicate autonomously in yeast (as it does not have the yeast 2 μm origin of replication or an autonomously replicating sequence (ARS) which allows its propagation in those yeast cells), however, the presence of the *HIS3* gene allows its integration into the mutated *HIS3* gene of the YM4271 yeast host strain. Integration was achieved by digesting the reporter plasmid with the restriction endonuclease *XhoI,* which cuts uniquely within the *HIS3* gene, since DNA ends are the favoured substrate for the recombination enzymes (Guthrie and Fink, 1991). To carry out the integration linear reporter plasmid carrying the putative binding site was introduced into competent yeast sphaeroplasts (a diagrammatic representation of the integration process can be seen in the materials and methods chapter). Integration of the plasmid DNA resulted in a stable duplication of the target gene into the yeast genome. This yeast reporter strain was then used as a host for the screening of a rat brain cDNA/AD fusion library to isolate the protein factor(s).

The cDNA/AD fusion library (Clontech Laboratories) is centred on RT-PCR products (from mRNA isolated from whole rat brains), to which *EcoR1* adapters were ligated, before being cloned via the *EcoR1* restriction endonuclease site located in the MCS of the plasmid pGAD10 which comprises of the target – independent GAL 4 activation domain and the target – specific cDNA binding domain (a diagrammatic representation of the cDNA/AD fusion library is shown in Figure 6c). The cDNA/AD fusion hybrid then interacts with the target element (binding site) resulting in the expression of the HIS3 reporter gene, which allows colony growth on minimal medium lacking histidine. Also incorporated into the growth medium

216

was 3-Aminotriazole (3-AT), which is a competitive inhibitor of the yeast

HIS3 protein (*HIS3*p), and was used to inhibit low levels of *HIS3*p expressed

in a background (or leaky) manner in the reporter strain.

The plasmid DNA encoding the protein factor(s) was sequenced

automatically, and following conversion of the nucleotide sequence to a

protein sequence the results were compared to other protein factor(s) by

FastA and BLAST searches in the Swissprot database. The protein factor(s)

have been identified as follows :

Clones 1 and 4 both contain protein sequence homology to the rat *c-jun*

transcription factor which belongs to the basic DNA-binding domain / leucine

zipper family of transcription factors. Using the BLAST2 sequence alignment

tool a significant sequence overlap between the two clones was also found.

From the expected probability value (E) it is highly unlikely these alignments

had occurred by chance. The reason the AP-1/*c-jun* transcription factor is

binding to the region – 609 to – 577 may be explained by the fact that this

region contains a partial homology to the AP-1 binding site (consensus

sequence TGAGTCAG). This may also explain the quite high background

expression of the *HIS3* gene in the reporter strain since yeast also have a

transcriptional activator GCN4 which binds to the consensus sequence

TGA(C/G)TCAG (Hattori *et al.*, 1988). Such similarities imply that the *c-jun*

product acts in the nucleus as a transcription factor, and that the high

structural identity with GCN4 suggests that the two proteins may recognise

similar DNA sequences (Latchman, 1998).

Clone 2 carries a sequence that at present has not yet been identified and

could possibly be a novel transcription factor.

Clone 3 carries a sequence identified as the PASK protein. The PASK protein is a proline-alanine rich STE20-related kinase.

Clone 5 belongs to the homeobox / POU family of transcription factors and has been identified as the pituitary specific positive transcription factor 1 (Pit-1), growth hormone factor-1 (GHF-1) or general transcription factor (GTF2.i). The homeobox is a highly conserved 180 nucleotide segment of DNA encoding a 60 amino acid homeodomain, which was first identified in several *Drosophila* genes important for controlling early development (McGinnis *et al.*, 1984). The observed homology of the homeodomain with the helix-turn-helix DNA-binding motif found in prokaryotic repressors led to the notion that homeodomain-containing proteins are sequence-specific DNA-binding proteins (Mihara and Kaiser, 1988). POU proteins, like the homeobox proteins, occur in a wide variety of organisms and play an extremely important role in the regulation of gene expression during development. These proteins may also be of particular importance in the development of the central nervous system (Latchman, 1998).

Clone 6 has been identified from BLAST sequence alignment searches as belonging to the EF transcription factor family, which are transcriptional activators binding to E2F sites.

BLAST alignment searches have identified clone 7 as the protein factor known as ATF-2 (the activating transcription factor 2), and like clones 1 and 4, ATF-2 belongs to the basic DNA-binding/leucine zipper family of transcription factors. Both of these protein factors mediate their effects by binding to the basic DNA-binding domain as either a homodimer or as a heterodimer.

In order to confirm the DNA-binding activity of the protein isolated from the clones it would be necessary to carry out an electrophoretic mobility shift assay (EMSA) using the purified protein (transcription factor) and the labelled binding site as probe.

Since the region (between nucleotides – 609 to – 577) lies outside the distal promoter region (which is located between nucleotides – 254 to – 53 as described by Tamura *et al.*, 1989 and Aoyama *et al.*, 1990) of the mouse myelin basic protein gene it may have some other function, for example, acting as an enhancer.

A characteristic feature of eukaryotic gene expression is the existence of sequence elements located some distance away from the transcriptional start site, which also influence the level of gene expression. These elements can be located upstream, downstream or within the transcription unit itself, and function irrespective of orientation relative to the transcription start site (Figure 38). Enhancers act by increasing the activity of the promoter, although they lack promoter activity (Hatzopoulos *et al.*, 1988 and Muller *et al.*, 1988). Some enhancers are active in all tissues increasing the activity of the promoter in all cell types, however, others function as tissue-specific enhancers which activate a particular promoter only in a specific cell type as it has been proposed for the region located between nucleotides –609 to –577 of the myelin basic protein promoter. The tissue-specific activity of such enhancer elements plays a fundamental role in mediating the observed tissue-specific pattern of expression of the corresponding gene. As with the promoter elements, enhancer elements also contain multiple binding sites for transcription factors which interact together (Carey 1998). In many cases

these elements are identical to those immediately upstream of gene promoters, for example the immunoglobulin heavy-chain enhancer contains a copy of the octamer sequence found in the immunoglobulin promoters. Enhancers, therefore, are made up of sequence elements that are present in similarly regulated promoters and may also be found within the enhancers associated with other control elements. Mutation of conserved sequences within the tissue-specific enhancers abolish the tissue-specific pattern of expression of those genes.

Models involving the binding of protein factors to the enhancer followed by direct interaction with proteins of the transcriptional apparatus are more difficult to resolve with the action at some distance, characteristic of most enhancers. Models hypothesising this suggest that the enhancer serves as a site of entry for a regulatory factor, this factor would then make contact with the promoter – bound transcriptional apparatus by one of three ways:

(a) sliding along the DNA

(b) via a continuous scaffold of other proteins, or

(c) by 'looping' out of the intervening DNA.

Of these possibilities, both the continuous scaffold and sliding models are difficult to square with the large distances over which enhancers act. It has also been shown that an enhancer can act on promoters located on two different DNA molecules linked only by a protein bridge. This would totally disrupt the scaffold or sliding molecules model (Muller *et al.* 1989). A more explicable model is one in which the proteins bound at the promoter and enhancer have an affinity for one another and make contact by looping out

220

of the intervening DNA, and such a model can also explain the critical importance that DNA structure has on enhancer action, for example, using the SV40 enhancer it has been shown that removing multiples of 10 bases (one helical turn) has no effect on enhancer activity. However, deletion of five DNA bases (corresponding to half a helical turn) does severely disrupt enhancer activity (Takahashi *et al.* 1989).

Enhancers have been considered to act in a positive manner, however, some enhancers produce specific patterns of gene expression by relieving a negative effect. In these cases it appears the enhancer binds a protein factor, which inhibits the promoter, following induction, or in the case of a specific tissue, this negative factor is inactivated and falls off the enhancer allowing a positive-acting factor to bind and activate gene expression. This positive acting factor may be present in all cell types, but cannot normally bind to the DNA because of the presence of the negative repressor factor (Goodbourn *et al.*, 1986). The most extreme example of an enhancer-like sequence acting in a totally negative way is a silencer. Silencers can act in either orientation to inhibit the activity of distant promoters without binding a regulatory protein, hence appearing to have a negative effect. To sum up, whether acting positively or negatively, enhancers play a key role in regulating eukaryotic gene expression, and often act together with related regulatory elements located adjacent to the promoter itself.

a) Distance

b) Orientation
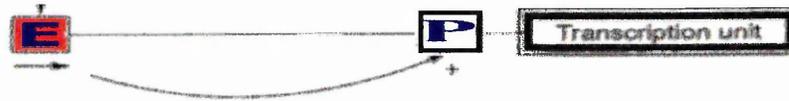
c) Position
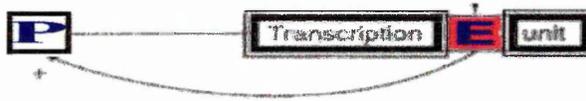
Figure 38. Characteristics of an enhancer element activating a promoter at:

a) a distance from the promoter

b) in either orientation relative to the promoter

c) or when positioned upstream, downstream or within a transcription

element.

KEY   E = Enhancer.   P = Promoter

(Adapted from Latchman D.S. 1990)

222

Awatramani *et al.* (1997) have investigated the patterns of postnatal

brain expression and DNA binding of Gtx (a homeodomain transcription

factor). The accumulation of Gtx mRNA parallels that of the RNAs encoding

the major structural proteins of myelin, for example, the myelin basic protein

and the proteolipid protein during postnatal brain development. Gtx mRNA

decreases in parallel with MBP and PLP in the brains of myelin deficient

rats. Gtx mRNA is expressed in differentiated, post-mitotic oligodendrocytes

but not in oligodendrocyte precursors, astrocytes or other glia. Using a

maltose-binding protein-Gtx fusion protein produced in *E.coli* Awatramani *et*

al. (1997) have established that Gtx is a sequence-specific DNA-binding

protein, which binds DNA sequences containing a core AT-rich

homeodomain. Immunoprecipitation of labelled DNA fragments encoding the

myelin basic protein or proteolipid protein promoter regions with the fusion

protein have identified several Gtx-binding fragments, which have been

confirmed by electrophoretic mobility shift assay. In this way four Gtx binding

sites have been identified within the first 750 bp of the myelin basic protein

promoter, three sites are located between nucleotides – 598 and – 656, and

the fourth site is located between nucleotides – 304 and – 331.The site

between nucleotides – 598 and – 621 binds Gtx with the highest affinity of

the four MBP sites. Four Gtx binding sites in the first – 1.3 kb of the

proteolipid protein promoter, the site located between nucleotides – 644 and

– 671 binds Gtx with the highest affinity. The sites located between – 409

and – 434, and – 1128 and –1214 bind with a lesser affinity and the site

located between – 259 and –286 binds with the least affinity. The locations

and DNA sequences of the Gtx binding sites located in the MBP and PLP promoters are shown in Table 4. Further analysis of the first kilobase of the PLP promoter sequence has since revealed the presence of an additional six putative Gtx binding sites. These data suggest that Gtx is important for the function of differentiated oligodendrocytes and may be involved in the regulation of myelin-specific gene expression.

Although Gtx is a regulator of oligodendrocyte development and myelin-specific gene expression it is not expressed in Schwann cells, which are the myelinating cells of the peripheral nervous system (PNS), whereas, both the major CNS proteins, MBP and PLP, are also expressed in myelinating Schwann cells. Comparison of the known mechanisms of development and regulation of myelin-specific gene expression in both oligodendrocytes and Schwann cells demonstrate several differences. Firstly the transcription factors necessary for normal Schwann cell maturation and PNS myelination, SCIP and Krox-20, are not required for normal oligodendrocyte maturation or myelination, and the developmental expression profile of these proteins is different in the two cell types e.g. SCIP is expressed in promyelinating Schwann cells committed to myelination, but it is also found in oligodendrocyte precursors and undifferentiated myelinating cells. Krox-20 is not expressed in oligodendrocytes. Secondly, different regions within the myelin basic protein promoter region have been shown to be required for oligodendrocyte and Schwann cell gene expression, e.g. myelin basic protein promoter sequences up to 3.2 kb are sufficient to drive developmentally regulated, oligodendrocyte-specific transgene expression but do not support transgene expression in Schwann

cells. (Gow *et al.*, 1992). Thirdly, comparison of the oligodendrocyte and

Schwann cell nuclear proteins that bind to the proximal MBP promoter

(Wrabetz *et al.*, 1993) demonstrates a number of significant differences,

suggesting there is a unique set of transcriptional activators present in the

two cell types. The data suggests that the molecular mechanism underlying

the regulation of Schwann cell and oligodendrocyte development are

different, and as a result the presence of Gtx in oligodendrocytes but not in

Schwann cells is further proof of this suggestion.

Type	Location of site	Sequence
MBP	-327 to -320	CATATTCT
MBP	-606 to -613	GTTAAATA
MBP	-630 to -623	AATAAATG
MBP	-650 to -643	ACTAAAAA
PLP	-262 to -269	TGTAATTG
PLP	-284 to -277	CTTAATTT
PLP	-428 to -421	GGTAATAT
PLP	-662 to -669	ATTAATGA
PLP	-1194 to -1201	AATAATTC
PLP*	-10 to -17	TGTAATCT
PLP*	-245 to -252	TATAAATG
PLP*	-296 to -303	CTTAAATC
PLP*	-331 to -324	TTTAAATG
PLP*	-559 to -566	AATAATTT
PLP*	-1020 to -1013	GATAATCC

* A putative Gtx binding site identified by homology to a known binding site.

Table 4. Gtx binding sites located within the MBP and PLP Promoter
regions.

226

In many respects the myelin basic protein gene is unique, it is among a relatively small group of genes > 100 kb in length and represent an example of both a gene within a gene as well as two overlapping genes, and in that respect it bears some resemblance to the neurofibromatosis gene in that both genes are large transcription units encompassing smaller genes. However, they are different in that the smaller myelin basic protein gene portions are not simply included within an intron of the larger gene but share alternatively spliced exons in common.

Another unusual feature of the myelin basic protein gene is its regulation. The MBP transcription start site (S3), shown in Figure 2, is under very tight developmental control and is only expressed in myelin-forming cells. Most of this regulation resides within the exon 5a region upstream of transcription start site S3. The *golli-mbp* transcription start site S1 is under less specific regulation, therefore, there appears to be a gradient of cell and tissue specificity amongst the three promoters of the gene, with the most downstream promoter being the most specific to myelin forming cells and the most upstream promoter exhibiting the least cell and tissue specificity. This combination of structure and expression significantly changes the notion of the myelin basic protein gene, which has generally been thought to be specific for myelin-forming cells. The genomic region encompassing exon 5A has been used by many workers to target transgenes specifically to oligodendrocytes. It is now suggested that this "myelin" gene is expressed in many neuronal cell populations, some of which are important in establishing the structure of the cerebral cortex. In addition, it has been suggested that increased expression of this gene within the immune system may be

responsible for the relapsing disease phase of experimental autoimmune encephalomyelitis, an animal model of multiple sclerosis (MacKenzie-Graham et al., 1997). Therefore, it has become clear that the myelin basic protein gene is important to neurons and immune cells as well as oligodendrocytes and must therefore perform a function that is not solely associated with myelination.

The work of Landry et al., (1998) has also contributed to the evolving notion that the myelin basic protein gene is one with broader functions in neurons and in myelin-forming cells in the nervous system than was first thought. They have identified a region upstream of the first transcription start site (S1), shown in Figure 2, that can target transgenes to a very limited number of neurons, including the very earliest-forming (pioneer) neurons in the cortex, the Cajal-Retzius cells and preplate neurons.

Since the early nineties there has been a shift away from the traditional notion that the MBP and PLP genes were expressed solely in myelin-forming cells, and their products were thought to be structural components of myelin. It has now been shown that the products of these genes have been identified in nonmyelinating cell types including the immune system and nervous system (Givogri et al., 2000). In the nervous system golli is produced in oligodendrocytes and specific subsets of neurons, however there is no evidence of golli expression in astrocytes or microglia (Landry et al., 1996). Neuronal expression of golli is prominent in the following regions: the spinal cord (interneurons of the anterior and posterior grey horns), the developing cortex (embryonic cortical neurons, presubplate and subplate neurons, Cajal-Retzius cells), the hippocampus

(pyramidal cells, granule cells of dentate gyrus), the cerebellum (Purkinje and granular cells), the olfactory system and the dorsal root ganglia (Landry *et al.*, 1996). Golli products of the *MBP* gene are expressed in the immune system (Campagnoni *et al.*, 1993). In the mouse thymus, golli appears to be expressed in the thymic epithelial cells and macrophages and to a lesser extent in the thymocytes (Campagnoni *et al.*, 1993). Expression of the MBP gene in the thymus might have an important role in influencing tolerance development (by influencing the positive and negative selection of the T-cell repertoire) which could lead to the opening of new research to understanding autoimmune diseases like experimental autoimmune encephalomyelitis (EAE) (Voskhul, 1998).

