Characterisation and genetic mapping of genes with potential relevance to neurodegenerative disease

by

Andrew Smyth McCallion BSc. (Hons)

Thesis presented to the University of Glasgow, for the degree of PhD, in the Faculty of Science

November, 1997

Institute of Biomedical and Life Sciences,
Division of Molecular Genetics,
Robertson Laboratory of Biotechnology

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

Acknowledgements

I wish to take this opportunity to thank those individuals who have helped and supported me throughout the duration my PhD project.

I must begin by thanking Professor R. Wayne Davies; for giving me the opportunity to perform this work under his supervision, for his confidence in my abilities, and for his constant demand for disciplined thought and high standards. I wish to thank everyone within "*Wayne's World*" for their help, advice and laughter, especially during my extended periods of work in Paris. In particular, I wish to thank Greg Stewart for his friendship and tireless work in the EST and *Mobp* sequencing projects. I also wish to acknowledge the great help and support provided by Lynn and Irene.

I wish to thank Dr Jean Louis Guénet (Institut Pasteur) for seeing the potential in our collaboration, for accepting me as a Stagière within his laboratory (1995-1997), for improving my spoken French and for acting as my mentor during my time at l'Institut. Pasteur. Thanks are also due to all within Génétique des Mammifères. In particular, I wish to thank Christophe for the hours of discussion, shared ideas and many cups of hot chocolate; Alexis, Anna and Patricia for their friendship and Ken for demonstrating that Irishmen get everywhere.

Thanks are due to Dr T.A. Glencorse and Professor I.R. Griffiths with whom I also established collaborative studies and who provided advice, support and the opportunity to perform publishable research.

I wish to thank dad, mum, Julie and the rest of the family for their tireless support and prayers throughout my years as a student and beyond. Thanks also to Peter & Carolyn, Iain & Kirstine and John & Lisa for their prayers and friendship.

Finally, I wish to thank Gill for her constant belief in me, calming words, love and patience during the preparation and writing of this thesis. My gem, you have shown me how good it feels to smile.

Sole deo gloria

Abstract

Section 1

I have established a PCR based single strand conformation polymorphism (SSCP) assay for the detection of polymorphisms in PCR amplified genomic fragments. I have used this assay in the analysis of PCR products up to 300 bp in length from putative 3' untranslated regions (3'UTR) of mouse expressed sequence tags (ESTs) and show that an SSCP difference allowing mapping of ESTs on *Mus spretus* × C57BL/6 crosses is found at >90%. These studies report the mapping of 46 new mouse brain expressed sequence tags (ESTs) and one anonymous sequence tag to localised regions (≥ 2 cM) of the mouse genetic map. Assignment of these ESTs to chromosomes was carried out by SSCP analysis of sets of typed DNAs from animals in several backcross generations of a breeding programme that generates consomic lines with particular chromosomes deriving from *M.spretus* (SEG/Pas) on a C57BL/6 background. This allowed specific sets of recombinants to be selected from the European Collaborative Interspecific Backcross (EUCIB) resource for precise mapping. These experiments establish an effective strategy for the genetic mapping of multiple mouse ESTs. The particular EST set which I am mapping is enriched for novel and locally expressed genes of the substantia nigra, and may prove to be a good source of candidate genes for loci involved with neurological disorders. These studies also establish a qualitative RT-PCR screen evaluating the presence or absence of an RNA species in a known tissue at a specific timepoint. This work constitutes a significant step towards the goal of establishing an integrated gene map, combining information on the position of a gene, its sequence and its expression pattern.

Abstract

Section 2

A new mouse oligodendrocyte-specific gene was discovered in a differential screen, and subsequently found to be the mouse homologue of the recently discovered rat gene encoding the myelin-associated oligodendrocytic basic proteins (MOBP); a family of related highly basic myelin protein isoforms. MOBP is a new major component of CNS myelin. I have used SSCP analysis of a radioactively labelled PCR product from the 3'UTR of this gene in an assay to allow rapid genetic mapping. I report that *Mobp* is localised between *D9Mit55* and *D9Mit19* at 63.6 ± 1cM distal to the centromere of chromosome nine. Three mouse neurological mutations, *spinner, tippy* and *ducky*, map in this region of the mouse genome. The syntenic region of the human genome is 3p21-22. I have cloned the mouse gene *Mobp* in an endeavour to elucidate its genomic organisation and regulation. These studies demonstrate that the *Mobp* gene is a complex transcriptional unit. It comprises at eight discrete exons encompassing a genomic region in excess of 14 kb. *Mobp* also contains alternative internal splice donor/acceptor sites within at least two of its exons. I have utilised an oligo-capping assay in an endeavour to identify the transcription start points (tsp) associated with the *Mobp* gene. These studies have also revealed the presence of a novel transcriptional unit within the *Mobp* gene. In excess of 10 kb of sequence has been generated across the genomic region encompassing *Mobp*. Finally I have generated a gene targeting construct, designed to disrupt the *Mobp* gene, in an endeavour to begin to elucidate the function of this novel gene. These studies provide the platform for a more detailed analysis of the function of the *Mobp* gene product and subsequent evaluation of its possible involvement in known neuropathies.

Figure	Table of figures	Page Number

List of abbreviations

ASO	Allele Specific Oligonucleotides
BAP	Bacterial Alkaline Phosphatase
cDNA	Complementary Deoxyribonucleic Acid
CMT	Charcot-Marie-Tooth
CNS	Central Nervous System
DCK	Deoxycytidine Kinase
DGGE	Denaturing Gradient Gel Electrophoresis
DNA	Deoxyribonucleic Acid
DS	Double stranded (DNA)
ER	Endoplasmic Reticulum
ES	Embryonic Stem
EST	Expressed Sequence Tag
EUCIB	European Collaborative Interspecific Backcross
Ftg	Fibre-tract gene
GFAP	Glial Fibrillary Acidic Protein
hET-BR	Human Endothelin Receptor type B
HNPP	Hereditary Neuropathy with liability to Pressure Palsies
Ig	Immunoglobulin
IPL	Intraperiod Line
ISH	*in situ* hybridisation
KO	Knockout
MAG	Myelin Associated Glycoprotein
MBP	Myelin Basic Protein
MDL	Major Dense Line
MOBP	Myelin-associated Oligodendrocytic Basic Protein
MOG	Myelin Oligodendrocyte Glycoprotein
N-CAM	Neural-Cell Adhesion Molecule
OMGP	Oligodendrocyte Myelin Glycoprotein
PCR	Polymerase Chain Reaction
PNS	Peripheral Nervous System
P0	Protein Zero
PLP	Proteolipid Protein
PMD	Pelizaeus Merzbacher Disease

List of abbreviations continued

PMP22	Peripheral Myelin Protein 22
RNA	Ribonucleic Acid
RT	Reverse Transcriptase
RFLP	Restriction Fragment Length Polymorphism
SSCP	Single Strand Conformation Polymorphism
SN	substantia nigra
TAP	Tobacco Acid Pyrophosphatase
tsp	Transcription Start Points
3'UTR	3' Untranslated Region
XSP	X-Linked Spastic Paraplegia
YAC	Yeast Artificial Chromosome

Chapter1 Introduction

Section 1

1.1 Mapping expressed genes

1.1.1 Relevance of gene mapping to the Human Genome

High resolution expression maps of the mouse and human genomes are among the most immediate needs of the human genome project (Gardiner and Brennan, 1992; Hochgeschwander, 1992). It is widely accepted that mapping genes in higher organisms e.g. man or mouse, will provide information permitting the elucidation of the higher levels of genome organisation (Polymeropoulos *et al.*, 1992). This approach permits the generation of data at a number of levels, ranging from the localisation of human genes and their rodent orthologues to the characterisation and classification of genes according to function and expression. Such a strategy will yield information critical to the understanding of gene evolution (Polymeropoulos *et al.*, 1992; Polymeropoulos *et al.*, 1993).

One goal of the Human Genome Project is the ability to integrate information on the genetic and physical position of every gene in the genome with information regarding their temporal and spatial expression (Gardiner and Brennan, 1992). This form of integrated map will permit elucidation of the organisation of mammalian genomes, particularly gene clustering, gene number and the overlap, in sets of expressed sequences, between tissues and developmental stages (Levitt, 1991a). The ultimate goal is to subsequently obtain structural and functional information for each gene in our genome (Strachan *et al.*, 1997). The generation and genetic mapping of large numbers of expressed sequence tags (ESTs) is now accepted as a technically feasible way to work towards these important goals (Stewart, 1995). ESTs are single-pass sequences from the end of cDNA clones, usually from tissue-specific directional libraries (Adams *et al.*, 1991; Hoog, 1991; Wilcox *et al.*, 1991; Adams *et al.*, 1992; Okubo, 1992).