4.2. Future Work

Future work would involve further studies on the role of the putative
transcription factors, which were identified in this project, in the expression of
the myelin basic protein gene in oligodendrocytes. To achieve this the
activity of the promoter in HeLa cells (control) and oligodendrocytes in the
presence and absence of the cDNA clone expressing the studied
transcription factor would be compared. For this purpose it would be
necessary to obtain the mouse myelin basic protein gene promoter in a
vector containing a reporter gene e.g. chloramphenicol acetyl transferase
(pBLCAT) or luciferase (pGL2-Basic vector), both vectors lack eukaryotic
promoter and enhancer sequences (the reporter plasmid). A full length copy
of the cDNA for a putative transcription factor in a plasmid vector containing
a ubiquitously active promoter e.g. CMV promoter (pcDNA 3.1) would be
also required. Experiments involving the co-transfection of the reporter
plasmid and the plasmid containing the transcription factor cDNA into HeLa
cells (control) and glial (oligodendrocyte) cells would be performed. If the
CAT assay were used to assess activity of the promoter, it would be
performed as follows: cell lysates would be incubated with ^{14}C-labelled
chloramphenicol and acetyl-CoA. In this reaction the acetyl group from the
acetyl-CoA is transferred by the CAT enzyme to the 2- or 3- position on the
chloramphenicol. The products of this reaction are detected by thin layer
chromatography (TLC) followed by autoradiography. The TLC separates the
acetylated chloramphenicol from the unreacted substrates, as a result the 2-
and 3- acetylated chloramphenicol migrate at slightly different rates
producing two spots if the promoter were active. The presence and intensity

of the spots reflects the activity of the promoter as the CAT enzyme is present only when active promoter is upstream of the reporter gene in the reporter plasmid.

Studies of how mutations in the region –609 to –577 affect the expression of the reporter gene in the presence of putative transcription factors could also be undertaken. To achieve this it would initially be necessary to transfect a plasmid with the wild type binding site (enhancer) linked to a promoter and a reporter gene (e.g. pBLCAT2 vector which has the *tk* HSV promoter upstream of the CAT gene or pGL2-promoter vector which has SV40 promoter upstream of the luciferase gene) into HeLa and glial (oligodendrocyte) cell lines to investigate further the biological function and the cell type-specificity of the binding site (enhancer). Similar experiments would be performed using co-transfection experiments with cDNA for the putative transcription factor. An increase in the expression of the reporter gene in the presence of a cDNA encoding a putative transcription factor is expected in an appropriate cell line.

Subsequently constructs with point mutations in the –609 to –577 nt region linked to a reporter gene would be prepared. These constructs would be transfected with and without the transcription factor cDNA into oligodendrocytes (and HeLa cells as control). In this case no increase in reporter gene transcription as determined by the CAT enzyme activity would be expected.

REFERENCES.

Ainger K., Avossa D., Morgan F., Hill S. J., Barry C., Barbarese E. and
Carson J. H. (1993)
Transport and localisation of exogenous myelin basic protein mRNA
microinjected into oligodendrocytes.
J. Cell Biol. **123**, 431 – 441.

Ainger K., Avossa D., Diana A. S., Barry C. and Barbarese E. (1997)
Transport and localisation elements in myelin basic protein mRNA.
J. Cell Biol. **138**, (5) 1077 – 1087.

Albers R. W., Siegel G. J. and Stahl W. L. (1994)
In Basic neurochemistry (Siegel G. J. Agranoff B. W., Albers R. W.
and Molinoff P. B., Eds) 5thed., 49 – 73. Raven Press.

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990)
Basic local alignment search tool.
J. Mol. Biol. **215**, 403 – 410.

Amur S.G., Shanker G., Cochran J.M., Ved H.S., and Pieringer R.A.
(1986)
Correlation between inhibition of myelin basic protein (arginine)
methyltransferase by sinefungin and lack of compact myelin
formation in cultures of cerebral cells from embryonic mice.
J. Neurosci. Res. **16**, 367 - 376.

Aoyama A., Tamura T., and Mikoshiba K.   (1990)
Regulation of brain-specific transcription of the mouse myelin basic
protein gene: function of the NF1-binding site in the distal promoter.
Biochem. Biophys. Res. Comm. **167**, 648 – 653.

Aruga J., Okano H., and Mikoshiba K.   (1991)
Identification of the new isoforms of mouse myelin basic protein: the
existence of exon 5a.
J. Neurochem. **56**, 1222 – 1226.

Asipu A., and Blair G.E,   (1994)
Regulation of myelin basic protein-encoding gene transcription in rat
oligodendrocytes.
Gene **150** 227 – 234.

Asipu A., and Blair G.E,   (1997)
    Regulation of transcription and translation in the Central Nervous
    System.
    Molecular Biology of Multiple Sclerosis (Russell W.C. editor)
    John Wiley and Sons pp 121 – 136.

Assa-Munt N., Mortishire-Smith R.J., Aurora R., Herr W. and Wright P.E.
    (1993)
    The solution structure of the Oct-1 POU-specific domain reveals a
    striking similarity to the bacteriophage lambda repressor DNA-binding
    domain.
    Cell 73, 193 – 205.

Awatramani R., Scherer S., Grinspan J., Collarini E., Skoff R., O'Hagan D.,
    Garbern J., and Kamholz J.   (1997)
    Evidence that the homeodomain protein Gtx is involved in the
    regulation of oligodendrocyte myelination.
    J. Neurosci. **17**, (17) 6657 – 6668.

Bacher G, Lütcke H., Jungnickel B., Rapoport T. A. and Dobberstein B.
    (1996)
    Regulation by the ribosome of the GTPase of the signal-recognition
    particle.
    Nature (London) **381**, 248 – 251.

Bansal G., Martenson R.E., Leveille P., and Campagnoni A.T.   (1987)
    Characterisation of a novel monoclonal anti-myelin basic protein
    antibody: use in immunoblotting and immunohistochemical studies.
    J. Neuroimmunol. **15**, 279 - 294.

Barbarese E., Braun P.E. and Carson J.H.   (1977)
    Identification of prelarge and presmall basic proteins in mouse myelin
    and their structural relationship to large and small basic proteins.
    Proc. Natl. Acad. Sci. USA. **74**, 3360 - 3364.

Barbarese E., and Pfeiffer S.E.   (1981)
    Developmental regulation of myelin basic protein in dispersed
    cultures.
    Proc. Natl. Acad. Sci. USA. **78**, 1953 - 1957.

Barbarese E., Nielson M.L., and Carson J.H.   (1983)
    The effect of the *shiverer* mutation on myelin basic protein expression
    in homozygous and heterozygous mouse brain.
    J. Neurochem. **40**, 1680 - 1686.

Barbarese E., Koppel D. E., Deutscher M. P., Smith C. L., Ainger K., Morgan F., and Carson J. H. (1995)
Protein translation components are colocalised in granules in oligodendrocytes.
J. Cell Sci. **108**, 2781 – 2790.

Benjamins J.A., and Morell P. (1987)
Proteins of myelin and their metabolism.
Neurochem. Res. **3**, 137 - 174.

Berg J. M. (1986)
Potential metal binding domains in nucleic acid binding proteins.
Science **232**, 485 – 492.

Berlet H. H., Bischoff H. and Weinhardt F. (1994)
Divalent metals of myelin and their differential binding by myelin basic protein of bovine central nervous system.
Neurosci. Lett. **179**, (1-2) 75 – 78.

Bhat S., Barbarese E., and Pfeiffer S.E. (1981)
Requirement for non-oligodendrocyte cell signals for optimal myelinogenic gene expression in long term cultures of purified rat oligodendrocytes.
Proc. Natl. Acad. Sci. USA. **78**, 1283 - 1287.

Bhat S., and Pfeiffer S.E. (1986)
Stimulation of oligodendrocytes by extracts from astrocyte-enriched cultures.
J. Neurosci. Res. **15**, 19 - 27.

Birnboim C., and Doly J. (1979)
Rapid alkaline extraction procedure for screening recombinant plasmid DNA.
Nucl. Acids Res. **7**, 1513 – 1523.

Blank V. and Andrews N.C. (1997)
The Maf transcription factors: regulators of differentiation.
Trends in Biochemical Sciences **22**, 437 – 441.

Bourre J.M., Jacque C., Delassale A., Nguyen-Largros J., Dumont O., Lachapelle F., Raoul M., Alvarez C., and Baumann N. (1980)
Density profile and basic protein measurements in the myelin range of particulate material from normal developing mouse brain and from neurological mutants.
J. Neurochem. **35**, 458 - 464.

Boyle W.J., Smeal T., Defize L.H.K., Angel P., Woodgett J.R., Karin M. and
      Hunter T. (1991)
      Activation of protein kinase C decreases phosphorylation of c-jun at
      sites that negatively regulate its DNA binding activity.
      Cell **64**, 573 – 584.

Breitbart R.E., Andreadis A., and Nadal-Ginard B.   (1987)
      Alternative splicing: a ubiquitous mechanism for the generation of
      multiple protein isoforms from single genes.
      Ann. Rev. Biochem. **56**, 467 - 495.

Brophy P. J., Bocaccio G. L. and Colman D. R. (1993)
      The distribution of myelin basic protein mRNA with myelinating
      oligodendrocyte.
      Trends Neurosci. **16**, 515 – 521.

Campagnoni A.T., Carey G.D., and Yu Y.-T.   (1980)
      In vitro synthesis of the myelin basic proteins: subcellular site of
      Synthesis.
      J. Neurochem. **34**, 677 - 686.

Campagnoni A.T., Campagnoni C.W., Bourre J.-M., Jacque C., and
      Baumann N. (1984)
      Cell-free synthesis of myelin basic proteins in normal and
      dysmyelinating mutant mice.
      J. Neurochem. **42**, 733 - 739.

Campagnoni A.T.,   (1985)
      Molecular Biology of myelination: gene expression of myelin basic
      protein.
      Gene Expression in Brain, (Zomzely - Neurath C. and Walker W.A.,
      eds),   John Wiley, pp 205 - 233.

Campagnoni A.T., Hunkeler M.J., and Moskaitis J.E.   (1987)
      Translation regulation of myelin basic protein synthesis.
      J. Neurosci. Res. **17**, 102 - 110.

Campagnoni A.T.,   (1988)
      Molecular biology of myelin proteins from the central nervous system.
      J. Neurochem. **51**, (1) 1 – 14.

Campagnoni A.T., and Macklin W.B.   (1988)
      Cellular and molecular aspects of myelin protein gene expression.
      Mol. Neurobiol. **2**, 41 – 89.

Campagnoni A.T., Pribyl T.M., Campagnoni C.W., Kampf K., Amar-Umarjee S., Landry C. F., Handley V.W., Newman S.L., Garbay B. and Kitamura K (1993)
Structure and developmental regulation of *Golli-mbp*, a 105-kilobase gene that encompasses the myelin basic protein gene and is expressed in cells in the oligodendrocyte lineage in the brain.
J. Biol. Chem. **268**, 4930 – 4938.

Carey M.   (1998)
Enhanceosome and transcriptional synergy.
Cell **92**, 5 – 8.

Carnegie P.R.   (1971)
Properties, Structure, and possible neuroreceptor role of the encephalitogenic protein of the human brain.
Nature. **229**, 25 - 28.

Carnow T.B., Carson J.H., Brostoff S.W., and Hogan E.L.   (1984)
Myelin basic protein gene expression in *quaking, jimpy* and *myelin synthesis-deficient* mice.
Dev. Biol. **106**, 38 - 44.

Carson J.H., Nielson M.L., and Barbarese E.   (1983)
Developmental regulation of myelin basic protein expression in mouse brain.
Dev. Biol. **96**, 485 - 492.

Cavatorta P., Giovanelli S., Bobba A., Riccio P., Scabo A. G. and Quagliariello E. (1994)
Myelin basic protein interaction with zinc and phosphate:
Fluorescence studies on the water-soluble form of the protein.
Biophys. J. **66**, 1174 – 1179.

Chen S. and Zack D. J. (2000)
Cloning and characterisation of retinal transcription factors, using target site-based methodology.
Methods in Enzymology **316**, 590 – 610.

Chernoff G.   (1981)
Shiverer. an autosomal recessive mutant mouse with myelin deficiency.
J. Hered. **72**, 128.

Clark R., Stewart M., Miskimins K. and Miskimins R. (1998)
Involvement of MAP kinase in the cyclic AMP induction of myelin basic protein gene expression.
Int. J. Devl. Neuroscience **16**, (5), 323 – 331.

Clewell D.B., and Helinski D.R,   (1969)
Supercoiled circular DNA-protein complex in *Escherichia coli*:
Purification and induced conversion to an open circular DNA form.
Proc. Natl. Acad Sci U.S.A. **62**, 1159 – 1166.

Clewell D.B. (1972)
Nature of Col E_1 plasmid replication in *Escherichia coli* in the
presence of   chloramphenicol.
J. Bacteriol. **110**, 667.

Clontech Laboratories (1995)
Matchmaker One-hybrid system protocol 1 – 29.

Cohen S.N., Chang A.C.Y. and Hsu L.   (1972)
Nonchromosomal antibiotic resistance in bacteria: Genetic
transformation of *Escherichia coli* by R - factor DNA.
Proc. Natl. Acad Sci U.S.A. **69**, 2110

Colman D.R., Kreibich G., Frey A.B., and Sabatini D.D.   (1982)
Synthesis and incorporation of myelin polypeptides into CNS myelin.
J. Cell Biol. **95**, 598 - 608.

Dawson S.J., Morris P.J. and Latchman D.S. (1996)
A single amino acid change converts a repressor into an activator.
Journal of Biological Chemistry **271**, 11631 – 11633.

Deibler G.E., Krutzsch C.H., and Kies M.W.   (1986)
A new form of myelin basic protein found in human brain.
J. Neurochem. **47**, 1219 - 1225.

Deibler G. E., Stoner A. I. And Kies M. W. (1990)
Role in phosphorylation in conformational adaptability of bovine
myelin basic protein.
Proteins **7**, 32 – 40.

Delassale A., Zalc B., Lachapelle R., Raoul M., Collier P., and Jacque C.
(1981)
Regional distribution of myelin basic protein in the CNS *quaking,
jimpy* and normal mice during development and aging.
J. Neurosci. Res. **6**,303 – 313.

Desmarais D., Filon M., Lapointe L., and Royal A. (1992)
Cell-specific transcription of the peripherin gene in neuronal cell lines
involves cis-acting element surrounding the TATA box.
EMBO J. **11**, 2971 –2980.

DesJardins K.C., and Morell P.   (1983)
   Phosphate group modifying myelin basic proteins are metabolically
   labile; methyl groups are stable.
   J. Cell Biol. **97**,438 - 446.

Devine-Beach K.A., Haas S., and Khalili K.   (1992)
   Analysis of the proximal transcriptional element of the myelin basic
   protein gene.
   Nucl. Acids Res. **20** 545 – 550.

Diehl H.-J., Schaich M., Budzinski R.-M., and Stoffel W.   (1986)
   Individual exons encode the integral membrane domains of human
   myelin proteolipid protein.
   Proc. Natl. Acad. Sci. USA. **83**, 9807 - 9811.

Doolittle D.P., and Schweikart K.M.   (1977)
   Myelin deficient, a new neurological mutant in the mouse.
   J. Hered. **68**, 331 - 332.

Doolittle D.P., Baumann N., and Chernoff G.   (1981)
   Allelism of two myelin deficient mutations in the mouse.
   J. Hered. **72**, 285.

Dubois-Delacq M., Behar T., Hudson L., and Lazzarini R.A. (1986)
   Emergence of three myelin proteins in oligodendrocytes cultured
   without neurons.
   J. Cell Biol. **102**, 384 - 392.

Duncan I. D., Lunn K. F., Holmgren B., Urba-Holmgren R. and Brignolo-
      Holmes L. (1992)
   The taiep rat: A myelin mutant with an associated oligodendrocyte
   microtubular defect.
   J. Neurocytol. **21**, 870 – 884.

Ebersole T. A., Rho O. and Artzt K. (1992)
   The proximal end of chromosome 17: New molecular markers identify
   a deletion associated with quaking[viable].
   Genetics **131**, 183 – 190.

Ebersole T. A., Chen Q., Justice M. J. and Artzt K. (1996)
   The quaking gene product necessary in embryogenesis and
   myelination combines features of RNA binding and signal
   transduction proteins.
Nat. Genet. **12**, 260 – 265.

Eichberg J. and Iyer S. (1996)
   Phosphorylation of myelin proteins: Recent advances.
   Neurochem. Res. **21**, (4) 527 – 535.

Eylar E.H., (1970)
Amino acid sequence of the basic protein of the myelin membrane.
Proc. Natl. Acad. Sci. USA. **67**, 1425 - 1431.

Farsetti A., Desvergne B., Hallenbeck P., Robbins J., and Nikodem V.M. (1992)
Characterisation of myelin basic protein thyroid hormone response element and its function in the context of native and heterologous promoter.
J. Biol. Chem. **267**, 15784 – 15788.

deFerra F., Engh H., Kamholz J., Puckett C., Molineaux S., and Lazzarini R.A. (1985)
Alternative Splicing accounts for the four forms of Myelin Basic Protein.
Cell **43**, 721 – 727.

Fors L., Hood L., and Saavedra R.A. (1993)
Sequence similarities of myelin basic protein promoters from mouse and shark: Implications for the control of gene expression in myelinating cells.
J. Neurochem **60**, 513 – 521.

Frederickson C. J. (1989)
Neurobiology of zinc and zinc-containing neurons.
Int. Rev. Neurobiol. **31**, 145 – 238.

Gandy G., Jacobson W., and Sidman R. (1973)
Inhibition of a transmethylation reaction in the central nervous system - an experimental model for subacute combined degeneration of the cord.
J. Physiol. (Lond). **233**,1P - 3P.

Garbay B., Heape A. M., Sargueil F. and Cassagne C. (2000)
Myelin synthesis in the peripheral nervous system.
Prog. Neurobiol. **61**, 267 – 304.

Gehring W. J. (1992)
The homeobox in perspective.
Trends Biochem. Sci. **17**, 277 – 280.

Ghosh S. K., Paik W. K. and Kim S. (1988)
Purification and molecular identification of two protein methylase I from calf brain.
J. Biol. Chem. **263**, 19024 – 19033.

Ginalski-Winkelman H., Guillermina A., and Matthieu J.-M.   (1983)
In vitro myelin basic protein synthesis in the PNS and CNS of myelin deficient (*mld*) mutant mice.
Brain Res. **277**, 386 - 388.

Givogri M. I., Bongarzone E. R. and Campagnoni A. T. (2000)
New insights on the biology of myelin basic protein gene: the neural-immune connection.
J. Neurosci. Res. **59**, 153 – 159.

Goodbourn S., Burnstein H., and Maniatis T. (1986)
The human β-interferon gene is under negative control.
Cell **45**, 601 - 610.

Gow A., Friedrich Jr. V.L., Lazzarini R.A.   (1992)
Myelin basic protein gene contains separate enhancers for oligodendrocyte and Schwann cell expression.
J. Cell Biol. **119** 605 – 616.

Guthrie C. and Fink G.R.   (1991)
Guide to yeast genetics and molecular biology.
In Methods in Enzymology (Academic Press, San Diego) **194**, 1-932.

Haas S., Steplewski A., Siracusa L.D., Amini S., and Khalili K.   (1995)
Identification of a sequence specific single-stranded DNA binding protein that suppresses transcription of the mouse myelin basic protein gene.
J. Biol. Chem. **270**, (21) 12503 – 12510.

Hardy R. J. (1998)
Molecular defects in the dysmyelinating mutant quaking.
J. Neurosci. Res. **51**, 417 – 422.

Hardy R. J., Loushin C. L., Friedrich Jr V. L., Chen Q., Ebersole T. A., Lazzarini R. A. and Artzt K. (1996)
Neural cell type-specific expression of QKI proteins is altered in quakingviable mutant mice.
J. Neurosci. **16**, 7941 – 7949.

Hattori K., Angel P., Lebeau M.M. and Karin M. (1988)
Structure and chromosomal localisation of the functional intronless human jun proto-oncogene.
Proc. Natl. Acad. Sci. USA **85**, 9148 – 9152.

Hatzopoulos A.K., Schlokat U. and Gruss P.   (1988)
Enhancers and other *cis*-acting sequences. In transcription and splicing, B. D. Hames and D. M. Glover (*eds*) 43 – 96 IRL press, Oxford

240

Hickson A. K., Payne D. M., Martino P. A., Rossomando A. J., Shabapowitz J., Weber M. J., Hunt D. J. and Sturgall T. W. (1990)
Identificationby mass spectrometry of threonine 97 in bovine myelin basic protein as a specific phosphorylation site for mitogen-activated protein kinase.
J. Biol. Chem. **265**, 19728 – 19735.

Hoek K. S., Kidd G. J., Carson J. H. and Smith R. (1998)
hnRNP A2 selectively binds to cytoplasmic transport sequence of myelin basic protein mRNA.
Biochemistry **37**, 7021 – 7029.

Holmes D.S. and Quigley M.   (1981)
A rapid boiling method for the preparation of bacterial plasmids.
Anal. Biochem. **114** 193.

Holz A., Schaeren_Wiemers N., Schafer C., Pott U., Colello R. J. and Schwab M. E. (1996)
Molecular and developmental characterisation of novel cDNAs of the myelin-associated/oligodendrocytic basic protein.
J. Neurosci. **16**, 467 – 477.

Ingraham H.A., Chen R., Mangalam H.J., Elsholtz H.P., Flynn S.E., Lin C.R., Simmons D.M., Swanson L. and Rosenfeld M.G. (1988)
A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype.
Cell **55**, 519 – 529.

Inoue Y., Nakamura R., Mikoshiba K., and Tsukada Y.   (1981)
Fine structure of the central myelin sheath in the myelin deficient mutant *shiverer* mouse, with special reference to the pattern of myelin formation by oligodendroglia.
Brain Res. **219**, 85 - 94.

Jacque C., Delassale A., Raoul M., and Baumann N. (1983)
Myelin basic protein deposition in the optic and sciatic nerves of dysmyelinating mutants *quaking, jimpy, trembler*, MLD, *shiverer* during development.
J. Neurochem. **41**, 1335 - 1340.

Jiricny J., Hughes M., Corman N., and Rudkin B.B.   (1988)
A human 200 kDa protein binds selectively to DNA fragments containing G.T mismatches.
Proc. Natl. Acad Sci U.S.A. **85** 8860-8864.

Johnson W.A. and Hirsh J.(1990)
Binding of Drosophila POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons.
Nature **343**, 467 – 470.

Kadonaga J.T., Carner K.R., Masiarz F.R. and Tjian R. (1987)
Isolation of cDNA encoding the transcription factor Sp1 and functional analysis of the DNA binding domain.
Cell **51**, 1079 – 1090.

Kageyama R., Sasai Y. and Nakanishi S. (1991)
Molecular characterisation of transcription factors that bind to the cAMP responsive region of the substance P precursor gene. cDNA cloning of a novel C/EBP-related factor.
J. Biol. Chem. **266**, 15525 –15531.

Kamholz J., deFerra F., Puckett C., and Lazzarini R. (1986)
Identification of three forms of human myelin basic protein by cDNA cloning.
Proc. Natl. Acad. Sci. USA **83**, 4962 - 4966.

Kenan D. J., Query C. C. and Keene J., D. (1991)
RNA recognition: towards identifying determinants of specificity.
Trends Biochem. Sci. **16**, (6), 214 – 220.

Kerlero de Rosbo N., Carnegie P.R., Bernard C.C.A., and Linthicum D.S. (1984)
Detection of various forms of myelin basic protein in vertebrates by electroimmunoblotting.
Neurochem. Res. **9**, 1359 - 1369.

Kerner A.-L., and Carson J.H. (1984)
Effect of the *jimpy* mutation on expression of myelin proteins in heterozygous and hemizygous mouse brain.
J. Neurochem. **43**, 1706 - 1715.

Kim S., Tuck M., Campagnoni A.T., and Paik W. (1984)
Studies of myelin basic protein-specific protein methylase I in various dysmyelinating mutant mice.
Biochem. Biophys. Res. Comm. **123**, 468 - 474.

Kim S., Tuck M., Ho L., Campagnoni A.T., Barbarese E., Knobler R.L., Lublin F.D., Chanderkar L.P., and Paik W. (1986)
Myelin Basic Protein-specific protein methylase I activity in *shiverer* mutant mouse brain.
J. Neurosci. Res. **16**, 357 - 365.

Kim S., Lim I. K., Park G-H. and Paik W. K. (1997)
Biological methylation of myelin basic protein: Enzymology and biological significance.
Int. J. Biochem. Cell Biol. **29**, (5) 743 – 751.

Kimbro K.S., Rosenberg P.A., and Saavedra R.A.   (1994)
Box I and II motif from myelin basic protein gene promoter binds to nuclear proteins from rodent brain.
J. Mol. Neurosci. **5**, (1) 27 – 37.

Kimura M., Inoko H., Katsuki M., Ando A., Sato T., Hirose T., Takashima H., Inayama S., Okano H., Takamatsu K., Mikoshiba K., Tsukada Y., and Watanabe I. (1985)
Molecular genetic analysis of myelin-deficient mice: *shiverer* mutant mice show deletion in gene(s) coding for myelin basic protein.
J. Neurochem. **44**, 692 - 696.

Kimura M., Sato M., Akatsuka A., Nazawa-Kimura S., and Takahashi R., (1989)
Restoration of myelin formation by a single type of myelin basic protein in transgenic *shiverer* mice.
Proc. Natl. Acad. Sci. U.S.A. **86**, 5661 - 5665.

Kitamura K., Newman S. L., Campagnoni C. W., Verdi J. M., Mohandas T., Handley V. W. and Campagnoni A. T. (1990)
Expression of a novel transcript of the myelin basic protein gene.
J. Neurochem. **54**, (6) 2032 – 2041.

Klemm J.D., Rould M.A., Aurora R., Herr W. and Pabo C.O. (1994)
Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules.
Cell **77**, 21 – 32.

Kondo T., Furuta T., Mitsunaga K., Ebersole T. A., Schichiri M., Wu J., Artzt K., Yamamura K. and Abe K. (1999)
Genomic organisation and expression analysis of the mouse QKI locus.
Mamm. Genome **10**, 662 – 669.

Kwon S., Barbarese E. and Carson J.H.   (1999)
The cis-acting RNA trafficking signal from myelin basic protein mRNA and its cognate trans-acting ligand hnRNP A2 enhance cap-dependent translation.
J. of Cell Biol. **147**, (2) 247 – 256.

Landry C.F., Ellison J.A., Pribyl T.M., Campagnoni C.W., Kampf K, and Campagnoni A.T.   (1996)
Myelin protein gene expression in neurons: developmental and regional changes in protein targeting within neuronal nuclei, cell bodies and processes.
J. Neurosci. Res. **50**, 265 – 271.

Landry C.F., Pribyl T.M., Ellison J.A., Givogri M.I., Kampf K, Campagnoni
C.W., and Campagnoni A.T.   (1998)
Embryonic expression of the myelin basic protein gene: Identification
of a promoter region that targets transgene expression to pioneer
neurons.
J. Neurosci. **18**, (18) 7315 – 7327.

Latchman D.S.   (1990)
Gene regulation: A eukaryotic perspective.
Pub: Unwin Hyman

Latchman D.S.   (1998)
Eukaryotic transcription factors.
Pub: Academic Press (3rd Edition)

Ledeen R. W. (1992)
Enzymes and receptors of myelin.
In Myelin biology and chemistry (Martenson R. P., ed.) 531 – 570.
CRC Press, Roca Raton. Fl.

Lees M. B. and Brostoff S. W. (1984)
In Myelin, 2nd edn. (Morell P., ed.) 174 – 244. Plenum Press

Lemke G. (1986)
Molecular biology of the major myelin genes.
TINS. (6) 266 – 270.

Liu H., Oteiza P. I., Gershwin M. E., Golub M. S. and Keen C. L. (1992)
Effects of maternal marginal zinc deficiency on myelin protein profiles
in the suckling rat and infant rhesus monkey.
Biol. Trace Elem. Res. **34**, (1) 55 – 66.

Lowery J. M., Berti-Martera L. N., Zhu X., Peterson R. G. and Eichberg J.
(1989)
Relationship of ATP turnover, polyphosphoinositide metabolism and
protein phosphorylation in sciatic nerve and derived myelin
subfractions from normal and diabetic rats.
J. Neurochem. **52**, 921 – 932.

Luskin M.B., Pearlman A.M., and Sanes J.R.   (1988)
Cell lineage in the cerebral cortex of the mouse studied *in vivo* and *in
vitro* with a recombinant retrovirus.
Neuron. **1**, 635 – 647.

MacKenzie-Graham A.J., Pribyl T.M, Kim S., Porter V.R., Campagnoni A.T.
and Voskuhl R.R.   (1997)
Myelin protein expression is increased in lymph nodes of mice with
relapsing experimental encephalomyelitis.
J. Immunol. **159**, 4602 – 4610.

Macklin W.B., Campagnoni C.W., Deininger P.L., and Gardinier M.V., (1987)
Structure and expression of the mouse myelin proteolipid protein gene.
J. Neurosci. Res. **18**, 383 - 394.

Marais R. and Marshall C.J., (1996)
Control of the ERK MAP kinase cascade by Ras and Raf.
Cancer Surveys **27**, 101 – 125.

Mathis J.M., Simmons D.M., He X., Swanson L.W. and Rosenfeld M.G. (1992)
Brain-4 – a novel mammalian POU-domain transcription factor exhibiting restricted brain-specific expression.
EMBO J. **11**, 2551 – 2561.

Matthieu J.-M., Ginalski H., Friede R.L., and Cohen S.R. (1980)
Low myelin basic protein levels and normal myelin in peripheral nerves of myelin deficient mutant mice, *shiverer* and *mld*.
Neurosci. **5**, 2315 - 2320.

McCarthy K.D., and deVellis (1980)
Preparation of separate astroglial and oligodendroglial cell cultures from rat cerebral tissues.
J. Cell Biol. **85**, 890 - 902.

McGinnis W, Levine M. S., Hafen E., Kuroiwa A. and Gehring W. J. (1984)
A conserved DNA sequence in homeotic genes of the *Drosophila antennapedia* and bithorax complexes.
Nature **308**, 428 – 433.

Mihara H. and Kaiser E.T. (1988)
A chemically synthesised *antennapedia* homeodomain binds to a specific DNA sequence.
Science **242**, 925 – 926.

Mirsky R., Winter J., Abney E. R., Pruss R. M., Gavrilovic J. and Raff M. C. (1980)
Myelin-specific proteins and glycolipids in rat Schwann cells and oligodendrocytes in culture.
J. Cell Biol. **84**, 483 – 494.

Miura M., Tamura T., Aoyama A., and Mikoshiba K. (1989)
The promoter elements of the mouse myelin basic protein gene function efficiently in NG108-5 neuronal/glial cells.
Gene **75**, 31-38.

Molineaux S.M., Engh H., deFerra F., Hudson L. and Lazzarini R.A.
(1986)
Recombination within the myelin basic protein gene created the
dysmyelinating *shiverer* mouse mutant.
Proc. Natl. Acad. Sci. USA. **83**, 7542 - 7546.

Morrel P., Quarles R. H. and Norton W. T. (1994)
Myelin formation, structure and biochemistry.
In Basic neurochemistry: Molecular, cellular and medical aspects.
(Siegel G. J., Agranoff B. W., Albers R. W. and Molinoff P. B. eds.).
Lippincott-Raven Press.

Moriguchi T., Gotoh Y. and Nishida E.   (1996)
Roles of the MAP kinase cascade in vertebrates.
Adv. Pharm. **36**, 121 – 137.

Moscarello M. A., Wood D. D., Ackerley C. and Boulias C. (1994)
Myelin in multiple sclerosis is developmentally immature.
J. Clin. Invest. **94**, 146 – 154.

Muller M. M., Gerster T. and Schaffner W.   (1988)
Enhancer sequences and the regulation of gene transcription.
European Journal of Biochemistry **176**, 485 – 495.

Muller H.-P., Soga J.W. and Schaffner W.   (1989)
An enhancer stimulates transcription in *trans* when linked to the
promoter via a protein bridge.
Cell **58**, 767 – 777.

Nakajima K., Ikenaka K., Kagawa T., Aruga J., Nakao J., Nakahira K., Shiota
C., Seung K.U., and Mikoshiba K.   (1993)
Novel isoforms of mouse myelin basic protein predominantly
expressed in embryonic stage.
J. Neurochem. **60**, 1554 – 1563.

Newman S.L., Kitamura K., and Campagnoni A.T. (1987a)
Identification of a cDNA coding for a fifth form of myelin basic protein
in mouse.
Proc. Natl Acad Sci USA **84**, 886 - 890.

Newman S.L., Kitamura K., Roth H.J., Kronquist K., Kelero de Rosbo N.,
Crandall B.F., and Campagnoni A.T.   (1987b)
Differential Expression of the Myelin Basic Protein Genes in mouse
and human.
J. Neurochem. **48,** S27.

Otting G., Qian Y., Müller M., Affolter M., Gehring W. and Wuthrich K. (1988)
Secondary structure determination for the *antennapedia*
homeodomain by nuclear magnetic resonance and evidence for a
helix-turn-helix motif.
EMBO J. **7**, 4305 – 4310.

Park G. -H., Chanderkar L.P., Paik W.K., and Kim S. (1986)
Myelin basic protein inhibits histone-specific protein methylase I.
Biochem. Biophys. Acta. **874**, 30 - 36.

Petryniak B., Staudt L.M., Postema C.E., McCormack W.T. and Thompson C.
(1990)
Characterisation of chicken octamer-binding proteins demonstrates
that POU domain-containing homeobox transcription factors have
been highly conserved during vertebrate evolution.
Proc. Natl Acad Sci USA **87**, 1099 – 1103.

Pfeiffer S.E., Barbarese E., and Bhat S. (1981)
Non-coordinate regulation of myelinogenic parameters in primary
cultures of dissociated foetal rat brain.
J. Neurosci. Res. **6**, 369 - 380.

Popko B., Puckett C., Lai E., Shine H.D., Readhead C., Takahashi N., Hunt
III S.W., Sidman R.L., and Hood L. (1987)
Myelin deficient mice: expression of myelin basic protein and
generation of mice with varying levels of myelin.
Cell. **48**, 713 - 721.

Popko B., Puckett C. and Hood L. (1988)
A novel mutation in myekin deficient mice results in unstable myelin
basic protein gene transcripts.
Neuron **1**, 221 – 225.

Pribyl T. M., Campagnoni C. W., Kampf K., Kashima T., Handley V. W.,
McMahon J. and Campagnoni A. T. (1993)
The human myelin basic protein gene is included within a 179-
kilobase transcription unit: Expression in the immune and central
nervous systems.
Proc. Natl. Acad. Sci. USA **90**, 10695 – 10699.

Pritzker L. B., Joshi S., Harauz G. and Moscarello M. A. (2000)
Deimination of myelin basic protein. 2. Effect of methylation of MBP
on its Deimination by peptidylarginine deiminase.
Biochemistry **39**, 5382 – 5388.

Qi Y. and Dawson G. (1993)
Effects of hypoxia on oligodendrocyte signal transduction.
J. Neurochem. **61**, 1097 – 1104.