1.1.2 Selection of the mouse as the species of choice for these studies

Analyses of spatial and temporal expression are more readily performed in the mouse (Hochgeschwander, 1992). The mouse presents much fewer of the ethical problems, encountered in man, to an investigator wishing to obtain embryonic tissues. For example, there are fewer restrictions on gaining access to specific tissues e.g. brain and also for

making comparative studies of this tissue at different developmental stages when generating specific cDNA libraries, and subsequently ESTs. Figure 1.1 illustrates the steps taken, in this study, from isolation of tissue from a particular developmental stage, through the generation of ESTs to their eventual localisation within the mouse genome. Such an approach permits the identification of gene candidates for involvement with previously identified pathologies (Levitt, 1991; Wilcox *et al.*, 1991). The utility of gene/EST mapping, to this end, is discussed in greater detail in Chapter 1, 1.2.

The mouse is the species of choice for the performance of functional assays of mammalian genes (Guénet, 1996). The development of transgenic and gene targeting technologies in the mouse, due partially to the high degree of evolutionary conservation between human and mouse genes (Guénet, 1996), has provided a useful tool for the Human Genome Project. Consequently, in order to utilise these technologies, in the mouse, any gene identified in man must also be identified in and subsequently cloned from the mouse genome (Guénet, 1996). It is noteworthy that the ESTs used in this study derive from subtracted cDNA libraries enriched for genes showing low level expression or localised expression in the mouse ventral midbrain (Savioz and Davies, 1995). It is the dopaminergic neurons of the substantia nigra pars compacta within the ventral midbrain that is specifically degenerated in the pathology of Parkinson's disease. A number of linkage groups, conserved between the mouse and human genomes, have already been identified. Elucidation of the full extent of these syntenic regions requires continuous refinement of their respective genetic and physical maps (Guénet, 1996). This will further aid the exploitation of the mouse for the functional analysis of genes identified in either species.

As mentioned above the generation and genetic mapping of large numbers of expressed sequence tags (ESTs) is a technically feasible way to work towards obtaining expression maps and furthering the Human Genome Project (Stewart, 1995). However, the generation of mouse ESTs has progressed more slowly than human EST sequencing, and the development of a simple, rapid and efficient approach to mapping ESTs remains of particular importance.

Until recently, mapping human cDNAs was only useful in generating low resolution physical maps (Polymerpoulos, 1992). The use of human-mouse somatic cell hybrids with a limited complement of human chromosomes, that have been fragmented by large doses of radiation, have increased thae rate at which genes and other markers can be localised within the human genome (Guénet, 1996). These refined hybrid panels have not been available for mapping markers in the mouse genome. One reason for this may be the

rarity of deletions and translocations within somatic cell hybrids for the mouse (Guénet, 1996). However, it is widely accepted that genetic mapping in the mouse is a very powerful tool (Avner *et al.*, 1988; Copeland and Jenkins, 1991; Takahashi and Ko, 1993). Any EST marker demonstrating a polymorphism between two mouse strains can be localised, and its position in the linear order of markers in that region can be readily determined. This may be established by the genetic typing of backcross progeny between the strains in question (Takahashi and Ko, 1993).

1.1.3 Selection of sequence from 3' untranslated regions for use in polymorphism assays

Coding sequences are more frequently conserved during evolution than non-coding sequences (Levitt, 1991; Wilcox *et al.*, 1991; Polymeropoulos *et al.*, 1992, 1993; Takahashi and Ko, 1993), and as a consequence gene markers are more productively sought in untranslated sequences (Levitt, 1991b; Wilcox *et al.*, 1991; Polymeropoulos *et al.*, 1992; Beier, 1993; Polymeropoulos *et al.*, 1993; Takahashi and Ko, 1993; Ko *et al.*, 1994; Beier *et al.*, 1995). Polymorphic variation occurs in non-coding (5' untranslated region, 5'UTR; 3' untranslated region, 3'UTR; intronic) sequences as frequently as 1 in 100 bp, between individuals (Antonarakis *et al.*, 1982). These polymorphisims are suggested (Levitt, 1991) to be readily detectable using PCR amplification in combination with established electrophoretic techniques (Orita *et al.*, 1989b; Sheffield *et al.*, 1989; White *et al.*, 1992).

3' untranslated sequence has been the most commonly used in the generation of gene markers for the human and mouse genomes (Levitt, 1991b; Wilcox *et al.*, 1991; Polymeropoulos *et al.*, 1992; Beier, 1993; Polymeropoulos *et al.*, 1993; Takahashi and Ko, 1993; Ko *et al.*, 1994; Beier *et al.*, 1995; McCallion *et al.*, 1996). With respect to mapping expressed sequences in the mouse, the use of sequence derived from 3'UTRs provides a number of well documented advantages. These include:

- The significant sequence variation between the mouse strains most commonly used in genetic mapping (Beier *et al.*, 1992; Beier, 1993; Takahashi and Ko, 1993).
- The availability of sequence that is virtually free of introns (Hawkins, 1988) and the consequent ability to predict the fragment size of amplified sequence in genomic DNA.
- The use of single-pass sequence information for the generation of oligonucleotides for PCR amplification and the subsequent generation of polymorphic gene markers (Levitt, 1991b; Podsulo *et al.*, 1991).

- The use of 3' UTR provides a consistent point of reference for mapping large genes e.g. the gene encoding dystrophin.
- 3' UTR sequences are frequently unique even among family members for example the gene family encoding actin proteins (Ko, 1990). Thus it permits the generation of gene-specific sequence-tagged sites (STS).

It is noteworthy that the use of 3' directed cDNA libraries makes 3' UTR sequence directly available.
Thus, the mapping of ESTs automatically generates STSs which can serve to aid the construction of physical maps of chromosomes e.g. in the establishment of cosmid, BAC and YAC (Burke *et al.*, 1987; McCormick *et al.*, 1989) contigs. The generation of sequence information around anonymous DNA markers, for the construction of a physical map of chromosomes, is both laborious and expensive (Hochgeschwander, 1992). This makes the mapping of ESTs particularly attractive.

1.1.4 The purpose of establishing genetic and physical maps of expressed genes

The ultimate goal of a gene map is to assign a genetic and physical map position to all genes, and a tissue-expression profile for all genes, in the mouse genome (Takahashi and Ko, 1993). However, a more immediate goal is to obtain a 1 cM resolution linkage map. YACs (Burke *et al.*, 1987; McCormick *et al.*, 1989) contain up to 1 Mb of genomic DNA. Hence an expression map, with genes positioned at a 1 cM resolution (approximately 1.6 Mb), may be used to order, and demonstrate overlaps among, YAC clones. A study by Copeland and Jenkins (1991) demonstrated that to obtain a 1 cM resolution linkage map at the 99% confidence interval, would require 8,000 cDNAs to be mapped on 500 backcross progeny. These calculations rely upon the assumption that genes are distributed randomly throughout the genome.

EST mapping may be performed at different levels ranging from chromosome assignment (50-250 Mb resolution) to the identification of a DNA clone or contig (< 1 Mb resolution) containing the sequence in question (Stewart, 1995). The establishment of a complete gene map requires the integration of information from different approaches. The result of combining sequence, mapping and expression data will prove to be useful in investigating biological phenomena and in identifying and cloning gene candidates for genetic disorders (Wilcox *et al.*, 1991; Polymeropoulos *et al.*, 1992, 1993; Takahashi and Ko, 1993).

1.2 Expression maps: Correlation with disease

Most mutations in genes are recognised solely as a consequence of the phenotype they cause (Levitt, 1991). It is widely acknowledged that establishing linkage with polymorphic markers is a crucial first step in the localisation and cloning of a gene associated with a pathological process (Levitt, 1991; Wilcox *et al.*, 1991). High resolution expression maps will expedite the process of localising disease genes (Levitt, 1991). The ability to detect polymorphisms in eukaryotic gene sequences facilitates their mapping and subsequent evaluation as candidate genes for involvement in an identified pathology (Levitt, 1991; Wilcox *et al.*, 1991; Takahashi and Ko, 1993). Mapping cDNAs, generated from RNAs of known spatial and temporal expression, will provide strong candidates genes for scrutiny in connection with genetic disorders.