Raff M.C., Miller R.H., and Noble M.   (1983)
  A glial progenitor cell that develops in vitro into an astrocyte or an
  oligodendrocyte depending on culture medium.
  Nature. **303**, 390 - 396.

Raible D.W. and McMorris F. A.   (1989)
  Cyclic AMP regulates the rate of differentiation of oligodendrocytes
  without changing the lineage commitment of their progenitors.
  Dev. Biol. **133**, 437 – 446.

Rajpurohit R., Lee O-S., Park J-O., Paik W. K. and Kim S. (1994)
  Enzymatic methylation of recombinant Heterogeneous nuclear RNP
  A1. Dual substrate specificity for *S*-adenosylmethionine: histone-
  arginine *N*-methyltransferase.
  J. Biol. Chem. **269**, 1075 – 1082.

Rauscher F.J., Cohen D. R., Curran T., Bos T. J., Bogt P. K., Bohmann D.,
  Tjian R. and Franza B. R. (1988)
  Fos-associated protein p39 is the product of the c-jun oncogene.
  Science **240**, 1010 – 1016.

Readhead C., Popko B., Takahashi N., Shine H.D., Saavedra R.A., Sidman
  R.L., and Hood L.   (1987)
  Expression of a myelin basic protein gene in transgenic *shiverer* mice:
  correction of the dysmyelinating phenotype.
  Cell. **48**, 703 - 712.

Riccio P., Giovannelli S., Bobba A., Romito E., Fasano A., Bleve-Zacheo T.,
  Favilla R., Quagliariello E. and Cavatorta P. (1995)
  Specificity of zinc binding to myelin basic protein.
  Neurochem. Res. **20**, (9) 1107 – 1113.

Roach A., Boylan K., Horvath S., Prusiner S.B., and Hood L. (1983)
  Characterisation of cloned cDNA representing rat myelin basic
  protein: absence of expression in brain of *shiverer* mutant mice.
  Cell. **34**, 799 - 806.

Roach A., Takahashi N., Pravtcheva D., Ruddle F., and Hood L.   (1985)
  Chromosomal mapping of mouse myelin basic protein gene and
  structure and transcription of the partially deleted gene in *shiverer*
  mutant mice.
  Cell. **42**, 149 - 155.

Roch J.-M., Brown-Luedi M., Cooper B.J., and Matthieu J.-M.
  (1986)
  Mice heterozygous for the *mld* mutation have intermediate levels of
  myelin basic protein mRNA and its translation products.
  Mol. Brain Res. **1**, 137 - 144.

248

Roth H.J., Hunkeler M.J., and Campagnoni A.T.   (1985)
    Expression of myelin basic protein genes in several dysmyelinating
    mouse mutants during early postnatal brain development.
    J. Neurochem. **45**, 572 - 580.

Roth H.J., Kronquist K., Pretorius P.J., Crandall P.J., Crandall B.F., and
    Campagnoni A.T.   (1986)
    Isolation and characterisation of a cDNA coding for a novel human
    17.3 k myelin basic protein (MBP) variant.
    J. Neurosci. Res. **16**, 227 - 238.

Roth H.J., Kronquist K.,Kelero de Rosbo N., Crandall B.F., and Campagnoni
    A.T.   (1987)
    Evidence for the expression of four myelin basic protein variants in
    the developing human spinal cord through cDNA cloning.
    J. Neurosci. Res. **17**, 321 - 328.

Ryder K., Lau L.F., and Nathans D. (1988)
    A gene activated by growth factors is related to the oncogene v-jun.
    Proc. Natl Acad Sci USA **85**, 1487 – 1491.

Saccomanno L., Loushin C., Jan E., Punkay E., Artzt K. and Goodwin E. B.
    (1999)
    The STAR protein QKI-6 is a translational repressor.
    Proc. Natl. Acad. Sci. USA **96**, 12605 – 12610.

Salzer J.L., Holmes W.P., and Colman D.R.   (1987)
    The amino acid sequences of the myelin-associated glycoproteins:
    homology to the immunoglobulin gene superfamily.
    J. Cell Biol. **104**, 957 – 965.

Sambrook J., Fritsch E.F., and Maniatis T.,   (1989)
    Molecular Cloning - A Laboratory Manual (Second Edition) **2** 11.31 -
    11.32.

Sanger F., Miklen S., and Coulson A.R. (1977)
    DNA sequencing with chain termination inhibitors.
    Proc. Natl. Acad Sci U.S.A. **74** 5463 - 5467

Saxe D.F., Takahashi N., Hood L., and Simon M.I.   (1985)
    Localisation of the human myelin basic protein gene (MBP) to region
    18q22->qter by *in situ* hybridisation.
    Cytogenet. Cell Genet. **39**, 246 - 249.

Schibler U., Sierra F.   (1987)
    Alternative promoters in developmental gene expression.
    Annu. Rev. Genet. **21** 237-257.

Schoenherr C.J., Paquette A.J., and Anderson D.J. (1996)
Proc Natl Acad Sci U.S.A. 93 9881 – 9886.

Schofield P. N. (1987)
Patterns, puzzles and paradigms – The riddle of the homeobox.
Trends Neurosci. **10**, 3 – 6.

Shan J., Moran-Jones K., Munro T. P., Kidd G. J., Winzor D. J., Hoek K. S.
and Smith R. (2000)
Binding of an RNA trafficking response element to heterogeneous
nuclear ribonucleoproteins A1 and A2.
J. Biol. Chem. **275**, (49) 38286 – 38295.

Sidman R.L., Conover D.S., and Carson J.H. (1985)
Shiverer gene maps near the distal end of chromosome 18 in the
house mouse.
Cytogenet. Cell Genet. **39**, 241 - 245.

Sorg B.J.A., Agrawal D., Agrawal H.C., and Campagnoni A.T. (1986)
Expression of myelin proteolipid protein and basic protein in normal
and dysmyelinating mutant mice.
J. Neurochem. **46**, 379 - 387.

Sorg B.J., Smith M.M., and Campagnoni A.T. (1987)
Development expression of the myelin proteolipid protein and basic
protein mRNAs in normal and dysmyelinating mutant mice.
J. Neurochem. **49**, 1146 - 1154.

Sparkes R.S., Mohandas T., Heinzmann C., Roth H.J., Klisak L., and
Campagnoni A.T. (1987)
Assignment of the myelin basic protein gene to human chromosome
18q22->qter.
Hum. Genet. **75**, 147 - 150.

Streicher R. and Stoffel W. (1989)
The organisation of the human myelin basic protein gene.
Biol. Chem. Hoppe-Seyler **370**, 503 – 510.

Surtees R. (1993)
Biochemical pathogenesis of sub acute combined degeneration of the
spinal cord and brain.
J. Inherited Metab. Dis. **16**, 762 – 770.

Sutcliffe J.G., (1987)
The genes for myelin.
Trends in Genetics. **3**, (3) 73 - 76.

Tabor S. And Richardson C.C. (1987)
DNA sequence analysis with a modified bacteriophage T7 DNA Polymerase.
Proc. Natl. Acad Sci U.S.A. **84** 4767

Takahashi N., Roach A., Teplow D.B., Prusiner S.B., and Hood L. (1985)
Cloning and characterisation of the myelin basic protein gene from mouse: one gene can encode both 14 kd and 18.5 kd MBPs by alternate use of exons.
Cell. **42**, 139 - 148.

Takahashi K., Vigneron M., Matthes H., Wildeman A., Zeake M. and Chambon P. (1986)
Requirement of stereospecific alignments for initiation from the simian virus 40 early promoter.
Nature **319**, 121 – 126.

Tamura T., Miura M., Ikenaka K., and Mikoshiba K. (1988)
Analysis of transcription control elements of the mouse myelin basic protein gene in HeLa cell extracts: demonstration of a strong NF1-binding motif in the upstream region.
Nuc. Acids Res. **16**, (24) 11441 – 11459.

Tamura T., Sumita K., Hirose S., and Mikoshiba K. (1990)
Core promoter of the mouse myelin basic protein gene governs brain-specific transcription *in vitro.*
EMBO J. **9**, 3101 – 3108.

Tatusova T.A. and Madden T.L. (1999)
Blast 2 sequences - a new tool for comparing protein and nucleotide sequences.
FEMS Microbiol Lett. **174**, 247 – 250.

Taveggia C., Pizzagalli A., Feltri M.L., Grinspan J.B., Kamholz J., and Wrabetz L. (1998)
MEBA depresses the proximal myelin basic protein promoter in oligodendrocytes.
J. Biol. Chem. **273**, (42), 27741 – 27748.

Temple S., and Raff M.C. (1985)
Differentiation of a bipotential glial progenitor cell in single cell microculture.
Nature. **313**, 223 - 225.

Theill L.E., Hattori K., Lazarro D., Castrillo J.-L. and Karin M. (1992)
Differential splicing of GHF-1 primary transcript gives rise to two functionally distinct homeoproteins.
EMBO J. **11**, 2261 – 2269.

251

Torron J., Ljubetic C.I., Huang L., Kimbro K.S., Dyer C. and Saavedra.
(1997)
Two proteins bind to a novel motif in the promoter of the myelin basic
protein gene from mouse.
J Mol. Neurosci. **8**, 181 – 191.

Townsend L.E., and Benjamins J.A.   (1983)
Effects of monensin on posttranslational processing of myelin
proteins.
J. Neurochem. **40**, 1333 - 1339.

Trapp B. D. (1990)
Distribution of myelin protein gene products in actively-myelinating
oligodendrocytes.
NATO ASI Series H **43**, 59 – 79.

Tretiakova A., Gallia G.L., Shcherbik N., Jameson B., Johnson E.M., Amini
S., and Khalili K.   (1998)
Association of Purα with RNAs homologous to 7SL determines its
binding ability to the myelin basic protein promoter DNA sequence.
J. Biol. Chem. **273**, (35), 22241 – 22247.

Tretiakova A., Steplewski A., Johnson E.M., Amini S., and Khalili K.(1999)
Regulation of myelin basic protein gene transcription by Sp1 and
Purα: evidence for association of Sp1 and Purα in brain.
J. Cell Physiol. **181**, (1), 160 – 168.

Tsang D., Tsang Y. S., Ho W. K. K. and Wong R. N. S. (1997)
Myelin basic protein is a zinc-binding protein in brain: Possible role in
myelin compaction.
Neurochem. Res. **22**, (7) 811 – 819.

Ueno S., Handley V. W., Byravan S. and Campagnoni A. T. (1994a)
Structural features of myelin basic protein mRNAs influence their
translational efficiencies.
J. Neurochem. **62**, 1254 – 1259.

Ueno S., Kotani Y., Kondoh K., Sano A., Kakimoto Y. and Campagnoni A. T.
(1994b)
The 3'-untranslated region of mouse myelin basic protein gene
increases the amount of mRNA in immortalised mouse
oligodendrocytes.
Biochem. Biophys. Res. Commun. **204**, 1352 – 1357.

Umemeori H., Kadowaki Y., Hirosawa K., Yoshida Y., Hironaka K., Okano H
and Yamamoto T. (1999)
Stimulation of myelin basic protein gene transcription by fyn tyrosine
kinase for myelination.
J. Neuroscience **19**, (4) 1393 – 1397.

Ushiro H., Tsutsumi T., Suzuki K., Kayahara T. and Nakano K. (1998)
Molecular cloning and characterisation of a novel Ste20-related
protein kinase enriched in neurons and transporting epithelia.
Arch. Biochem. Biophys. **355**, 233 – 240.

Vernet C. and Artzt K. (1997)
STAR, a gene family involved in signal transduction and activation of
RNA.
Trends Genet. **13**, 479 – 484.

Vogt P. K. and Bos T.J. (1990)
jun: oncogene and transcription factor.
Adv. Cancer Res. **55**, 1 – 35.

Voskhul R. R.　　　(1998)
Myelin protein expression in lymphoid tissues: implications for
peripheral tolerance.
Immunol. Rev. **164**, 81 – 92.

Waxman, S.G. and Ritchie, J.M. (1993)
Molecular dissection of the myelinated axon.
Ann. Neurol. **33**, 121 – 136.

Whitaker J. N., Layton B. A., Herman P. K., Kachelhofer R. D., Burgard S.
and Bartolucci A. A. (1993)
Correlation of myelin basic protein-like material in cerebrospinal fluid
of multiple sclerosis patients with their response to glucocorticoid
treatment.
Ann. Neurol. **33**, 10 – 17.

Whitaker J. N., Williams P. H., Layton B. A., M°Farland H. F., Stone L. A.,
Smith E. M., Kachelhofer R. D., Bradley E. L., Burgard S., Zhao G.
and Paty D. W. (1994)
Correlation of clinical features and findings on cranial magnetic
resonance imaging with urinary myelin basic protein-like material in
patients with multiple sclerosis.
Ann. Neurol. **35**, 577 – 585.

Wilson R. and Brophy P. J. (1989)
Role for the oligodendrocyte cytoskeleton in myelination.
J. Neurosci. Res. **22**, 439 – 448.

Wolf M. K., Nunnari J. N. and Billings-Gagliardi S. (1999)
Quaking*Shiverer double mutant mice survive for at least 100 days
with no CNS myelin.
Dev. Neurosci. **21**, 483 – 490.

Wood D. D., Bilbao J., O'Connor P. and Moscarello M. A. (1996)
   Acute multiple sclerosis (Marburg Type) is associated with
   developmentally immature myelin basic protein.
   Ann. Neurol. **40**, 18 – 24.

Wrabetz L., Shumas S., Grinspan J., Feltri M.L., Bozyczko D., McMorris
   F.A., Pleasure D., and Kamholz J.   (1993)
   Analysis of the human MBP promoter in primary cultures of
   oligodendrocytes: positive and negative cis-acting elements in the
   proximal MBP promoter mediate oligodendrocyte-specific expression
   of MBP.
   J. Neurosci. Res. **36**, 455 – 471.

Young P. R., Vacante D. A. and Waickus C. M. (1987)
   Mechanism of the interaction between myelin basic protein and the
   myelin membrane: The role of arginine methylation.
   Biochem. Biophys. Res. Commun. **145**, 1112 – 1118.

Yu Y.-T., and Campagnoni A.T.   (1982)
   In vitro synthesis of the four mouse myelin basic proteins: evidence
   for the lack of a metabolic relationship.
   J. Neurochem. **39**, 1559 - 1568.

Yu J-S. and Yang S. D. (1994)
   Protein kinase F_A/glycogen synthase kinase 3 predominantly
   phosphorylates the *in vivo* site Thr97-Pro in myelin basic protein:
   evidence for the Thr-Pro and Ser-Arg-X-X-Ser as consensus
   sequence motifs.
   J. Neurochem. **62**, 1596 – 1603.

Yung S. D. (1986)
   Identification of the ATP-Mg dependent protein phosphatase activator
   (F_A) as a myelin basic protein kinase in the brain.
   J. Biol. Chem. **261**, 11786 – 11791.

Zelenika D., Grima B. and Pessac B. (1993)
   A new family of transcripts of the myelin basic protein gene:
   Expression in brain and in immune system.
   J. Neurochem. **60**, 1574 – 1577.

Zhang X., and Miskimins R.   (1993)
   Binding at an NF1 site is modulated by cyclic AMP-dependent
   activation of myelin basic protein gene expression.
   J. Neurochem. **60**, 2010 – 2017.

APPENDIX I.

EQUIPMENT SUPPLIERS.

Mini-Gel Submarine Electrophoresis Units.

     Biometra Ltd, P.O. Box 42, Sale, Manchester.

     Gallenkamp Ltd. (Fisons Plc Scientific Equipment Division), Bishop Meadow, Loughborough, Leicester.

DNA Sequencing Gel Apparatus.

     Biometra Ltd, P.O. Box 42, Sale, Manchester.

Mini-Gel Submarine Electrophoresis Units.

     Gallenkamp Ltd. (Fisons Plc Scientific Equipment Division), Bishop Meadow, Loughborough, Leicester.

Mini-Gel Electrophoresis Power Supply.

     Biometra Ltd, P.O. Box 42, Sale, Manchester.

     Gallenkamp Ltd. (Fisons Plc Scientific Equipment Division), Bishop Meadow, Loughborough, Leicester.

Sequencing Power Supply.

     Pharmacia Ltd, 351 Midsummer Boulevard, Milton Keynes, Buckinghamshire.

PCR Machines.

     Biometra Ltd, P.O. Box 42, Sale, Manchester.

     MWG Biotech Ltd,

APPENDIX II.

MEDIA AND BUFFER COMPOSITION.

2.2.   Large scale isolation of plasmid DNA by a modified Clewell and Helinski technique.

Luria Broth:

Tryptone	10 g
Yeast Extract	5 g
Sodium Chloride	5 g
GLUCOSE	2 g
Water	make up to 1 litre.

Buffer I:

Glucose	50	mmol l^{-1}
Tris-HCl (pH 8.0)	25	mmol l^{-1}
EDTA (pH8.0)	10	mmol l^{-1}

SOLUTION I can be prepared in batches of 100 ml, autoclaved for 15 min at 10 lb/in^2 and stored at 4° C.

3X Triton X-100 Lysis Solution:

Triton X-100 (10 %)	300	μl
EDTA (2.5 M)	750	μl
Tris-HCl (pH 8.0) (1M)	1500	μl
WATER	700	μl

2.3.   Purification of plasmid DNA by centrifugation to equilibrium in caesium chloride-ethidium bromide gradient.

T.E. Buffer (pH8.0):

Tris-HCl (pH 8.0)	10	m mol l^{-1}
EDTA (0.5M)	1	m mol l^{-1}

2.4. Analytical agarose gel electrophoresis of DNA.

10X Tris acetate buffer:

Tris (SIGMA 7-9)	32.3 g
EDTA	2.48 g
Glacial Acetic Acid	11.7 ml
WATER	make up to 1 litre.

Loading Buffer:

Glycerol	25 %
Xylene Cyanol FF	0.02 %
Bromophenol Blue	0.05 %

2.6.3. Ligation.

10X Ligation Buffer:

Tris-HCl	200	mmol l^{-1}
Magnesium Chloride	100	mmol l^{-1}
Dithiothreitol	100	mmol l^{-1}
WATER (Sterile)	600	µl

Dispense into 200 µl aliquots and store frozen until required.
Fresh A.T.P. is added (from a 10 mmol l^{-1} stock) just before use.

Incubation Buffer:

Tris-HCl	20	mmol l^{-1}
Magnesium Chloride	10	mmol l^{-1}
Dithioerythritol	10	mmol l^{-1}
A.T.P.	0.6	mmol l^{-1}
Adjust the pH TO 7.6		

2.6.4. Introduction of plasmid DNA into *E. coli* cells.

M9 Minimal Medium:
5X M9 Salts:

Disodium hydrogen orthophosphate ($7H_2O$)	64	gl^{-1}
Potassium dihydrogen orthophosphate	15	gl^{-1}
Sodium chloride	2.5	gl^{-1}
Ammonium chloride	5.0	gl^{-1}

The salt solution is divided into 200 ml aliquots prior to sterilisation by autoclaving at 15 psi for 15 min.

To 750 ml of sterile deionised water cooled to 50° C the following was added aseptically:

200 ml of 5X M9 salts
20 ml of 20 % Glucose (filter sterilised)
Make up to 1 litre with sterile deionised water added aseptically.

20X SSC Buffer:

Sodium Chloride	175 g
Tri-sodium Citrate	88.2 g

Dissolve in 800 ml of water, adjust the pH to 7.0 with 10 normal sodium hydroxide and then make up to volume (1 litre).
Sterilize by autoclaving.

AIX Medium:

Allow L-B agar to cool to 50 °C before adding:

I.P.T.G. (1M)	300 µl
X-GAL (10%)	80 µl
Ampicillin (1%)	2.5 ml

for each 100 ml OF L-B agar.

L-B Agar:

Tryptone	10 g
Yeast Extract	5 g
Sodium Chloride	10 g
Agar	10 g

Add 900 ml of water, adjust the pH TO 7.5 with sodium hydroxide and then make up to volume (1 litre).

Stock I.P.T.G. (1 M):
> Dissolve 1 g I.P.T.G. in 4.2 ml of sterile distilled water and store frozen until required.

Stock X-GAL (10%):
> Dissolve 100 mg X-GAL in 1000 μl of Dimethylformamide.

2.6.5.1.Preparation of plasmid DNA on a small-scale.

STET Buffer:

Sodium Chloride	100 mmol l^{-1}
Tris-HCl (pH 8.0)	10 mmol l^{-1}
EDTA (pH 8.0)	1 mmol l^{-1}
Triton X-100	5 %

3.1.1. Alkaline lysis mini-preparation using the Pharmacia kit.

SB Medium:

Tryptone	12 g
Yeast Extract	24 g

Dissolved in 900 ml of distilled water.

Prepare 0.9 mol l^{-1} Potassium Phosphate buffer (pH 7.3) by dissolving 2.3 g of KH_2PO_4 and 12.5 g of K_2HPO_4 in 100 ml of water, sterilize both solutions separately by autoclaving and mix them aseptically together and add 4 ml of sterile glycerol.

Add ampicillin to a final concentration of 50 μg/ml for selection of the plasmid of interest.

Solution I:

Tris-HCl (pH 8.0)	50	mmol l^{-1}
Glucose	50	mmol l^{-1}
EDTA	10	mmol l^{-1}
KATHON® CG/ICP	0.15 %	

Solution II:

Sodium Hydroxide	0.8 mol l^{-1}
Triton X-100	4 % (v/v)

It should be noted that before using solution II for the first time, 75 ml of sterile distilled water should be added, this will bring the final concentration to 0.2 mol l^{-1} NaOH and 1% TRITON X-100.

Solution III:

Sodium Acetate (pH 4.8)	3 mol l^{-1}

Column Buffer:

Tris-HCl (pH 8.0)	10 mmol l^{-1}
EDTA	1 mmol l^{-1}

4.0. Sequencing of DNA.

T.E. Buffer pH 7.5 (for reconstitution of oligonucleotides only).

Tris-HCl pH7.5	10 m mol l^{-1}
EDTA	1 m mol l^{-1}

4.2. DNA sequencing using the Sequenase kit.

Sequenase Buffer (5X concentrate).

Tris-HCl pH7.5	200	mM
Magnesium chloride	100	mM
Sodium chloride	250	mM

Dithiothreitol	0.1 M

Enzyme Dilution Buffer.

Tris-HCl pH7.5	10	mM
DTT	5	mM
BSA	0.5 mg/ml	

Stop Solution.

Formamide	95	%
EDTA	20	mM
Bromophenol Blue	0.05	%
Xylene Cyanol FF	0.05	%

4.4.1. Preparation of polyacrylamide gels.

40 % Acrylamide Solution:

Acrylamide	114 g
Bis-acrylamide	6 g

Make up to 300 ml with deionised water.
When dissolved stir gently with 6 g of amberlite MB-1 mixed bed resin. Filter through glass wool to remove resin and store solution at 4° C until required.

10X TBE Buffer:

Tris	54	g
Boric Acid	27.5	g
EDTA (0.5 mol l^{-1} STOCK)	20	ml

Make up to 500 ml with sterile distilled water.

1X TBE Gel Mix:

10X TBE	20.8	ml
Acrylamide (40%)	31.2	ml
Urea (Ultrapure Grade)	100	g

Make up to 208 ml with sterile distilled water.

6% Acrylamide Sequencing Gel:

1X TBE GEL MIX	70	ml
Ammonium Persulphate (25%)	140	μl
TEMED	90	μl

The gel should set within 15 min otherwise the whole procedure must be repeated.

5.0. Electrophoretic mobility shift assay.

5X Annealing Buffer:

100	mM	Sodium chloride
10	mM	Tris.Cl (pH 7.8)
1	mM	EDTA
100	mM	Magnesium chloride
5	mM	DTT

Binding Buffer:

25 mM Hepes.KOH (pH 8.0)
0.5 mM EDTA
10 % (v/v) Glycerol
0.01 mM zinc chloride
0.5 mM dithiothreitol

5.2. Preparation of nuclear extracts.

Buffer A:

10	mM Hepes (pH 7.9)
1.5	mM magnesium chloride
10	mM potassium chloride
0.5	mM PMSF
0.5	mM dithiothreitol

Buffer C:

20	mM Hepes (pH 7.9)
0.42	mM sodium chloride
1.5	mM magnesium chloride
0.2	mM EDTA
0.5	mM dithiothreitol
25	% v/v Glycerol

Buffer D:

20	mM Hepes (pH 7.9)
100	mM potassium chloride
0.2	mM EDTA
0.5	mM PMSF
0.5	mM dithiothreitol
20	% v/v Glycerol

262

It should be noted that DTT and PMSF were added fresh at the time of use.

7.0. Large scale preparation of yeast vector containing the binding site.

Terrific broth:

Tryptone	12	g
Yeast extract	24	g
Glycerol	4.0	ml
Water	900	ml

Autoclave to sterilise, when cooled to room temperature aseptically add 100 ml of phosphate salt solution.

Phosphate salt solution:

Potassium dihydrogen orthophosphate	2.31	g
Dipotassium hydrogen orthophosphate	12.54	g
Water	100	ml
Autoclave to sterilise		

Solution A:

Glucose	9.0	g
1M Tris (pH 8.0)	12.5	ml
0.5 M EDTA	10.0	ml
Water	475	ml

Autoclave to sterilise.

Solution B: (Prepared fresh just prior to use)

10M Sodium hydroxide	1.6	ml
10% SDS	8.0	ml
Water	70.4	ml

Solution C:

Potassium acetate	147	g
Acetic acid	58	ml

Make up to a final volume of 500 ml
Autoclave to sterilise, and store at 4° C

PEG 8000

Dissolve 5 g of Polyethylene glycol 8000 in 15 ml of water then make up to a final volume of 25 ml.

8.2. Preparation of yeast competent cells.

YPD Medium (For Liquid Cultures):

20 gl^{-1} Difco Peptone
10 gl^{-1} Difco Yeast Extract

Adjust the pH to 5.8 and make up to a final volume of 950 ml.
Autoclave to sterilise.
Cool to 55 °C prior to aseptically adding 50 ml of filter sterilised 40% glucose

YPD Medium (For Plates):

20 gl^{-1} Difco Peptone
10 gl^{-1} Difco Yeast Extract
20 gl^{-1} Difco Agar

Adjust the pH to 5.8 and make up to a final volume of 950 ml.
Autoclave to sterilise.
Cool to 55 °C prior to aseptically adding 50 ml of filter sterilised 40% glucose

10 X TE Buffer:

0.1 M   Tris.Cl
10 mM EDTA

Adjust the pH to 7.5 and make up to 100 ml with distilled water.
Autoclave to sterilise.

10 X LiAc:

Lithium Acetate   10.2 g
Adjust the pH to 7.5 with dilute acetic acid and make up to 100 ml with distilled water.
Autoclave to sterilise.

264

1 X TE/LiAc Solution (prepare fresh just before use):
    Water   8 ml of Sterile Water
    TE Buffer   1 ml of 10 X TE
    LiAc   1 ml of 10 X LiAc

50 % PEG 4000:
    Polyethylene glycol (avg. mol. wt. = 3,350)   50 g
    Make up to 100 ml with distilled water.
    Autoclave to sterilise.

1 X PEG/LiAc Solution (prepare fresh just before use):
    PEG 4000   8 ml of 50 % PEG
    TE Buffer   1 ml of 10 X TE
    LiAc   1 ml of 10 X LiAc

TE Buffer:
        1M   Tris.Cl (pH 8.0)   1000 μl
     500 mM   EDTA (pH 8.0)   200 μl

Make up to 100 ml with sterile distilled water.

SD Medium (For Liquid Medium):
        6.7 gl^{-1} Difco yeast nitrogen base without amino acids

Adjust the pH to 5.8 and make up to a final volume of 850 ml.
Autoclave to sterilise.
Cool to 55 °C prior to aseptically adding:
        100 ml of 10X dropout medium
        50 ml of filter sterilised 40% glucose

SD Medium (For Plates):
        6.7 gl^{-1} Difco yeast nitrogen base without amino acids
    20   gl^{-1} Difco agar

Adjust the pH to 5.8 and make up to a final volume of 850 ml.
Autoclave to sterilise.
Cool to 55 °C prior to aseptically adding:
        100 ml of 10X dropout medium

50 ml of filter sterilised 40% glucose

For medium containing 3 - AT, powdered 3 - Aminotriazole was added to the appropriate concentration after the medium had cooled and before the plates were poured.

10X Dropout Medium:

L-Isoleucine	300 mg l^{-1}
L-Valine	1500 mg l^{-1}
L-Adenine hemisulphate salt	200 mg l^{-1}
L-Arginine HCl	200 mg l^{-1}
L-Leucine	1000 mg l^{-1}
L-Lysine HCl	300 mg l^{-1}
L-Methionine	200 mg l^{-1}
L-Phenylalanine	500 mg l^{-1}
L-Threonine	2000 mg l^{-1}
L-Tryptophan	200 mg l^{-1}
L-Tyrosine	300 mg l^{-1}
L-Uracil	200 mg l^{-1}

Make up to a final volume of 1000 ml.
Autoclave to sterilise.
Keep refrigerated until required.

3 - AT Concentrations:

For each 100ml of SD medium add:

Concentration (mM)	Amount 3 - AT (g/100 ml of medium)
5 mM	0.043 g
10 mM	0.086 g
15 mM	0.13 g
30 mM	0.25 g
45 mM	0.38 g
60 mM	0.5 g

8.4. Screening of putative reporter strain constructs (rapid isolation of chromosomal DNA).

Breaking Buffer:

	2% (v/v)	Triton X-100
	1% (v/v)	Sodium Dodecyl Sulphate (SDS)
100 mM	NaCl	
10 mM	Tris.Cl (pH 8.0)	
1 mM	EDTA (pH 8.0)	

Can be stored at room temperature for up to 1 yr.

APPENDIX III.

PREPARATION OF SEPHACRYL S-400 COLUMNS.

Each spun column was prepared as follows:

The column was inverted several times to resuspend the Sephacryl S-400 gel and then set upright in a rack to allow the gel to settle to form a continuous bed.

The Column was allowed to drain prior to the addition of 2 ml of Column Buffer.

The column was placed in a 15 ml Corex tube and centrifuged for 2 min at approximately 400 x g in a swinging bucket rotor and the eluted buffer discarded from the tube.

An uncapped 1.5 ml Eppendorf tube was placed in the bottom of the Corex tube, the column was replaced inside the Corex tube with the tip inside the Eppendorf tube.

The column was then used for DNA purification.

APPENDIX IV.

OLIGONUCLEOTIDE SYNTHESIS.

The procedure used for the synthesis of oligonucleotides was the
phosphoramidite version of the phosphite triester method carried out on a
solid phase support. The support material was inert towards all the reagents
and solvents used in the synthesis cycle.

Four fully protected nucleotides corresponding to the bases A, C, G
and T, (the basic building blocks in the synthesis), and referred to as
amidites, were required,. An example of the structure of a fully protected
amidite is shown in Figure 39. The complete reaction cycle itself consisted of
four main steps, each separated by column washing as shown in Figure 40:

a. DEPROTECTION.

This is the removal of the dimethoxytrityl (DMTr) group to expose the
hydroxyl group at position 5 of the 2-deoxyribose unit, and the subsequent
activation of the phosphoramidite by tetrazole.

b. COUPLING.

The formation of an internucleotide bond between the 5' hydroxyl group and
the phosphorous containing group at the 3' position on the incoming
protected nucleotide.

269

c. CAPPING.

The termination of any unreacted 5' hydroxyl groups.

d. OXIDATION.

The phosphite linkage is oxidised to phosphate by iodine.

Each time this procedure is repeated one nucleotide is added to the growing chain, until an oligonucleotide of the required sequence and length has been synthesised.
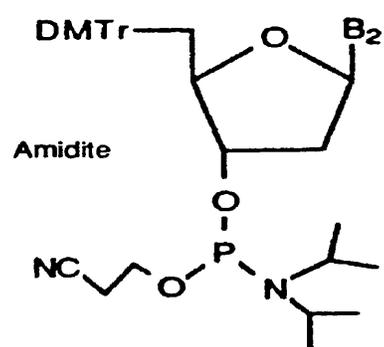
FIGURE 39: The structure of a fully protected amidite.
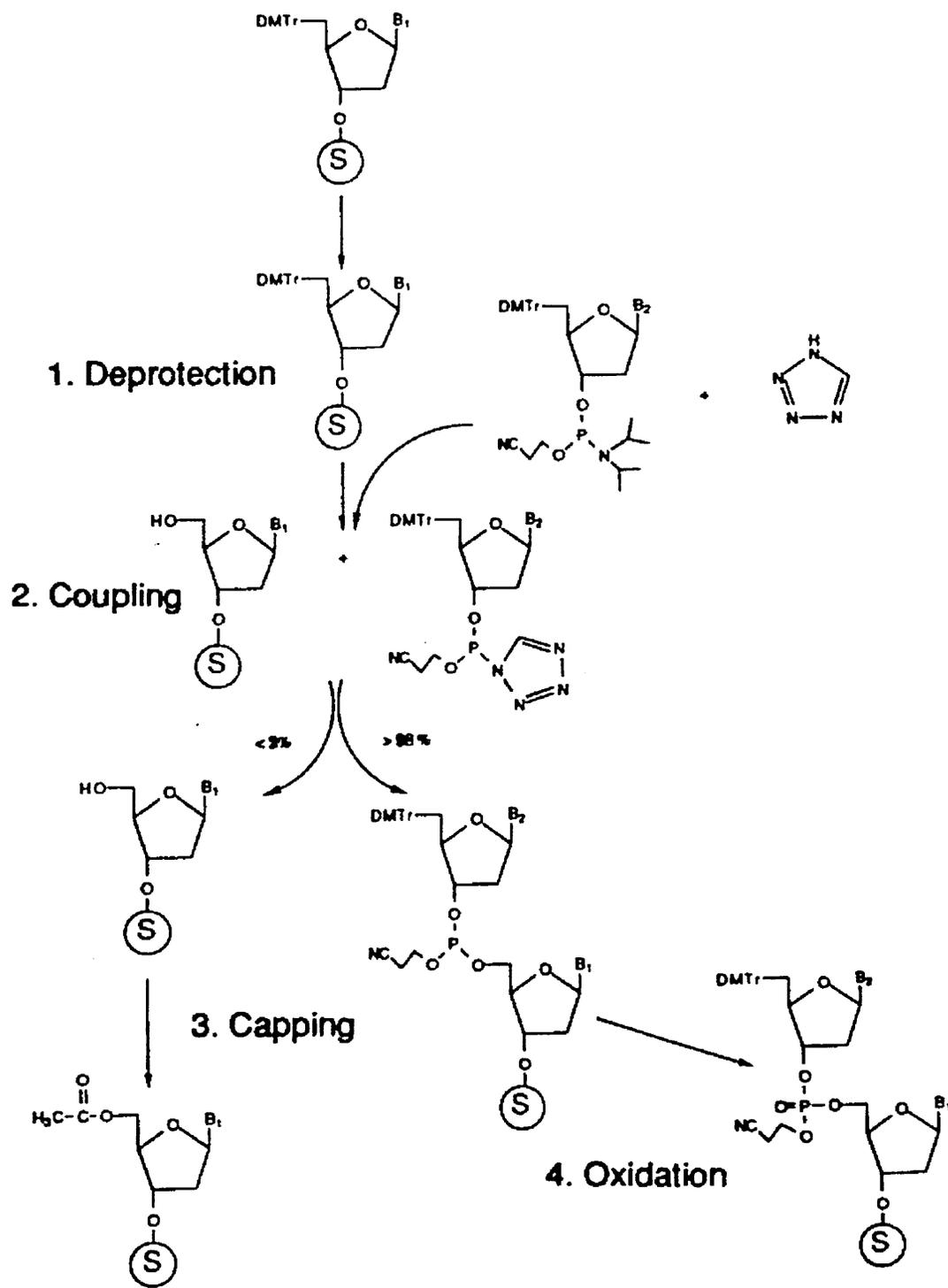
FIGURE 40: The complete synthesis cycle.