Most published reports of EST sequencing utilise tissue-specific cDNA libraries derived from human tissues (Adams *et al.*, 1991, 1992; Wilcox *et al.*, 1991; Höög *et al.*, 1991; Okubo *et al.*, 1992). Likewise, the mouse ESTs mapped in this study (Section 1, 1.3) derive from subtracted libraries enriched for genes showing low level expression or localised expression in the mouse ventral midbrain (Savioz and Davies, 1995). The use of subtracted brain libraries as a source of ESTs significantly increases the probability of detecting candidate genes for neurological disorders by genetic mapping. A number of reports have already recognised the utility of brain-expressed ESTs in the study of neurogenetic disease (Wilcox *et al.*, 1991; Polymeropoulos *et al.*, 1992, 1993).

This candidate gene approach has been validated in a number of recent studies in the mouse and in man (Deutsch, 1991; Chalepakis *et al.*, 1991; Gibson *et al.*, 1995). Such an approach entails the localisation of a gene, known in the form of a mutant allele, on the mouse map whilst a cDNA is also known to map to the region. A plausible connection is first suspected and then examined. In one such example (Gibson *et al.*, 1995) the murine mutant *shaker-1* (*sha1*) was found to be a consequence of a mutation within a gene encoding a novel myosin protein. It has also been shown (Chalepakis *et al.*, 1991) that the homeobox containing the gene *Pax1*, found to map to the region of mouse chromosome 2 previously implicated in the undulated (*un*) mutation, was indeed the gene mutated in affected *un* animals. The process of mapping and subsequent identification of plausible candidate cDNA sequences was performed as outlined above. The result was confirmed by a genetic study which demonstrated that the homologous gene in man was responsible for Usher syndrome Type 1B (Weil *et al.*, 1995).

1.3 Identification and selection of genes for further study

In combined EST sequencing and genetic mapping projects a number of arbitrary criteria must be established in order to permit the identification and selection of gene sequences for further study. Such criteria may be based on the correlation of an EST location with a locus associated with a known pathology, or on the identification of ESTs corresponding to transcripts which are restricted to or preferentially present within a tissue affected by a known pathology. ESTs which are identified as satisfying both of these criteria are obvious candidate sequences for further study (Hochgeschwander and Brennan, 1993). As mentioned above, the mouse ESTs in these studies (Section 1, 1.3, 1.4 and 1.5) derive from subtracted libraries enriched for genes exhibiting a low level of expression or localised expression in the mouse ventral midbrain (Savioz and Davies, 1995), including in particular the Substantia nigra which is involved in movement control and the pathology of Parkinson's Disease. Consequently, genes exhibiting low level or localised expression to the ventral midbrain in mouse are of primary interest. Genes demonstrating expression patterns localised within defined regions of the brain or restricted to CNS tissues are also of interest and may be selected for further study. Genes which satisfy the above mentioned criteria, in relation to other pathological processes that do not specifically target the nervous system, are of secondary interest to these studies.

1.4 Selection of the *Mobp* gene for detailed analysis

A novel EST, corresponding to an oligodendrocyte-specific cDNA fragment, was isolated in the course of a combined differential screening and sequencing programme to discover genes expressed preferentially in particular brain regions (Davies *et al.*, 1994; Savioz and Davies, 1995). When mouse ventral midbrain and cerebellum cDNA libraries are compared, the former are found to be strongly enriched in oligodendrocyte-derived cDNAs (Savioz and Davies, 1995). One clone (*Ftg*; fibre tract gene) demonstrated an mRNA distribution, as determined by *in situ* hybridisation, in which all CNS white matter fibre tracts were labelled, exactly as for myelin basic protein (MBP). This contrasts with most other myelin proteins which are found to be localised in the oligodendrocyte cell bodies. It was initially suggested that this sequence may either correspond to a novel *Mbp* splice variant or that it may represent a novel gene encoding a new protein that is incorporated into CNS myelin (Davies, pers. comm.). While this work was in progress, a new oligodendrocyte cDNA was discovered in rat spinal cord (Yamamoto *et al.*, 1994), which showed strong sequence similarity to mouse *Ftg*. The product of this rat gene was shown to be abundant in CNS myelin, at a level exceeded only by PLP and MBP, and was named myelin-associated oligodendrocytic basic protein (*Mobp*; Yamamoto *et al.*, 1994). The

mouse gene *Ftg* has subsequently been renamed *Mobp* in accordance with its homologue in rat (Yamamoto *et al.*, 1994). Subsequent analysis (Montague *et al.*, 1997) confirmed that the *Mobp* gene is specifically expressed in oligodendrocytes in concert with the development of myelination and that its product represented a new major CNS-specific myelin component. Myelin-associated oligodendrocytic specific basic protein (MOBP) is distributed throughout CNS compact myelin. A decision was taken to select this gene for further study rather than to wait for the eventual identification of a gene whose expression was restricted to the substantia nigra. In the context of these studies, waiting for the identification of such a gene may have resulted in insufficient time for its study. The work reported in Section 1 of this thesis represents a series of pilot studies and as such the search for a gene demonstrating expression that is restricted to or enhanced within the substantia nigra continues.

1.5 Aims of the studies performed in Section 1

The aim of this study was to identify gene candidates for involvement in neuropathies affecting man and mouse, specifically genes whose expression was restricted to the substantia nigra. I have established that this goal may be achieved by identifying genes which demonstrate linkage to a locus associated with a known pathology, whose expression is restricted to the tissues affected by a known pathology or which satisfy both of these criteria (Chapter 1, 1.1-1.4). I have endeavoured to establish a technically feasible and efficient way to assay large numbers of expressed sequences with respect to these criteria. In order to satisfy the aim of this study I defined a number of interim goals. These include:

- To establish an SSCP assay that will identify polymorphisms in 3'UTR sequences, between the mouse species *Mus spretus* (SEG/Pas and SPR) and *Mus domesticus* (C57BL/6), with a high frequency.
- To utilise this SSCP assay to establish a genetic mapping strategy for genes expressed in the central nervous system (CNS) of mouse and subsequently to increase the rate at which ESTs were mapped in the mouse genome via increased cooperation between the Davies laboratory (University of Glasgow) and the Guénet laboratory (Institut Pasteur, Paris).
- To establish a qualitative RT PCR assay of the spatial and temporal expression patterns of ESTs selected from the genetic mapping study

Figure 1.1 illustrates the steps taken, in this study, from isolation of tissue from a particular developmental stage, through the generation of ESTs to their eventual localisation within the mouse genome.