272

COUPLING

The phosphite triester method makes use of the great reactivity of the trivalent phosphorous, in this state it will be electron deficient and therefore readily accepts electrons from an electron donating group such as the 5'-hydroxyl group of the deoxyribose unit. Unfortunately any other similarly donating group(s) would interfere with the efficiency of this reaction. However, the group of most concern is the hydroxyl group found in water and as a consequence all coupling reactions were carried out under totally anhydrous conditions. All reagents and solvents used in this reaction were also stored under anhydrous conditions.

The first step was the removal of the dimethoxytrityl (DMTr) protecting group from the 5' hydroxyl of the initial nucleoside (as illustrated in Figure 41). The absorbance value of the orange DMTr ion produced was monitored at 436 nm and used to calculate the efficiency of coupling in the previous cycle. Obviously, no such calculation could be performed for the first detritylation.

The acidic conditions used in detritylation can also cause depurination, (this is the cleavage of the purine base from its deoxyribose at the glycosidic bond). Such a detrimental reaction would not be picked up by the DMTr assay since normal release of the DMTr cation would have taken place at the same time.
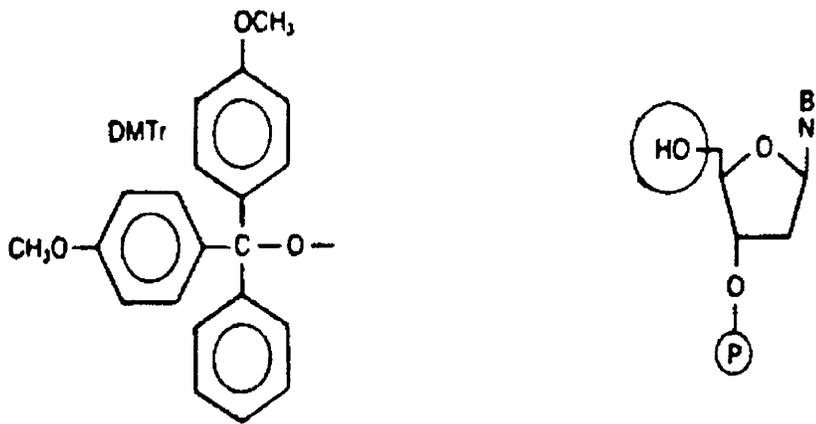
FIGURE 41: The detritylation reaction.

POST-SYNTHETIC PROCEDURES

Three further steps were needed before the final oligonucleotide was ready for use:

(a.) Deprotection.

(b.) Cleavage of the oligonucleotide from its support.

(c.) Purification.

It was at this stage that the advantage of the β-cyanoethyl group became apparent for it allowed both deprotection and cleavage to be carried out at the same time.

Deprotection is the removal of the protecting amide groups from the bases and the β-cyanoethyl group from the phosphorous, this process would normally take 16 h at 50° C, however, the use of PAC amidites and supports allows total deprotection and cleavage in 1 h at 70° C.

Ammonia cleaves the oligonucleotide from its support by breaking the ester linkage between the spacer arm and the 3' hydroxyl group of the initial nucleoside, thus leaving the cleaved DNA with a free 3' hydroxyl group.

During the synthesis cycle all 5' hydroxyls, which did not undergo successful coupling, were capped as acetates. These were now removed by the ammonia to leave a free hydroxyl group. The full length molecule, which of course was not acetylated, has its 5' hydroxyl protected by a DMTr group. Usually this DMTr is cleaved automatically as the final step in the synthesis

275

cycle but it may have been left uncleaved if reverse phase HPLC or other purification methods requiring its presence were to be employed.

This deprotected and detritylated synthetic oligonucleotide DNA has free 5' and 3' hydroxyls and is biologically active. Ion-exchange FPLC, reverse phase HPLC or PAGE electrophoresis are the usual methods of purification, but as the majority of oligonucleotides produced are less than 30 bases in length then nothing more complicated than desalting is usually required. This can be accomplished by size exclusion chromatography on a Sephadex column (for example NAP-10) or by ethanol precipitation.

APPENDIX V.

COP PURIFICATION OF 5' TRITYL-ON OLIGONUCLEOTIDES

The COP purification cartridges separate molecules using the same principle as reverse phase chromatography, that is the less hydrophobic the molecule the quicker it will elute from the column.

Since the desired oligonucleotide product possesses the hydrophobic dimethoxytrityl (DMTr) protecting group on the 5' hydroxyl group, and the truncated sequences did not, it gives the desired product more hydrophobicity and therefore enabled purification on the COP cartridge.

Following synthesis the trityl-ON oligonucleotide was deprotected as previously described in section 2.13.3.

After removing the plunger from a 5 ml disposable syringe the main body of the syringe was then connected to the COP cartridge and using the graduations 2 ml of acetonitrile was passed through the column at a rate of 1 to 2 drops per second using the plunger.

The cartridge was then flushed with 2 ml of 2.0 M triethylamine acetate (pH 7.0). Prior to loading the oligonucleotide onto the column 3 ml of water was added to the ammonium hydroxide solution containing the deprotected trityl-ON oligonucleotide.

The oligonucleotide solution was passed through the column, the eluate collected and then repassed through the column a second time.

Dilute ammonium hydroxide (3 ml) was flushed through the column followed by 2 ml of deionised water, 2 ml of 2 % trifluoroacetic acid to detritylate the support-bound oligonucleotide and a further 2 ml of water.

Recovery of the oligonucleotide was achieved by flushing the column with 2 ml of 20 % (v/v) acetonitrile/water.

The absorbance units recovered was determined at 260 nm from a dry sample resuspended in water.

Vector Maps and Sequences.
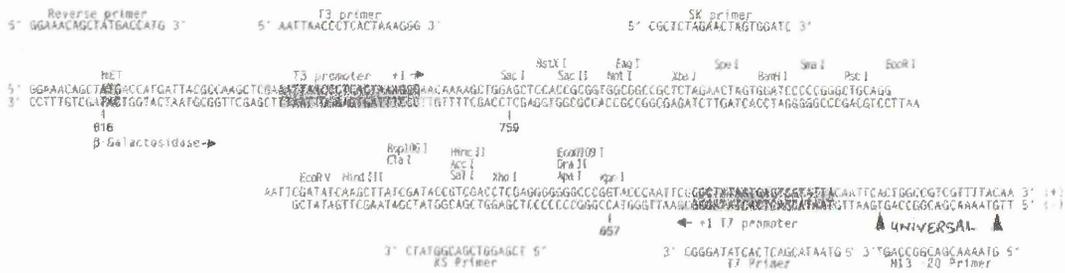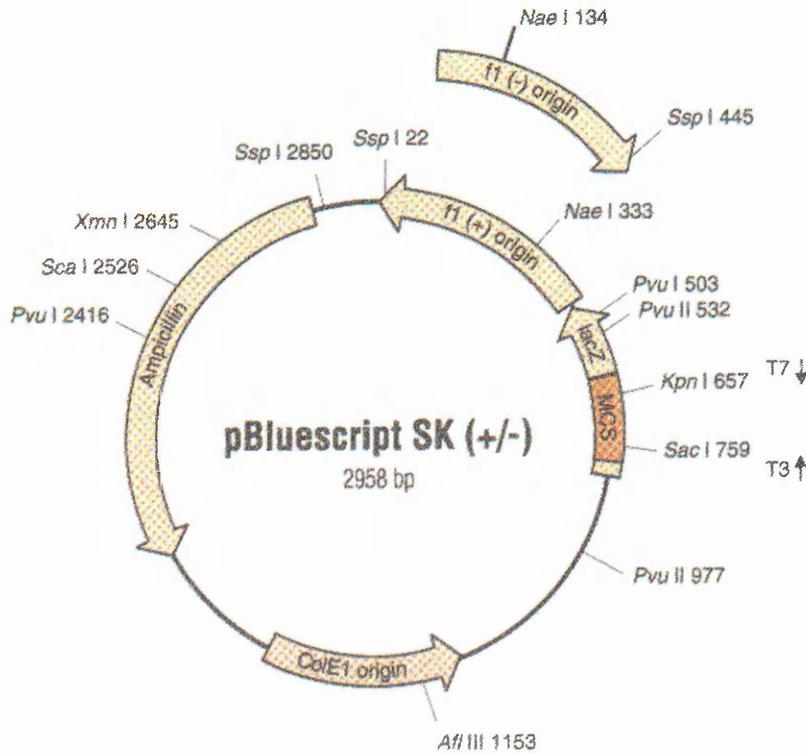
FIGURE 42.Map and multiple cloning site of plasmid pBluescript.

FIGURE 43. Map of pHISi-I and multiple cloning site.
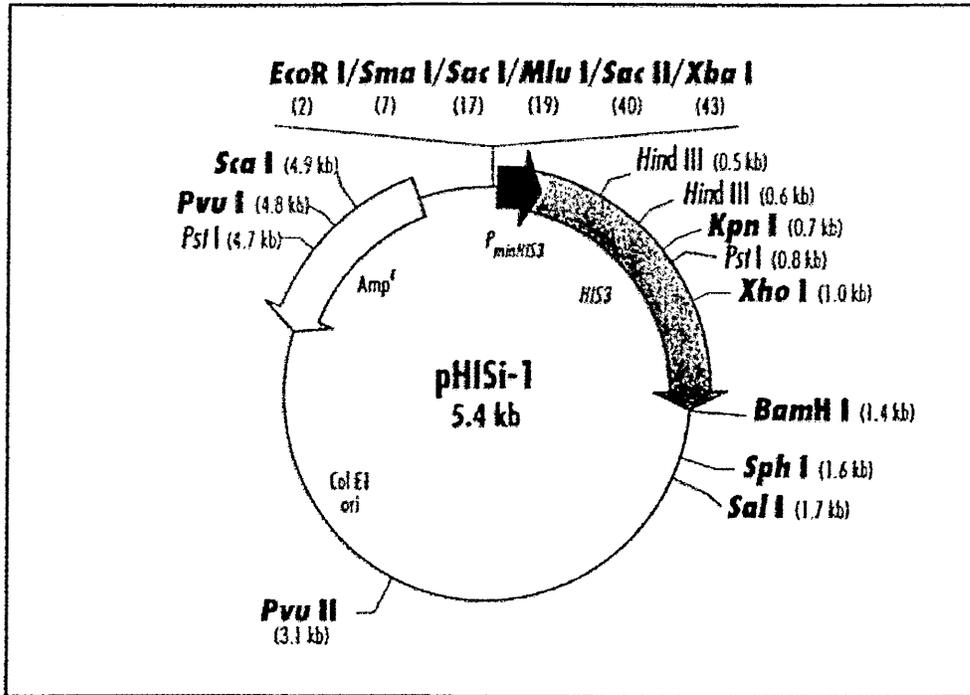
```
        EcoRI     SmaI      SacI          NruI  ClaI      SstII  XbaI
         |         |         |             |     |        SacII
GAATTCCCGGGGAGCTCACGCGTTCGCGAATCGATCCGCGGTCTAGAAATTCCTGGCATTATCACATAAT
CTTAAGGGCCCCTCGAGTGCGCAAGCGCTTAGCTAGGCGCCAGATCTTTAAGGACCGTAATAGTGTATTA
        10        20        30        40        50        60        70

GAATTATACATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAA
CTTAATATGTAATATATTTCATTACACTAAAGAAGCTTCTTATATGATTTTTTACTCGTCCGTTCTATTT
        80        90        100       110       120       130       140

CGAAGGCAAAGATGACAGAGCAGAAAGCCCTAGTAAAGCGTATTACAAATGAAACCAAGATTCAGATTGC
GCTTCCGTTTCTACTGTCTCGTCTTTCGGGATCATTTCGCATAATGTTTACTTTGGTTCTAAGTCTAACG
        150       160       170       180       190       200       210

        DraI
GATCTCTTTAAAGGGTGGTCCCCTAGCGATAGAGCACTCGATCTTCCCAGAAAAAGAGGCAGAAGCAGTA
CTAGAGAAATTTCCCACCAGGGGATCGCTATCTCGTGAGCTAGAAGGGTCTTTTTCTCCGTCTTCGTCAT
        220       230       240       250       260       270       280

                                                                NdeI
GCAGAACAGGCCACACAATCGCAAGTGATTAACGTCCACACAGGTATAGGGTTTCTGGACCATATGATAC
CGTCTTGTCCGGTGTGTTAGCGTTCACTAATTGCAGGTGTGTCCATATCCCAAAGACCTGGTATACTATG
        290       300       310       320       330       340       350

ATGCTCTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGTGACTTACACATAGACGACCA
TACGAGACCGGTTCGTAAGGCCGACCAGCGATTAGCAACTCACGTAACCACTGAATGTGTATCTGCTGGT
        360       370       380       390       400       410       420

                                       DraI
                               HindIII               BglI
TCACACCACTGAAGACTGCGGGATTGCTCTCGGTCAAGCTTTTAAAGAGGCCCTACTGGCGCGTGGAGTA
AGTGTGGTGACTTCTGACGCCCTAACGAGAGCCAGTTCGAAAATTTCTCCGGGATGACCGCGCACCTCAT
        430       440       450       460       470       480       490

                                                        BglII
AAAAGGTTTGGATCAGGATTTGCGCCTTTGGATGAGGCACTTTCCAGAGCGGTGGTAGATCTTTCGAACA
TTTTCCAAACCTAGTCCTAAACGCGGAAACCTACTCCGTGAAAGGTCTCGCCACCATCTAGAAAGCTTGT
        500       510       520       530       540       550       560

    BsiWI                                      BglII
GGCCGTACGCAGTTGTCGAACTTGGTTTGCAAAGGGAGAAAGTAGGAGATCTCTCTTGCGAGATGATCCC
CCGGCATGCGTCAACAGCTTGAACCAAACGTTTCCCTCTTTCATCCTCTAGAGAGAACGCTCTACTAGGG
        570       580       590       600       610       620       630

    HindIII
GCATTTTCTTGAAAGCTTTGCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGCGAGGCAAGAAT
CGTAAAAGAACTTTCGAAACGTCTCCGATCGTCTTAATGGGAGGTGCAACTAACAGACGCTCCGTTCTTA
        640       650       660       670       680       690       700
```

```
BclI
|
GATCATCACCGTAGTGAGAGTGCGTTCAAGGCTCTTGCGGTTGCCATAAGAGAAGCCACCTCGCCCAATG
CTAGTAGTGGCATCACTCTCACGCAAGTTCCGAGAACGCCAACGGTATTCTCTTCGGTGGAGCGGGTTAC
     710       720       730       740       750       760       770

     KpnI                                                    DraI      PstI
     |                                                       |         |
GTACCAACGATGTTCCCTCCACCAAAGGTGTTCTTATGTAGTGACACCGATTATTTAAAGCTGCAGCATA
CATGGTTGCTACAAGGGAGGTGGTTTCCACAAGAATACATCACTGTGGCTAATAAATTTCGACGTCGTAT
     780       790       800       810       820       830       840

CGATATATATACATGTGTATATATGTATACCTATGAATGTCAGTAAGTATGTATACGAACAGTATGATAC
GCTATATATATGTACACATATATACATATGGATACTTACAGTCATTCATACATATGCTTGTCATACTATG
     850       860       870       880       890       900       910

               NsiI
               |
TGAAGATGACAAGGTAATGCATCATTCTATACGTGTCATTCTGAACGAGGCGCGCTTTCCTTTTTTCTTT
ACTTCTACTGTTCCATTACGTAGTAAGATATGCACAGTAAGACTTGCTCCGCGCGAAAGGAAAAAGAAA
     920       930       940       950       960       970       980

          XhoI
          |
TTGCTTTTTCTTTTTTTTTCTCTTGAACTCGAGAAAAAAAATATAAAAGAGATGGAGGAACGGGAAAAAG
AACGAAAAAGAAAAAAAAAGAGAACTTGAGCTCTTTTTTTTTATATTTTCTCTACCTCCTTGCCCTTTTC
     990       1000      1010      1020      1030      1040      1050

TTAGTTGTGGTGATAGGTGGCAAGTGGTATTCCGTAAGAACAACAAGAAAAGCATTTCATATTATGGCTG
AATCAACACCACTATCCACCGTTCACCATAAGGCATTCTTGTTGTTCTTTCGTAAAGTATAATACCGAC
     1060      1070      1080      1090      1100      1110      1120

AACTGAGCGAACAAGTGCAAAATTTAAGCATCAACGACAACAACGAGAATGGTTATGTTCCTCCTCACTT
TTGACTCGCTTGTTCACGTTTTAAATTCGTAGTTGCTGTTGTTGCTCTTACCAATACAAGGAGGAGTGAA
     1130      1140      1150      1160      1170      1180      1190

AAGAGGAAAACCAAGAAGTGCCAGAAATAACATGAGCAACTACAATAACAACAACGGCGGCTACAACGGT
TTCTCCTTTTGGTTCTTCACGGTCTTTATTGTACTCGTTGATGTTATTGTTGTTGCCGCCGATGTTGCCA
     1200      1210      1220      1230      1240      1250      1260

GGCCGTGGCGGTCGGCAGCTTCTTTAGCAACAACCGTCGTGGTGGTTACGGCAACGGTGGTTTCTTCGGTG
CCGGCACCGCCACCGTCGAAGAAATCGTTGTTGGCAGCACCACCAATGCCGTTGCCACCAAAGAAGCCAC
     1270      1280      1290      1300      1310      1320      1330

               BglII                              ClaI
               |                                  |
GAAACAACGGTGGCAGCAGATCTAACGGCCGTTCTGGTGGTAGATGGATCGATGGCAAACATGTCCCAGC
CTTTGTTGCCACCGTCGTCTAGATTGCCGGCAAGACCACCATCTACCTAGCTACCGTTTGTACAGGGTCG
     1340      1350      1360      1370      1380      1390      1400
```

282

```
                                                   BamHI
                                                     |
TCCAAGAAACGAAAAGGCCGAGATCGCCATATTTGGTGTCCCCGAGGATCCTCTACGCCGGACGCATCGT
AGGTTCTTTGCTTTTCCGGCTCTAGCGGTATAAACCACAGGGGCTCCTAGGAGATGCGGCCTGCGTAGCA
     1410       1420      1430      1440      1450      1460      1470

            NarI                      NarI
              |                         |
GGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGAT
CCGGCCGTAGTGGCCGCGGTGTCCACGCCAACGACCGCGGATATAGCGGCTGTAGTGGCTACCCCTTCTA
     1480       1490      1500      1510      1520      1530      1540

            BspHI  Eco 47
              |      |
CGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGG
GCCCGAGCGGTGAAGCCCGAGTACTCGCGAACAAAGCCGCACCCATACCACCGTCCGGGGCACCGGCCCC
     1550       1560      1570      1580      1590      1600      1610

         NarI              SphI
           |                 |
GACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACT
CTGACAACCCGCGGTAGAGGAACGTACGTGGTAAGGAACGCCGCCGCCACGAGTTGCCGGAGTTGGATGA
     1620       1630      1640      1650      1660      1670      1680

                                           SalI
                                             |
ACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAAC
TGACCCGACGAAGGATTACGTCCTCAGCGTATTCCCTCTCGCAGCTGGCTACGGGAACTCTCGGAAGTTG
     1690       1700      1710      1720      1730      1740      1750

CCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCGGCACTTATGACTGTCTTCTTTATCA
GGTCAGTCGAGGAAGGCCACCCGCGCCCCGTACTGATAGCAGCGGCGTGAATACTGACAGAAGAAATAGT
     1760       1770      1780      1790      1800      1810      1820

                       Eco 47
                         |
TGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGC
ACGTTGAGCATCCTGTCCACGGCCGTCGCGAGACCCAGTAAAAGCCGCTCCTGGCGAAAGCGACCTCGCG
     1830       1840      1850      1860      1870      1880      1890

        GsuI
          |
GACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGT
CTGCTACTAGCCGGACAGCGAACGCCATAAGCCTTAGAACGTGCGGGAGCGAGTTCGGAAGCAGTGACCA
     1900       1910      1920      1930      1940      1950      1960

                                         EglI
                                           |
CCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACG
GGGCGGTGGTTTGCAAAGCCGCTCTTCGTCCGGTAATAGCGGCCGTACCGCCGGCTGCGCGACCCGATGC
     1970       1980      1990      2000      2010      2020      2030

        NruI
          |
TCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCAT
AGAACGACCGCAAGCGCTGCGCTCCGACCTACCGGAAGGGGTAATACTAAGAAGAGCGAAGGCCGCCGTA
     2040       2050      2060      2070      2080      2090      2100
```

283

```
                              EspMI
                               |
CGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGA
GCCCTACGGGCGCAACGTCCGGTACGACAGGTCCGTCCATCTACTGCTGGTAGTCCCTGTCGAAGTTCCT
    2110      2120      2130      2140      2150      2160      2170


TCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGCCT
AGCGAGCGCCGAGAATGGTCGGATTGAAGCTAGTGACCTGGCGACTAGCAGTGCCGCTAAATACGGCGGA
    2180      2190      2200      2210      2220      2230      2240

BglI                                        NarI
 |                                           |
CGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTT
GCCGCTCGTGTACCTTGCCCAACCGTACCTAACATCCGCGGCGGGATATGGAACAGACGGAGGGGCGCAA
    2250      2260      2270      2280      2290      2300      2310


GCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCA
CGCAGCGCCACGTACCTCGGCCCGGTGGAGCTGGACTTACCTTCGGCCGCCGTGGAGCGATTGCCTAAGT
    2320      2330      2340      2350      2360      2370      2380

                                                      FspI
                                                       |
CCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAA
GGTGAGGTTCTTAACCTCGGTTAGTTAAGAACGCCTCTTGACACTTACGCGTTTGGTTGGGAACCGTCTT
    2390      2400      2410      2420      2430      2440      2450

    GsuI
     |
CATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCA
GTATAGGTAGCGCAGGCGGTAGAGGTCGTCGGCGTGCGCCGCGTAGAGCCCGTCGCAACCCAGGACCGGT
    2460      2470      2480      2490      2500      2510      2520

    FspI
     |
CGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGC
GCCCACGCGTACTAGCACGAGGACAGCAACTCCTGGGCCGATCCGACCGCCCCAACGGAATGACCAATCG
    2530      2540      2550      2560      2570      2580      2590


AGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAAACGTCTGCGACCTGAGCAAC
TCTTACTTAGTGGCTATGCGCTCGCTTGCACTTCGCTGACGACGACGTTTTGCAGACGCTGGACTCGTTG
    2600      2610      2620      2630      2640      2650      2660


AACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAAACGCGGAAGTCAGCGCCCTGCACCATT
TTGTACTTACCAGAAGCCAAAGGCACAAAGCATTTCAGACCTTTGCGCCTTCAGTCGCGGGACGTGGTAA
    2670      2680      2690      2700      2710      2720      2730


ATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGC
TACAAGGCCTAGACGTAGCGTCCTACGACGACCGATGGGACACCTTGTGGATGTAGACATAATTGCTTCG
    2740      2750      2760      2770      2780      2790      2800
```

Eco 47
|
GCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACA
CGACCGTAACTGGGACTCACTAAAAGAGACCAGGGCGGCGTAGGTATGGCGGTCAACAAATGGGAGTGT
      2810   2820   2830   2840   2850   2860   2870

ACGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGT
TGCAAGGTCATTGGCCCGTACAAGTAGTAGTCATTGGGCATAGCACTCGTAGGAGAGAGCAAAGTAGCCA
      2880   2890   2900   2910   2920   2930   2940

ATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCC
TAGTAATGGGGGTACTTGTCTTTAGGGGGAATGTGCCTCCGTAGTCACTGGTTTGTCCTTTTTTGGCGGG
      2950   2960   2970   2980   2990   3000   3010

                                                               GsuI
                                                               |
TTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGA
AATTGTACCGGGCGAAATAGTCTTCGGTCTGTAATTGCGAAGACCTCTTTGAGTTGCTCGACCTGCGCCT
      3020   3030   3040   3050   3060   3070   3080

                XmnI   PvuII
                | |
TGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGT
ACTTGTCCGTCTGTAGACACTTAGCGAAGTGCTGGTGCGACTACTCGAAATGGCGTCGACGGAGCGCGCA
      3090   3100   3110   3120   3130   3140   3150

TTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGG
AAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCC
      3160   3170   3180   3190   3200   3210   3220

ATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGAC
TACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGCGTCGGTACTG
      3230   3240   3250   3260   3270   3280   3290

CCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAG
GGTCAGTGCATCGCTATCGCCTCACATATGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTC
      3300   3310   3320   3330   3340   3350   3360

     NdeI
     |
TGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGC
ACGTGGTATACGCCACACTTTATGGCGTGTCTACGCATTCCTCTTTTATGGCGTAGTCCGCGAGAAGGCG
      3370   3380   3390   3400   3410   3420   3430

TTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
AAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCCGC
      3440   3450   3460   3470   3480   3490   3500

GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG
CATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGTACACTCGTTTTCCGGTCGTTTTCC
      3510   3520   3530   3540   3550   3560   3570

```
CCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAA
GGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAACGTATCCGAGGCGGGGGGACTGCTCGTAGTGTT
     3580      3590      3600      3610      3620      3630      3640


AAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA
TTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCT
     3650      3660      3670      3680      3690      3700      3710


AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
TCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCC
     3720      3730      3740      3750      3760      3770      3780


GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCT
CTTCGCACCGCGAAAGAGTATCGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGA
     3790      3800      3810      3820      3830      3840      3850


GGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCC
CCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGG
     3860      3870      3880      3890      3900      3910      3920


AACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATG
TTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCATAC
     3930      3940      3950      3960      3970      3980      3990


TAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTAT
ATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATA
     4000      4010      4020      4030      4040      4050      4060


CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
GACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGG
     4070      4080      4090      4100      4110      4120      4130


GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATC
CGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAG
     4140      4150      4160      4170      4180      4190      4200


                                                                  BspHI
                                                                  |
CTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG
GAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTC
     4210      4220      4230      4240      4250      4260      4270


                            DraI                      DraI
                            |                         |
ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATA
TAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATAT
     4280      4290      4300      4310      4320      4330      4340
```

286

```
TATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT
ATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATA
     4350      4360      4370      4380      4390      4400      4410


TTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGG
AAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACC
     4420      4430      4440      4450      4460      4470      4480


                         GsuI
                          |
CCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGT
     4490      4500      4510      4520      4530      4540      4550


     EglI
      |
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCC
CGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTAACAACGG
     4560      4570      4580      4590      4600      4610      4620


                                  FspI                        PstI
                                   |                           |
GGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCGT
CCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGACGTCCGTAGCA
     4630      4640      4650      4660      4670      4680      4690


GGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGA
CCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACT
     4700      4710      4720      4730      4740      4750      4760


                                     PvuI
                                      |
TCCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCG
AGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGC
     4770      4780      4790      4800      4810      4820      4830


CAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTT
GTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAA
     4840      4850      4860      4870      4880      4890      4900


     ScaI
      |
TTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC
AAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACG
     4910      4920      4930      4940      4950      4960      4970


                                   DraI                        XmnI
                                    |                           |
CCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTT
GGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAA
     4980      4990      5000      5010      5020      5030      5040
```

287

```
CTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACC
GAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGG
         5050      5060      5070      5080      5090      5100      5110
```

```
CAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCC
GTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGG
         5120      5130      5140      5150      5160      5170      5180
```

```
GCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAA
CGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTT
         5190      5200      5210      5220      5230      5240      5250
```

```
                            BspHI
                              |
GCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGG
CGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCC
         5260      5270      5280      5290      5300      5310      5320
```

```
                                          AatII                BspHI
                                            |                    |
GGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCAT TATTATCATGACATTAACC
CCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGG
         5330      5340      5350      5360      5370      5380      5390
```

```
TA TAAAAAATAGGCGTATCATGAGCCCCTTTCGTCTTCAA
ATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGAAGTT
         5400      5410      5420      5430
```

FIGURE 44. Sequence of plasmid pHISi-I.

288
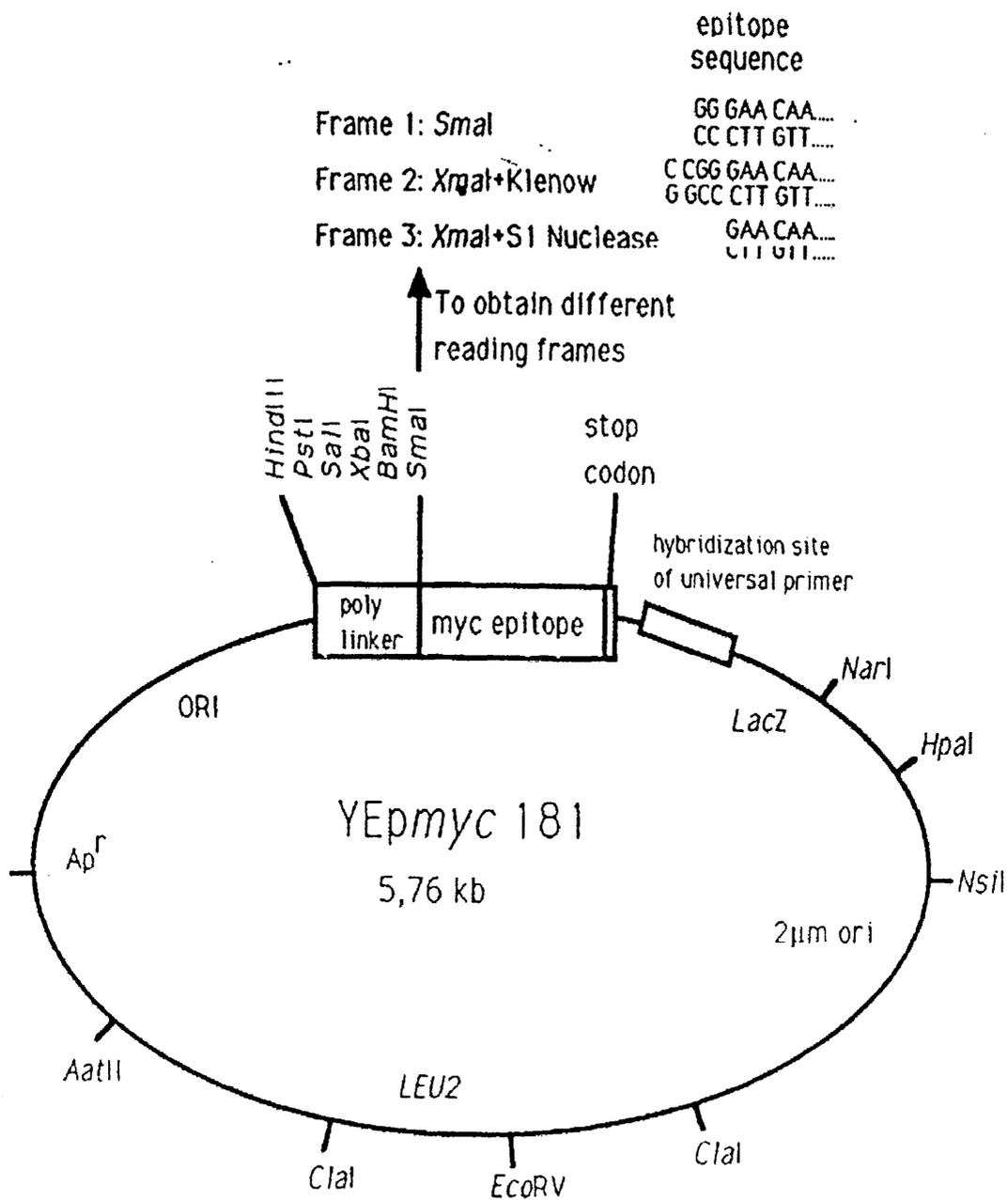
FIGURE 45. Map of autonomously replicating plasmid YEp*myc*181.

```
   1      gcttgcatgc  aacttctttt  cttttttttt  cttttctctc
  41      tcccccgttg  ttgtctcacc  atatccgcaa  tgacaaaaaa
  81      aatgatggaa  gacactaaag  gaaaaaatta  acgacaaaga
 121      cagcaccaac  agatgtcgtt  gttccagagc  tgatgagggg
 161      tatcttcgaa  cacacgaaac  tttttccttc  cttcattcac
 201      gcacactact  ctctaatgag  caacggtata  cggccttcct
 241      tccagttact  tgaatttgaa  ataaaaaag   tttgccgctt
 281      tgctatcaag  tataaataga  cctgcaatta  ttaatctttt
 321      gtttcctcgt  cattgttctc  gttccctttc  ttccttgttt
 361      cttttctgc   acaatatttc  aagctatacc  aagcatacaa
 401      tcaactccaa  gctttgcaaa  gatggataaa  gcggaattaa
 441      ttcccgagcc  tccaaaaaag  aagagaaagg  tcgaattggg
 481      taccgccgcc  aattttaatc  aaagtgggaa  tattgctgat
 521      agctcattgt  ccttcacttt  cactaacagt  agcaacggtc
 561      cgaacctcat  aacaactcaa  acaaattctc  aagcgctttc
 601      acaaccaatt  gcctcctcta  acgttcatga  taacttcatg
 641      aataatgaaa  tcacggctag  taaaattgat  gatggtaata
 681      attcaaaacc  actgtcacct  ggttggacgg  accaaactgc
 721      gtataacgcg  tttggaatca  ctacagggat  gtttaatacc
 761      actacaatgg  atgatgtata  taactatcta  ttcgatgatg
 801      aagatacccc  accaaaccca  aaaaaagaga  tctctcgagg
 841      atccgaattc  cagatctatg  aatcgtagat  actgaaaaac
 881      cccgcaagtt  cacttcaact  gtgcatcgtg  caccatctca
 921      atttctttca  tttatacatc  gttttgcctt  cttttatgta
 961      actatactcc  tctaagtttc  aatcttggcc  atgtaacctc
1001      tgatctatag  aatttttaa   atgactagaa  ttaatgccca
1041      tctttttttt  ggacctaaat  tcttcatgaa  aatatattac
1081      gagggcttat  tcagaagctt  tggacttctt  cgccagaggt
1121      ttggtcaagt  ctccaatcaa  ggttgtcggc  ttgtctacct
1161      tgccagaaat  ttacgaaaag  atggaaaagg  gtcaaatcgt
1201      tggtagatac  gttgttgaca  cttctaaata  agcgaatttc
1241      ttatgattta  tgatttttat  tattaaataa  gttataaaaa
1281      aaataagtgt  atacaaattt  taaagtgact  cttaggtttt
1321      aaaacgaaaa  ttcttgttct  tgagtaactc  tttcctgtag
1361      gtcaggttgc  tttctcaggt  atagcatgag  gtcgctctta
1401      ttgaccacac  ctctaccggc  atgcccgaaa  ttcccctacc
1441      ctatgaacat  attccatttt  gtaatttcgt  gtcgtttcta
1481      ttatgaattt  catttataaa  gtttatgtac  aaatatcata
1521      aaaaaagaga  atctttttaa  gcaaggattt  tcttaacttc
1561      ttcggcgaca  gcatcaccga  cttcggtggt  actgttggaa
1601      ccacctaaat  caccagttct  gatacctgca  tccaaaacct
1641      ttttaactgc  atcttcaatg  gccttacctt  cttcaggcaa
1681      gttcaatgac  aatttcaaca  tcattgcagc  agacaagata
1721      gtggcgatag  ggtcaacctt  attctttggc  aaatctggag
1761      cagaaccgtg  gcatggttcg  tacaaaccaa  atgcggtgtt
1801      cttgtctggc  aaagaggcca  aggacgcaga  tggcaacaaa
```