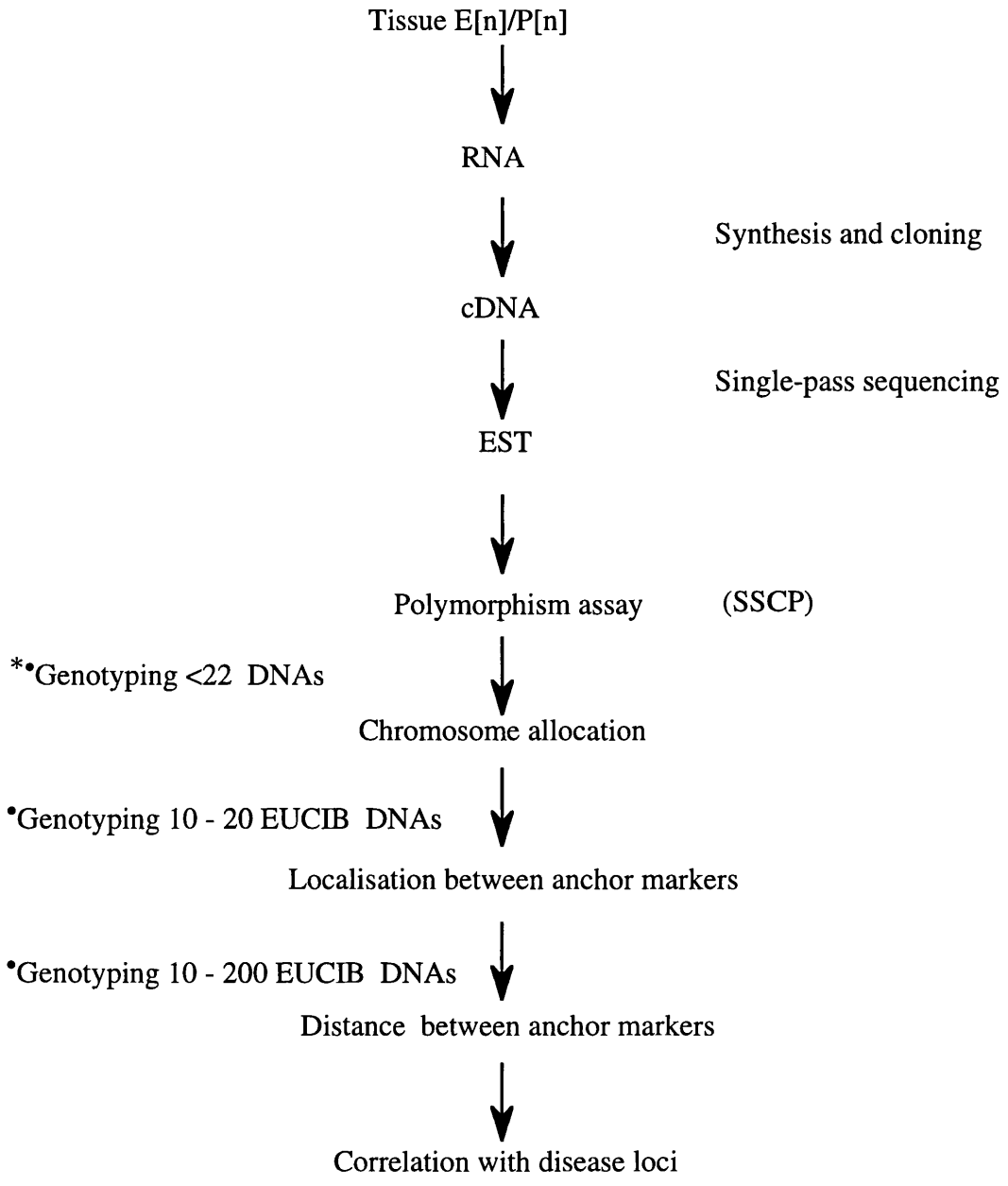


Figure 1.1 Schematic representation of the processes involved in the generation of ESTs from a particular tissue of mouse and their subsequent localisation within the mouse genome. E[n], Embryonic day [n]; P[n], Postnatal day [n]; EST, Expressed Sequence Tag; SSCP, Single Strand Conformation Polymorphism; EUCIB, European Collaborative Interspecific Backcross. •, subset panels of DNAs used in localisation of ESTs within the mouse genome. *, Intermediate generations of breeding programme established to generate consomic lines.

Section 2 Introduction

1.6 Myelin

1.6.1 Myelin structure and formation

Myelin is a unique cellular organelle of the central (CNS) and peripheral (PNS) nervous systems (Mikoshiba *et al.*, 1991; Lemke, 1992; Nave, 1994; Scherer and Chance, 1995; Staugaitis *et al.*, 1996). It forms a multilamellar membrane that ensheaths nerve axons and facilitates saltatory conduction of impulses (Mikoshiba *et al.*, 1991; Lemke, 1992; Staugaitis *et al.*, 1996). Myelin's unusual structure is readily recognised both macroscopically in the white matter tracts of the brain and spinal cord, and microscopically in the spiral rings of membrane that envelope all rapidly conducting axons (Lemke, 1992).

Myelin is formed by two specialised types of glia, oligodendrocytes in the CNS and myelin forming Schwann cells in the PNS (Jacobson, 1991; Lemke, 1992; Scherer and Chance, 1995). Figure 1.2 illustrates the myelination of multiple axons by a single oligodendrocyte and the characteristic structure of the myelin sheath. Embryologically, oligodendrocytes derive from multipotent progenitor cells of the neural tube. Schwann cells derive from the migratory cells of the neural crest, as do their neuronal partners in the PNS (Jacobson, 1991; Lemke, 1992). It is noteworthy that these cell types represent only a subset of peripheral and central glia (Lemke, 1992).

The mechanics of myelin deposition by oligodendrocytes and myelinating Schwann cells appear, at first aproximation, to be the same (Lemke, 1992): each synthesises a large sheet of plasma membrane which is spirally wrapped around target axons and subsequently tightly compacted (Mikoshiba *et al.*, 1991; Scherer and Chance, 1995; Staugaitis *et al.*, 1996). In the PNS this process results in a sheath comprising up to one hundred layers (Lemke, 1992). When viewed in cross section, through an electron microscope, the final structure of myelin resembles a highly regular array of concentric rings of plasma membrane bilayers (Jacobson, 1991; Lemke, 1992; Staugaitis *et al.*, 1996). This structure arises as a consequence of two different membrane appositions. One of these corresponds to the apposition of cytoplasmic membrane surfaces. The other corresponds to the apposition of extracellular plasma membrane surfaces. These membrane appositions have been termed major dense lines (MDL) and intraperiod lines (IPL) relating to the apposition of cytoplasmic and extracellular surfaces, respectively, and based on their appearance under the electron microscope (Privat, 1980; Lemke, 1992; Staugaitis *et al.*,

1996). The distance between adjacent MDL defines one repeat period of the myelin sheath. This distance is exceedingly small; approximately 170 Angstroms and 150 Angstroms in the PNS and CNS respectively. Membrane bilayers are known to be approximately 50 Angstroms wide (Lemke, 1992). Each MDL requires the apposition of two bilayers, each approximately 50 Angstroms wide. This indicates that there is very little cytoplasm in the spirals of compact myelin (Lemke, 1992; Staugaitis *et al.*, 1996).

1.6.2 Myelin function and neural evolution

During the propogation of an action potential very few ions move across axonal membranes in regions that are insulated by myelin. Current flow is instead restricted to regions of bare axonal membrane (Nodes of Ranvier), regularly spaced between adjacent myelin sheaths (Lemke, 1992; Staugaitis *et al.*, 1996). These Nodes of Ranvier contain a high density of voltage sensitive sodium channels, required for impulse conduction (Jacobson, 1991; Lemke, 1992), and are exposed to a relatively low resistance environment. This geometry compells the current to jump from one node to another in a saltatory fashion, a much faster mode of conduction than in unmyelinated axons (Lemke, 1992; Staugaitis *et al.*, 1986). It should be noted that a myelinated axon of fixed radius conducts nerve impulses at a rate 10 times faster than an unmyelinated axon of the same radius. As the volume occupied by an axon is directly proportional to the square of it radius, myelinated axons occupy on average only 1/100 the volume of an unmyelinated axon conducting impulses at the same rate (Lemke, 1992; Staugaitis *et al.*, 1996). This single factor is considered to be the driving force behind the evolution of myelinating cells and the appearance of myelin, constituting the last major cellular advance in evolution (Lemke, 1992). Phylogenetically, myelin is first observed in elasmobranch fish (Lemke, 1992). One critical difference between myelination by Schwann cells and myelination by oligodendrocytes relates to the number of axons that can be myelinated by any single cell. One Schwann cell can only myelinate a single axon, whereas a single oligodendrocyte can provide independent myelin sheaths for up to 50 different axons (Jacobson, 1991; Lemke, 1992). This permits a reduction in cell number and consequently affords a considerable saving in brain volume dedicated to cell bodies (Lemke, 1992; Staugaitis *et al.*, 1996). The development of a more complex and versatile nervous system may have been the driving force behind the transition, in the CNS, from myelination by Schwann cells to myelination by oligodendrocytes.

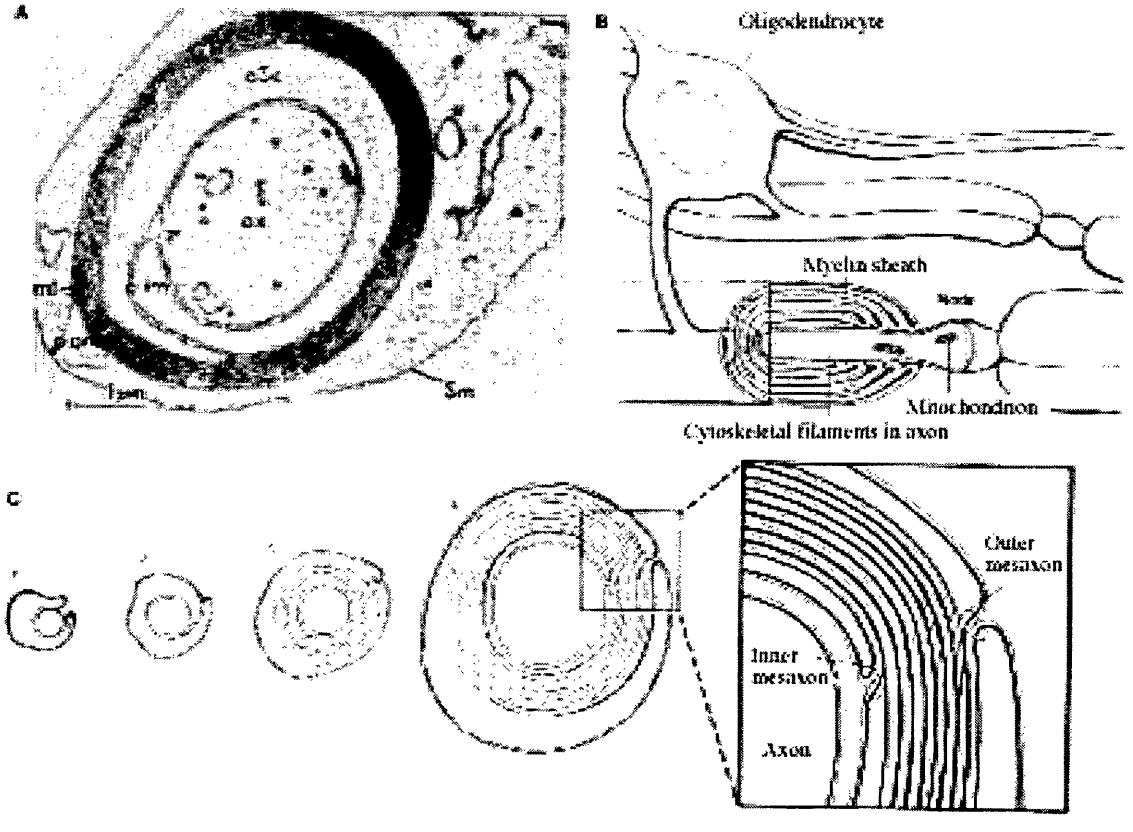


Figure 1.2 The processes of an oligodendrocyte forming a myelin sheath around an axon of the central nervous system. Figure 1.2 A illustrates the characteristic structure of myelin as observed microscopically. Figure 1.2 B illustrates the myelination of >1 axon by a single oligodendrocyte. Figure 1.2 C illustrates the process of formation of a myelin sheath around an axon. Figure adapted from Principles of Neural Science; Kandell, Schwartz and Jessell, 1991.

1.7 Major and minor myelin genes

During their differentiation into myelin producing cells, oligodendrocytes and Schwann cells activate expression of a group of myelin specific genes. These genes encode proteins that are involved in the mechanics of myelination: induction of myelination, initial contact with axons and deposition of the myelin sheath, and the wrapping and compaction of myelin about an axon. These proteins, grouped into major and minor sets (refer also to Table 1.1) based on their abundance in myelinating cells, have proved relatively straight forward targets for molecular cloning. At the time this study was initiated nine of these myelin genes had been cloned (Campagnoni, 1988; Lemke, 1988; Mikoshiba *et al.*, 1991; Suter *et al.*, 1993; Nave, 1994).

The major component proteins of myelin are listed in Table 1.1 (suffixed by *). These include myelin protein zero (P0), peripheral myelin protein 22 (PMP22), proteolipid protein (PLP) and myelin basic protein (MBP). P0 is exclusively expressed in peripheral myelin (Campagnoni, 1988; Lemke, 1988; Lemke, 1992). It is a 30 kD integral membrane glycoprotein and accounts for 50% of PNS myelin. P0 is is a member of the immunoglobulin (Ig) super family (Lemke, 1992) which includes the neural cell adhesion molecule (N-CAM). Its extracellular Ig domain appears to function as a homophillic cell adhesion molecule and as such promotes the formation of the intraperiod line. PLP is also a 30 kD integral membrane protein (Lemke, 1988; Mikoshiba *et al.*, 1991; Lemke, 1992). It is the major structural protein of central myelin (Lees and Brostoff, 1984; Mikoshiba *et al.*, 1991; Lemke, 1992) accounting for 50% of the protein therein (Lees and Brostoff, 1984). Another proteolipid protein DM20, which is produced by the same transcriptional unit as PLP, demonstrates immunological cross reactivity with the major PLP (Lees and Brostoff, 1984). PLP expression is primarily restricted to the oligodendrocytes of the CNS. However, it can be detected at low levels in the myelinating Schwann cells of higher vertebrates. PLP is believed to be crucial in the compaction and maintenance of the intraperiod line of CNS myelin (Duncan *et al.*, 1987; Mikoshiba *et al.*, 1991). It has been suggested that PLP may promote the apposition of extracellular surfaces of myelin lamellae (Mikoshiba *et al.*, 1991). MBP is actually a family of closely related protein isoforms (Lemke, 1992; Staugaitis *et al.*, 1996) and is expressed in both central and peripheral myelin. It comprises approximately 30% of protein in central myelin (Omli *et al.*, 1982) and 5-15% in the PNS (Lemke, 1992). Unlike P0 and PLP, MBP is not an integral membrane protein. It is, in fact, a highly cationic membrane- associated protein (Lees and Brostoff, 1984; Mikoshiba *et al.*, 1991) that is believed to interact with the negatively charged phospholipids of the cytoplasmic membrane (Mikoshiba *et al.*, 1991).

MBP is translated on free ribosomes in both oligodendrocytes and Schwann cells (Coleman *et al.*, 1982; Griffiths *et al.*, 1989). There are at least four isoforms of MBP in rat, and six in the mouse (Lees and Brostoff, 1984; Newman *et al.*, 1987; Nakajima *et al.*, 1993). They are all translated from mRNAs generated by the alternative splicing of a primary transcript from the single *Mbp* gene. MBP is thought to be involved in the maintenance and compaction of the major dense line i.e. the apposition of the cytoplasmic surfaces of the membrane (Lees and Brostoff, 1984). PMP22 is a 22 kD glycosylated integral membrane protein, which is similar in topology, with four transmembrane (TM) domains, to PLP (Suter *et al.*, 1993).

Thus far five minor myelin genes (Table 1.1), encoding the enzyme 2', 3' cyclicnucleotide 3' phosphodiesterase (CNP), myelin-associated glycoprotein (MAG) Myelin / oligodendrocyte glycoprotein (MOG), the fatty acid transport protein (P2) and the oligodendrocyte-myelin glycoprotein (OMgp), have been cloned (Lemke, 1992; Suter *et al.*, 1993).

MAG is a heavily glycosylated 100 kD integral membrane protein. It is expressed at low levels in both central and peripheral myelin. Like P0, MAG is also a member of the Ig super family and related to N-CAM (Salzer *et al.*, 1987). It mediates adhesion of the processes of myelinating cells to their target axons but is finally excluded from compact myelin (Salzer *et al.*, 1987; Bartsch, 1996). OMgp is also a large integral membrane glycoprotein and has a structure similar to known adhesion molecules (Lemke, 1992). CNP catalyses hydrolysis of 2', 3' cyclic nucleotides. Despite being enriched in myelin, the role of CNP has not yet been elucidated. P2 is a small (14 kD, Table 1.1) myelin specific cytoplasmic protein. It possesses an amino acid sequence that demonstrates strong similarity to fatty acid binding and transport proteins (Lemke, 1992). It is thought, in some way, to facilitate the exponential increase in the plasma membrane during the process of myelination. MOG is a 24-26 kD minor component of CNS myelin. It is primarily located on the external surfaces of the peripheral lamellae of CNS myelin (Pham-Dinh *et al.*, 1993). Its location on the external membrane surface of compact myelin suggest a role for MOG in the latter stages of myelination and maintainance of the myelin sheath (Pham-Dinh *et al.*, 1993).