290

```
1841    cccaaggaac ctgggataac ggaggcttca tcggagatga
1881    tatcaccaaa catgttgctg gtgattataa taccatttag
1921    gtgggttggg ttcttaacta ggatcatggc ggcagaatca
1961    atcaattgat gttgaacctt caatgtagga aattcgttct
2001    tgatggtttc ctccacagtt tttctccata atcttgaaga
2041    ggccaaaaca ttagctttat ccaaggacca aataggcaat
2081    ggtggctcat gttgtagggc catgaaagcg gccattcttg
2121    tgattctttg cacttctgga acggtgtatt gttcactatc
2161    ccaagcgaca ccatcaccat cgtcttcctt tctcttacca
2201    aagtaaatac ctcccactaa ttctctgaca acaacgaagt
2241    cagtaccttt agcaaattgt ggcttgattg gagataagtc
2281    taaaagagag tcggatgcaa agttacatgg tcttaagttg
2321    gcgtacaatt gaagttcttt acggattttt agtaaacctt
2361    gttcaggtct aacactacct gtaccccatt taggaccacc
2401    cacagcacct aacaaaacgg catcaacctt cttggaggct
2441    tccagcgcct catctggaag tgggacacct gtagcatcga
2481    tagcagcacc accaattaaa tgattttcga aatcgaactt
2521    gacattggaa cgaacatcag aaatagcttt aagaacctta
2561    atggcttcgg ctgtgatttc ttgaccaacg tggtcacctg
2601    gcaaaacgac gatcttctta ggggcagaca ttagaatggt
2641    atatccttga aatatatata tatattgctg aaatgtaaaa
2681    ggtaagaaaa gttagaaagt aagacgattg ctaaccacct
2721    attggaaaaa acaataggtc cttaaataat attgtcaact
2761    tcaagtattg tgatgcaagc atttagtcat gaacgcttct
2801    ctattctata tgaaaagccg gttccggcct ctcacctttc
2841    cttttctcc caatttttca gttgaaaaag gtatatgcgt
2881    caggcgacct ctgaaattaa caaaaaattt ccagtcatcg
2921    aatttgattc tgtgcgatag cgcccctgtg tgttctcgtt
2861    atgttgagga aaaaaataat ggttgctaag agattcgaac
3001    tcttgcatct tacgatacct gagtattccc acagttgggg
3041    atctcgactc tagctagagg atcaattcgt aatcatggtc
3081    atagctgttt cctgtgtgaa attgttatcc gctcacaatt
3121    ccacacaaca tacgagccgg aagcataaag tgtaaagcct
3161    ggggtgccta atgagtgagg taactcacat taattgcgtt
3201    gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc
3241    cagctggatt aatgaatcgg ccaacgcgcg gggagaggcg
3281    gtttgcgtat tgggcgctct tccgcttcct cgctcactga
3321    ctcgctgcgc tcggtcgttc ggctgcggcg agcggtatca
3361    gctcactcaa aggcggtaat acggttatcc acagaatcag
3401    gggataacgc aggaaagaac atgtgagcaa aaggccagca
3441    aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt
3481    ttccataggc tccgcccccc tgacgagcat cacaaaaatc
3521    gacgctcaag tcagaggtgg cgaaacccga caggactata
3561    aagataccag gcgtttcccc ctggaagctc cctcgtgcgc
3601    tctcctgttc cgaccctgcc gcttaccgga tacctgtccg
3641    cctttctccc ttcgggaagc gtggcgcttt ctcatagctc
3681    acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc
3721    aagctgggct gtgtgcacga accccccgtt cagcccgacc
3761    gctgcgcctt atccggtaac tatcgtcttg agtccaaccc
```

291

```
3801   ggtaagacac gacttatcgc cactggcagc agccactggt
3841   aacaggatta gcagagcgag gtatgtaggc ggtgctacag
3881   agttcttgaa gtggtggcct aactacggct acactagaag
3921   gacagtattt ggtatctgcg ctctgctgaa gccagttacc
3961   ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa
4001   ccaccgctgg tagcggtggt ttttttgttt gcaagcagca
4041   gattacgcgc agaaaaaaag gatctcaaga agatcctttg
4081   atcttttcta cggggtctga cgctcagtgg aacgaaaact
4121   cacgttaagg gattttggtc atgagattat caaaaaggat
4161   cttcacctag atcctttaa attaaaaatg aagttttaaa
4201   tcaatctaaa gtatatatga gtaaacttgg tctgacagtt
4241   accaatgctt aatcagtgag gcacctatct cagcgatctg
4281   tctatttcgt tcatccatag ttgcctgact ccccgtcgtg
4321   tagataacta cgatacggga gggcttacca tctggcccca
4361   gtgctgcaat gataccgcga gacccacgct caccggctcc
4401   agatttatca gcaataaacc agccagccgg aagggccgag
4441   cgcagaagtg gtcctgcaac tttatccgcc tccatccagt
4481   ctattaattg ttgccgggaa gctagagtaa gtagttcgcc
4521   agttaatagt ttgcgcaacg ttgttgccat tgctacaggc
4561   atcgtggtgt cacgctcgtc gtttggtatg gcttcattca
4601   gctccggttc ccaacgatca aggcgagtta catgatcccc
4641   catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg
4681   atcgttgtca gaagtaagtt ggccgcagtg ttatcactca
4721   tggttatggc agcactgcat aattctctta ctgtcatgcc
4761   atccgtaaga tgcttttctg tgactggtga gtactcaacc
4801   aagtcattct gagaatagtg tatgcggcga ccgagttgct
4841   cttgcccggc gtcaatacgg gataataccg cgccacatag
4881   cagaacttta aaagtgctca tcattggaaa acgttcttcg
4921   gggcgaaaac tctcaaggat cttaccgctg ttgagatcca
4961   gttcgatgta acccactcgt gcacccaact gatcttcagc
5001   atcttttact ttcaccagcg tttctgggtg agcaaaaaca
5041   ggaaggcaaa atgccgcaaa aaagggaata agggcgacac
5081   ggaaatgttg aatactcata ctcttccttt ttcaatatta
5121   ttgaagcatt tatcagggtt attgtctcat gagcggatac
5161   atatttgaat gtatttagaa aaataaacaa ataggggttc
5201   cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga
5241   aaccattatt atcatgacat taacctataa aaataggcgt
5281   atcacgaggc cctttcgtct cgcgcgtttc ggtgatgacg
5321   gtgaaaacct ctgacacatg cagctcccgg agacggtcac
5361   agcttgtctg taagcggatg ccgggagcag acaagcccgt
5401   cagggcgcgt cagcgggtgt tggcgggtgt cggggctggc
5441   ttaactatgc ggcatcagag cagattgtac tgagagtgca
5481   ccataacgca tttaagcata aacacgcact atgccgttct
5521   tctcatgtat atatatatac aggcaacacg cagatatagg
5561   tgcgacgtga acagtgagct gtatgtgcgc agctcgcgtt
5601   gcattttcgg aagcgctcgt tttcggaaac gctttgaagt
5641   tcctattccg aagttcctat tctctagcta gaaagtatag
5681   gaacttcaga gcgcttttga aaaccaaaag cgctctgaag
5721   acgcactttc aaaaaaccaa aaacgcaccg gactgtaacg
```

292

```
5761    agctactaaa atattgcgaa taccgcttcc acaaacattg
5801    ctcaaaagta tctctttgct atatatctct gtgctatatc
5841    cctatataac ctacccatcc acctttcgct ccttgaactt
5881    gcatctaaac tcgacctcta cattttttat gtttatctct
5921    agtattactc tttagacaaa aaaattgtag taagaactat
5961    tcatagagtg aatcgaaaac aatacgaaaa tgtaaacatt
6001    tcctatacgt agtatataga gacaaaatag aagaaaccgt
6041    tcataatttt ctgaccaatg aagaatcatc aacgctatca
6081    ctttctgttc acaaagtatg cgcaatccac atcggtatag
6121    aatataatcg gggatgcctt tatcttgaaa aaatgcaccc
6161    gcagcttcgc tagtaatcag taaacgcggg aagtggagtc
6201    aggctttttt tatggaagag aaaatagaca ccaaagtagc
6241    cttcttctaa ccttaacgga cctacagtgc aaaaagttat
6281    caagagactg cattatagag cgcacaaagg agaaaaaaag
6321    taatctaaga tgctttgtta gaaaaatagc gctctcggga
6361    tgcatttttg tagaacaaaa aagaagtata gattctttgt
6401    tggtaaaata gcgctctcgc gttgcatttc tgttctgtaa
6441    aaatgcagct cagattcttt gtttgaaaaa ttagcgctct
6481    cgcgttgcat ttttgtttta caaaaatgaa gcacagattc
6521    ttcgttggta aaatagcgct ttcgcgttgc atttctgttc
6561    tgtaaaaatg cagctcagat tctttgtttg aaaaattagc
6601    gctctcgcgt tgcatttttg ttctacaaaa tgaagcacag
6641    atgcttcgtt
```

KEY:

Blue   Locations of the AD10[S] and complement of AD10[R] PCR and sequencing primers.

Red   EcoR1 restriction endonuclease site used for subcloning the rat brain library cDNAs.

FIGURE 46. Sequence of the cDNA/AD fusion library plasmid pGAD10.

293

APPENDIX VII.

In order to determine the volume of nuclear extract used in the electrophoretic mobility shift assay and competition binding assay, to give a protein concentration of 10 µg per reaction, it was first necessary to determine the concentration of protein in the nuclear extracts. To achieve this a microtitre plate method utilising bicinchoninic acid was used and bovine serum albumin (0 to 10 µg/µl) served as the protein standard. The BCA reagent was mixed with the protein dilution or nuclear extract sample, incubated at room temperature and read on a microplate reader at 562 nm. The data for the preparation of a standard curve to determine the protein concentration (µg/µl) of the nuclear extracts is shown in table 5 and the standard curve is shown in Figure 47.

From cultures containing approximately 1×10^6 cells, 200 µl of nuclear extract were prepared producing a protein yield of between 1.0 and 3.0 µg/µl depending on the source of the nuclear extract (as shown in table 6)

Concentration (µg/µl)	Absorbance at 562 nm					
	Data set 1	Data set 2	Data set 3	Data set 4	Mean Absorbance	Corrected Absorbance
Reagent Blank	0.05	0.06	0.05	0.06	0.05	
0 µg/µl	0.05	0.06	0.05	0.06	0.05	0.00
1 µg/µl	0.29	0.28	0.27	0.27	0.28	0.23
2 µg/µl	0.50	0.48	0.46	0.47	0.48	0.43
3 µg/µl	0.65	0.64	0.57	0.61	0.62	0.57
4 µg/µl	0.81	0.81	0.80	0.78	0.80	0.75
5 µg/µl	0.98	0.99	1.00	0.97	0.99	0.94
6 µg/µl	1.15	1.20	1.16	1.19	1.17	1.12
7 µg/µl	1.36	1.33	1.39	1.39	1.36	1.31
8 µg/µl	1.53	1.50	1.51	1.50	1.51	1.46
9 µg/µl	1.80	1.79	1.70	1.68	1.74	1.69
10 µg/µl	2.03	1.93	1.81	1.81	1.90	1.85

Table 5.   Data obtained for the preparation of a standard curve to determine the protein concentration (µg/µl) of the nuclear extracts used in the EMSA and competition binding assays.
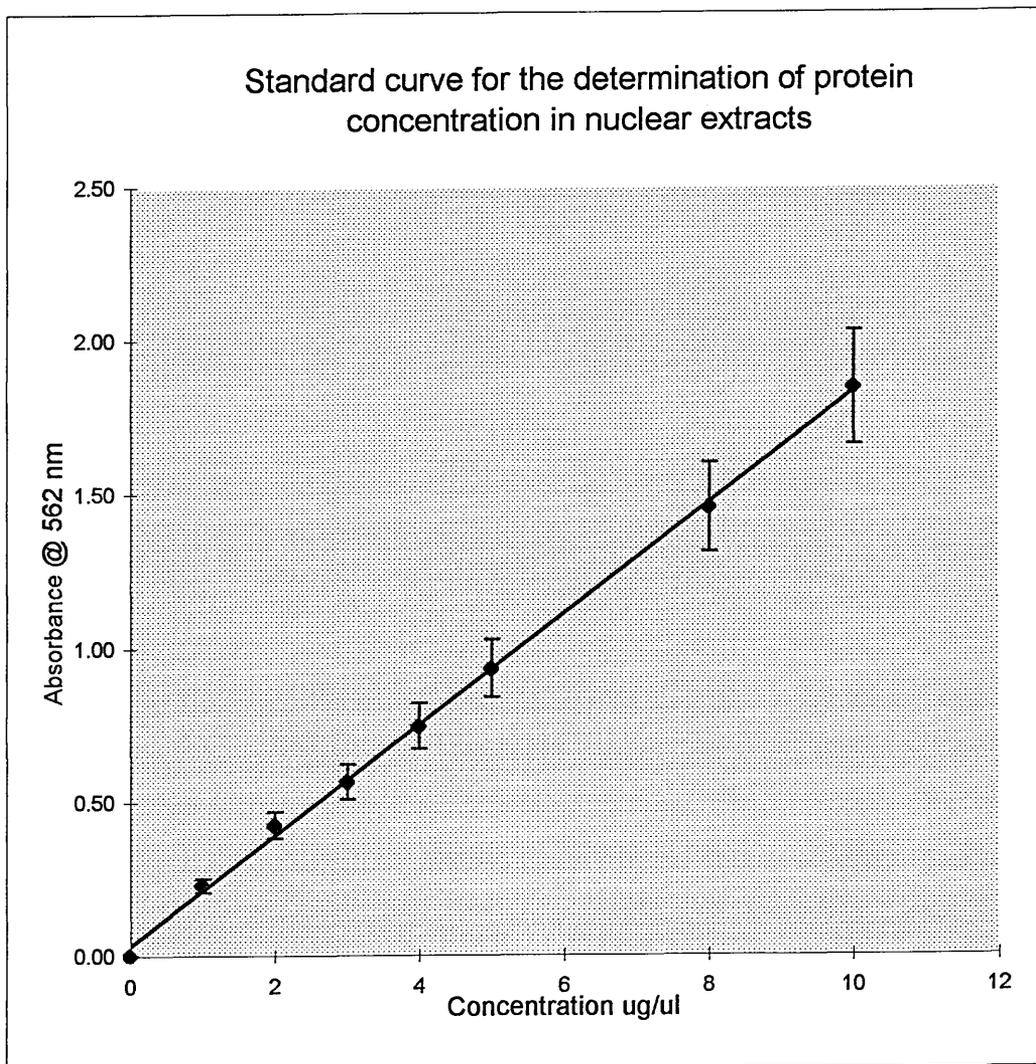
Figure 47. Standard curve to determine the concentration of protein ($\mu g/\mu l$) in nuclear extracts.

Sample	Absorbance at 562 nm				
	Data set 1	Data set 2	Mean Absorbance	Corrected Absorbance	Calculated Concentration (μg/μl)
Reagent Blank	0.05	0.06	0.06		
C$_6$ glioma	0.25	0.26	0.26	0.20	1.1 μg/μl
HeLa	0.24	0.23	0.24	0.18	1.0 μg/μl
RBX	0.61	0.64	0.62	0.56	3.0 μg/μl
33B oligodendrocytoma	0.32	0.29	0.31	0.24	1.3 μg/μl

Table 6. The concentration (μg/μl) of protein in the nuclear extracts as determined from the graph shown in figure 47.