Myelin protein	CNS	PNS	Molecular weight (kD)[a]	Mouse chromosome, mutation	Associated neuropathy in man
Protein zero* (P_0)	/	> 50%	25	1 *$P_0^{-/-}$*	CMT (Type 1B and Déjérne Sottas)
Proteolipid protein* (PLP)	50%	< 1%	30 26	X, *jimpy* *$Plp^{-/-}$*	CMT (Type 1A), PMD,XSP
Peripheral myelin protein* (PMP-22)	/	ND	22	11, *trembler* *$Pmp22^{-/-}$*	HNPP, CMT (Type 1A)
Myelin basic protein* (MBP)	≈ 30%	≈ 10%	14-21 (multiple forms)	18, *shiverer*	/
Myelin-associated glycoprotein•(MAG)	1%	< 1%	60 64	7 *$Mag^{-/-}$*	/
Myelin / oligodendrocyte glycoprotein• (MOG)	ND	ND	24-26	17	/
P2 basic protein• (P_2)	/	variable	14	ND	/
2'3'-cyclic nucleotide 3' phosphdiesterase• (CNP)	40%	< 1%	48 46	ND	/
Oligodendrocyte-myelin glycoprotein• (OMgp)	1%	/	46	17	/

Table 1.1 Myelin proteins known prior to the identification of MOBP

[a]Molecular weights are based on cDNA sequence and do not reflect posttranslational modifications such as glycosylation (P_0, MAG, OMgp) or phosphorylation (MBP, CNP). The major component proteins of myelin are listed above, suffixed by *. The minor component proteins of myelin are listed above, suffixed by •.
CMT, Charcot-Marie-Tooth disease; PMD, Pelizaeus-Merzbacher disease; XSP, X-linked spastic paraplegia; HNPP, hereditary neuropathy with liability to pressure palsies; $^{-/-}$, indicates those genes in which null mutations have been generated. ND, No Data (Note that this does not indicate that the corresponding protein is absent form the tissue in question). MOBP is absent from the PNS but present in the CNS and is estimated to comprise up to 20%, of the myelin protein present therein (Montague, pers. comm.)

1.8 Mutations in myelin genes

Genetic defects of development have provided insight into the possible roles of individual genes involved in cellular differentiation and brain function (Nave, 1994). In recent years a group of myelin specific proteins, expressed late in glial development, has been implicated in a number of mutant phenotypes (Nave, 1994). This has permitted great progress to be made in the elucidation of myelin disorders, from mutant gene to the whole organism. Most of these phenotypes demonstrate near total dysmyelination (Campagnoni, 1988; Lemke, 1988; Mikoshiba *et al.*, 1991). Gaps still remain in our understanding of the cellular consequences of some mutations within myelin genes. However the availability of a range of mutants, demonstrating perturbation of the same systems, now provides us with the ability to ask more precise questions. Rodents demonstrating dysmyelination exhibit a characteristic behavioural phenotype, a body tremor beginning around the third postnatal week (Nave, 1994). It is noteworthy that all spontaneous mutants were originally preselected for study on the basis of their phenotypes. Gene targeting is providing and will continue to provide novel mutations in the remaining myelin genes, for which no phenotype may be predicted. In a recent review Nave (1994) stated that estimation of the modulation and compensation of a phenotype required a detailed knowledge of all other proteins pertinent to the function of the cell in question.

Mutations in genes encoding myelin components have been shown to underlie a number of human inherited neuropathies, including Charcot-Marie-Tooth 1A (PMP22 mutations) and 1B (Myelin-P0 mutations) and Pelizaeus-Merzbacher disease (PLP mutations). Mutations in several mouse genes encoding myelin components have been well characterised (Mikoshiba *et al.*, 1991), including MBP (*shiverer, myelin-deficient*), PLP (*jimpy, rumpshaker*) and PMP22 (*trembler*), which are useful models of human neuropathies. The *shiverer* mouse was the first dysmyelination mutant to be characterised at the molecular genetic level (Roach *et al.*, 1985) and remains the best understood. A review of known mutations in genes encoding myelin components is given below and is also listed in Table 1.1.

1.8.1 Mutations in the *Mbp* gene encoding myelin basic protein

Shiverer (*shi*) and myelin deficient (*shi-mld*) are autosomal recessive murine mutants deficient in the expression of MBP (Mikoshiba *et al.*, 1991). The *shi* mutation is the result of a 20 kb deletion, caused by a genomic recombination event on mouse chromosome 18, of a region encompassing exons 3-7 (Roach *et al.*, 1985). It is noteworthy that these exons

reside within a larger transcriptional unit, *golli-Mbp*, which does not encode a myelin protein (Campagnoni *et al.*, 1993). None of the six MBP isoforms can be synthesised by *shiverer* mice. Consequently, oligodendrocytes fail to assemble compact myelin, resulting in hypomyelination of the CNS (Rosenbluth, 1980). The phenotype and myelination of *shiverer* mice can be rescued by the introduction of a transgene (1.3 kb of *Mbp* promoter sequence and a cDNA fragment encoding the 14 kD isoform of MBP; Kimura *et al.*, 1989).

Shiverer^{mld} (myelin deficient) is caused by a duplication of the entire *Mbp* gene and inversion of the region between exon 2 and the 3' end of the gene. The absence of MBP is a consequence of the formation of an endogenous antisense RNA which prevents translation of the *Mbp* RNA (Fremeau and Popko, 1990). The expression of *Shiverer*^{mld} MBP is described as being mosaic. That is to say that patches of MBP^{+} myelin sheath still form in the CNS. The amount of MBP in each oligodendrocyte varies widley (Akowitz *et al.*, 1987). The mechanism of this mosaicism is not yet fully understood (Mikoshiba *et al.*, 1991).

Both mutants demonstrate the absence of the MDL (Mikoshiba *et al.*, 1991). MBP is suggested to play a role in the maintenance of this structure (Campagnoni, 1988; Lemke, 1988; Mikoshiba *et al.*, 1991; Nave, 1994). The peripheral myelin of *shiverer* mice is morphologically normal (Mikoshiba *et al.*, 1991). It is believed that another myelin component e.g. P0, which has a basic intracellular domain, plays a role in compacting PNS myelin (Mikoshiba *et al.*, 1991; Nave, 1994). MBP production is rate limiting in CNS myelination. This is exemplified by the results of the above mentioned transgene complementation study (Kimura *et al.*, 1989), whereby increasing levels of *Mbp* transcript in mutant mice (*shiverer, shiverer*^{mld}) restores the degree of myelin assembly and formation of the MDL. It has been demonstrated that the single 14 kD MBP isoform is sufficient to serve most of the function of the entire MBP family of isoforms (Kimura *et al.*, 1989). Analyses of *shiverer* mice have helped to define the function of MBP in compact myelin (Mikoshiba *et al.*, 1991; Nave, 1994).

1.8.2 Mutations in the *Plp* gene encoding proteolipid protein

Plp has proved difficult to define genetically (Nave, 1994). However, a number of allelic mutations, allelic with *Plp*, are now available. Most information gained about *Plp* function

has been provided by the detailed analysis of rodents (Mikoshiba *et al.*, 1991; Nave, 1994). The majority of mutations reported in the *Plp* gene result in a lethal phenotype. These alterations span a range of mutation types, from major alterations of the protein structure e.g. in *jimpy* mice (which lack the PLP C-terminal domain), to conservative substitutions e.g. in the allele *jimpy*^{msd} (for myelin synthesis deficient; Ala 242→Val) (Nave *et al.*, 1986; Hudson *et al.*, 1987; Gencic and Hudson, 1990).

jimpy mice lack all functional PLP. The *Plp* RNA of *jimpy* mice demonstrates a 74 base deletion (Nave *et al.*, 1986). The deletion event is a consequence of an A→G transition event at the 3' splice site of exon 5 (*Plp*) and a subsequent failure to utilise exon 5 (*Plp*). The resulting protein contains an altered C-terminal as a consequence of the shift in reading frame. In *jimpy* mice the failure of oligodendrocytes to differentiate is followed by premature cell death prior to incorporation of the structural protein into myelin (Vermeesch *et al.*, 1990; Mikoshiba *et al.*, 1991). Oligodendrocytes which escape cell death may still assemble myelin-like structures which demonstrate slightly reduced spacing of the IPL. A report by Duncan (1989) suggests that PLP is not required for the adhesion of extracellular membranes but instead acts like a spacer or "strut" in the intraperiod line (IPL) of compact myelin (Duncan *et al.*, 1989; Yang and Skoff, 1997).

jimpy rumpshaker (*jimpy* ^{rsh}) mice arise as a consequence of a substitution event (Ile 186→Thr). The result is a marked dysmyelination of the CNS, yet oligodendrocytes are normal. Many *rumpshaker* oligodendrocytes generate myelin sheaths which demonstrate immunoreactivity with PLP antibodies. It has been suggested that mutations in the *Plp* gene can interfere with oligodendrocyte differentiation and without affecting the architecture of the myelin sheath (Schneider *et al.*, 1992).

Aberrant folding of PLP results in it being largely retained within the endoplasmic reticulum (ER) and consequently unable to reach the cell surface (Roussel *et al.*, 1987). Mutations in the gene encoding PLP result in interference with the differentiation of oligodendrocytes prior to myelination (Yang and Skoff, 1997). It has also been suggested that proliferating oligodendrocyte precursors, carrying mutations in the *Plp* gene, demonstrate an abnormal cell cycle and have an increased cell number (Knapp and Skoff, 1987). A number of recent studies (Barres *et al.*, 1992; Lachapelle *et al.*, 1992) have indicated that mutant cell death may be restricted to a specific stage of development and that extrinsic forces exist that may overide, at least in part, the lethality of *Plp* mutations

(Nave, 1994). However, these extrinsic factors remain to be elucidated.

Myelination is very sensitive to alteration in PLP dosage (Readhead *et al.*, 1994; Griffiths *et al.*, 1995). Twofold overexpression of the wildtype *Plp* gene is in itself lethal, causing CNS specific dysmyelination (Readhead *et al.*, 1994; Griffiths *et al.*, 1995). It is noteworthy that mice with a targeted disruption of the *Plp* gene demonstrate remarkably mild clinical signs. It has also been suggested that the lethal phenotype associated with most mutations in *Plp* is a complex combination of loss-of-function and gain-of-function effects (Griffiths *et al.*, 1995). Oligodendrocyte sensitivity to increased dosage of PLP underlies the human neuropathy Pelizaeus-Merzbacher disease (PMD). Mutations in the *Plp* gene have also been demonstrated to be associated with the human neuropathy X-linked spastic paraplegia (Griffiths *et al.*, 1995).

1.8.3 Mutations in the *P0* gene encoding myelin protein zero

P0 has been implicated in the compaction of peripheral myelin subsequent to analysis of mice carrying a targeted disruption of the *P0* gene (Martini *et al.*, 1995) and as a consequence of its membership of the Ig superfamily (Lemke, 1988). *P0* null ($P0^{-/-}$) mutants (Giese *et al.*, 1992; Martini *et al.*, 1995b; Zielasek *et al.*, 1996) demonstrate severe PNS specific dysmyelination. This mutation type has been associated with the human peripheral neuropathy Charcot-Marie-Tooth disease (CMT) type 1B and the more severe Déjérne Sottas disease (CMT type 3) (Zielasek *et al.*, 1996). Mice that are heterozygous ($P0^{+/-}$) for this mutation are much less affected and the resulting pathology also resembles the more common peripheral human neuropathy Charcot-Marie-Tooth disease (CMT) type 1B. Myelin structure is normal in younger individuals and progressive demyelination ensues with age (Martini *et al.*, 1995). This suggests that P0 may also demonstrate a gene dosage effect (Zielasek *et al.*, 1996). In $P0^{+/-}$ mice a characteristic and progressive tremor and abnormal gait are observable within 2-3 weeks of birth. The failure of Schwann cells to form normal myelin sheaths demonstrates a great degree of heterogeneity among individual cells (Martini *et al.*, 1995). These glia may fail to initiate wrapping of myelin sheaths around axons or the assembled myelin sheath may lack compaction. It is noteworthy that P0 function has been demonstrated to be developmentally downstream of glia neuron recognition; cells carrying this null mutation do still interact with axons and segregate in a 1:1 ratio (Martini *et al.*, 1995). Lack of one myelin component can, however, have a marked effect on the distribution of other myelin proteins and recognition molecules (Martini *et al.*, 1995). In this instance N-CAM and MAG are no longer

excluded from compact myelin and PLP, which is normally prevented from entering the myelin compartment of Schwann cells, is assembled into the the completed myelin structure. A recent report has also implicated a point mutation in the *P0* gene in the peripheral human neuropathy Charcot-Marie-Tooth disease (CMT) type 1B (Hayasaka *et al.*, 1993).

Axonal degeneration has been reported in $P0^{-/-}$ null mutant mice (Giese *et al.*, 1992) and in transgenic mice with an increased gene dosage (Readhead *et al.*, 1994). In neither case has the mechanism responsible been elucidated.

1.8.4 Mutations in the *Pmp22* gene encoding peripheral myelin protein 22

In the mouse mutants *trembler* and *trembler-J,* the predicted transmembrane (TM) domains of PMP22 are disrupted by a non-conservative substitution. This mutation results in a marked peripheral dysmyelination (Suter *et al.*, 1993). The cellular mechanisms which underlie this lack of myelin have not yet been elucidated (Nave, 1994). However, both of these alleles are semidominant, indicating either that mutant PMP22 behaves in a dominant negative manner or that one normal copy of the *Pmp22* gene is insufficient for myelin assembly (Suter *et al.*, 1993).

Pmp22 has been implicated in the human peripheral neuropathy Charcot-Marie-Tooth disease (CMT) type 1A and the more severe Déjérne Sottas disease (CMT type 3) (Scherer and Chance, 1995; refer also to Section 2, 2.6) and in the hereditary neuropathy with liability to pressure palsies (HNPP) (Scherer and Chance, 1995). The dominant disorder CMT 1A is associated with point mutations within the *Pmp22* gene or duplication of the chromosome (17) region encompassing the gene (Suter *et al.*, 1993).

1.9 The *Mobp* gene encodes a novel major myelin component

As described in Chapter 1, 1.4 a new oligodendrocyte-specific cDNA clone, corresponding to the gene now known as *Mobp*, was isolated in the course of a combined differential screening and sequencing programme to discover genes expressed preferentially in particular brain regions (Davies *et al.*, 1994; Savioz and Davies, 1995; Stewart *et al.*, 1996). Subsequent analysis (Montague *et al.*, 1997) confirmed that the *Mobp* gene is specifically expressed in oligodendrocytes in concert with the development of myelination and that its product represented a new CNS myelin component. While this work was in

progress, a new oligodendrocyte cDNA was discovered in rat spinal cord (Yamamoto *et al.*, 1994), which showed strong sequence similarity (>90%) to the mouse gene *Ftg*. The product of this rat gene was shown to be abundant in CNS myelin, at a level exceeded only by PLP and MBP, and was named myelin-associated oligodendrocyte basic protein (MOBP; (Yamamoto *et al.*, 1994)). The mouse gene *Ftg* has subsequently been renamed *Mobp* in accordance with its homologue in rat. MOBP is distributed throughout CNS compact myelin. Five isoforms have been identified (Yamamoto *et al.*, 1994; Holz *et al.*, 1996) which, in the rat, are 69, 71, 81, 99 and 170 amino acids long respectively. These small proline-rich basic proteins have features in common with MBP, and like MBP they are likely to be located on the inner surface of the cell membrane and to be involved in stabilising and compacting the multilayered structure (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). A more complete introduction to the identification of this gene and previous studies associated with it can be found in Section 2, 2.1.

1.10 Aims of the studies performed in Section 2

The aim of this study was to genetically map *Mobp* gene encoding the novel major component of CNS myelin (MOBP); to initiate studies in order to examine its structure and function and to evaluate its possible role in known neuropathies. In order to satisfy the aim of this study I defined a number of interim goals. These include:

- To genetically map the murine gene (*Mobp*) encoding myelin-associated oligodendrocytic basic protein, using the PCR-SSCP based genetic mapping strategy established in Section 1.

- To elucidate the complex splicing events required to generate the variety of splice variants produced from the *Mobp* transcript.

- To isolate the murine *Mobp* gene, encoding myelin-associated oligodendrocytic basic protein, and elucidate its genomic organisation.

- To examine the promoter region and identify the transcriptional start points (tsp) of the murine *Mobp* gene

- To establish a platform for studies examining the function of the *Mobp* gene product via the generation of a targeted disruption of *Mobp* in embryonic stem (ES) cells.

Chapter 2 Materials and methods

2.1 Bacteriological materials

2.1.1 Bacterial strains

Strain	Description	Genotype	Reference
TG-1	an EcoK$^-$ derivative of JM101 restriction deficient host for plasmids	*supE, hsdΔ5, thi, Δ (lac-proAB) F' [traD36, proAB$^+$, lacI^q, lacZΔ M15]*	(Gibson, 1984)
TG-2	a recombination deficient derivative of TG-1	*supE, hsdΔ5, thi, Δ (lac-proAB) Δ (srl-recA) 306 :: Tn10 (tet^r) F' [traD36, proAB$^+$, lacI^q, lacZΔ M15]*	M. Biggin, pers. comm.
Q358	A *supE* host used for growth of bacteriophage λ vectors	*supE, hsdR φ 80^r*	(Karn *et al.*, 1980)
INVα F'	TA cloning® vector for the cloning of PCR products	*F' endA1, recA1, hsdR17, (r_k^- m_k^+), supE44, relA1, thi-1, gyrA96, φ80lacZΔ M15Δ (lacZYA-argF)Y169λ$^-$*	Invitrogen® version B 150626 25-0024
BNN 132	a derivative of JM107 containing a lysogen (recombinant bacteriophage λ, expressing *cre* recombinase)	*(endA1, gyr96,thi, hsdR17, supE44,relA1, Δ (lac-proAB) F' [traD36, proAB$^+$, lacI^q, lacZΔ M15]*	(Elledge *et al.*, 1991)

Table 2.1 *Escherichia coli* strain information. Bacterial strains used in this work are listed by name, description, genotype and associated reference. Media for bacterial growth were prepared as described in Sambrook et al., (1989).

2.1.2 Commercially available, gifted and engineered vectors

Vector / construct	Description	Reference
pBS KS+ II	Phagemid derived from pUC 19	(Short *et al.*, 1988)
pCR™ 2.1	TA cloning® vector - permits the cloning of dA-tailed PCR products	Invitrogen® version B 150626 25-0024
pNT	Vector backbone for the generation of gene disruption constructs. Contains the bacterial neomycin resistance and HSV Thymidine Kinase genes.	Skarnes pers. comm.
λPS	Large insert vector which facilitates rapid mapping and targeted disruption of mammalian genes	(Nehls *et al.*, 1994)
λ2001	A new selective phage cloning vector, lambda 2001, with sites for *Xba* I, *Bam* HI, *Hin* dIII, *Eco* RI, *Sst* I and *Xho* I	Karn *et al.*, 1984
Mobp KO	Derived from pNT; a targeting construct designed to disrupt the murine gene *Mobp*	McCallion, unpublished

Table 2.2 Vector and construct information

Vectors and constructs used in this work are listed by name, description and associated reference.

2.2 Commercial sources of chemicals and enzymes

Chemical / Enzyme	Source
General chemicals and solvents	BDH, ICN
Biochemicals	Sigma, Gibco BRL
Agarose MP	Boehringer Mannheim
X-ray film	Kodak, Dupont
Restriction enzymes	Promega
Modifying enzymes	Promega
PCR product purification systems	Promega
Acrylamides	Applied biosystems (ABI)

Table 2.3 Sources of chemicals and enzymes
General chemicals and enzymes utilised in this study are listed with their commercial source. Bacterial growth media were purchased from Media preparation, Division of Molecular Genetics, Glasgow University. Solutions and protocol specific chemicals and enzymes are defined in the text.

2.3 Molecular biological materials

2.3.1 Sources of genomic and cDNA libraries

2.3.1.1 Genomic DNA libraries

Amplified libraries prepared in the bacteriophage vectors λ2001 (~0.8×10^5 plaque forming units, *pfu*) (Warren *et al.*, 1994) and λPS (~0.4×10^5 *pfu*) (Nehls *et al.*, 1994) were supplied by Dr Andrew Smith (University of Edinburgh, Scotland).

2.3.1.2 Subtracted cDNA libraries

Details of the construction of the cDNA libraries utilised in this study have been reported previously (Savioz and Davies, 1995). The libraries utilised in this study were enriched for rare transcripts by subtractive hybridisation, yielding two subtracted libraries. In both cases the target cDNA population was single-stranded (sense) DNA derived from the cDNA library. In the case of the first subtracted library, designated A (Savioz and Davies, 1995), the driver nucleic acid population was single-stranded first strand (antisense) cDNA derived from mouse cerebellar polyA$^+$ mRNA. The second library,

designated B, was constructed using antisense RNA synthesised *in vitro* from a directional mouse cerebellar cDNA library as the driver population. Both libraries were stored as 900 colony grids on 10 cm nutrient agarose plates, overlaid with Whatman 3 MM paper soaked in 60% Glycerol in SM (Sambrook *et al.*, 1989) at -70°C. The A library consists of 12905 clones and the B library of 17545 clones.

2.3.2 Sources of ribonucleic acids (RNAs)

2.3.2.1 Source of messenger (poly A^+) ribonucleic acids (mRNAs)

Messenger RNAs, from the cerebellum of adult male C57BL/6 mice, were kindly supplied by Dr T.A.Glencorse and Mrs A.B.Roberts (University of Glasgow).

2.3.3 Source of first strand cDNAs derived from total ribonucleic acids (RNAs)

RNAs, for RT-PCR assays, from the whole brain (postnatal day 4 [P4], P5, P16, P43); spinal cord (P1, P5); sciatic nerve (P15); kidney (P15); spleen (P15); testis (P15) and liver (P15) of adult male C57BL/6 mice, were kindly supplied at 2.5 ng/µl by Dr P. Montague (University of Glasgow, Scotland).

2.4 Methods

2.4.1 Basic molecular biological techniques

The following standard molecular biological techniques were performed as described by Sambrook *et al.* (1989): bacterial growth, isolation of genomic, bacteriophage and plasmid DNAs, DNA purification (i.e. phenol extraction, ethanol precipitation etc.), restriction enzyme digestion, agarose / polyacrylamide gel electrophoresis and the transfer of nucleic acids to nylon membranes via Southern blotting. Details of the above procedures, specific to this study, are outlined below.

2.5 Bacterial growth

2.5.1 Growth of bacterial host for bacteriophage lambda

Overnight cultures of *E. coli* strain Q358 (Karn *et al.*, 1980) were established in NZCYM liquid medium. Maltose was present in the medium at a final concentration of 20% weight/volume (w/v).

2.5.2 Growth of bacteria used in molecular cloning strategies

Overnight cultures of *E. coli* strain BNN 132 (Elledge *et al.*, 1991) were established in LB liquid medium. Overnight cultures of *E. coli* strains TG-1 (Gibson, 1984) and TG-2 (Biggin, pers. comm.) were established in 2 × YT liquid medium.

2.6 Propagation of automatically subcloned plasmids from bacteriophage λPS

2.6.1 Cre-*loxP* automatic subcloning

10 ml of fresh overnight culture of BNN 132 were centrifuged at 5,000 rpm (JA-20 rotor, Beckman J2-21 centrifuge) for 10 minutes, the supernatant aspirated and the pellet resuspended gently in 6 ml of 10 mM $MgSO_4$. 300 µl aliquots were removed and combined with 100 µl of each high titre bacteriophage stock (AS; 50, 51, 53, 55, 57, 58 and 59) in clean 1.5 ml centrifuge tubes. These were incubated at 30°C for one hour without agitation. 1 ml of LB was then added to each and the tubes incubated with agitation at 30°C for one hour. 200 - 400 µl of cells from each tube were then spread on 90 mm LB plates with ampicillin at 50 µg/ml, 0.1% glucose and incubated overnight at

37°C. Ampicillin resistant colonies were scraped from these plates into 1.5 ml of LB for small scale isolation of plasmid DNA by alkaline lysis and PEG precipitation (Chapter 2, 2.8.2).

2.7 DNA purification

2.7.1 Phenol extraction

Phenol (TE saturated: 10 mM Tris, 1 mM EDTA) was added to each tube in a ratio of 1:1 (v/v) and mixed vigorously by rapid inversion. The tubes were centrifuged (Heraeus, Biofuge; unless otherwise stated) at 15,000 rpm for two minutes and the aqueous phase removed to a clean tube.

2.7.2 Chloroform extraction

Chloroform:isoamyl alcohol (24:1) was added to each tube in a ratio of 1:1 (v/v), mixed by inversion and then centrifuged at 15,000 rpm for two minutes The upper phase was then removed to a clean tube.

2.7.3 Ethanol and Isopropanol precipitation

To each tube 0.1 volumes of 3 M sodium acetate and either 2.5 volumes of ethanol or an equal volume of isopropanol were added. The tubes were then mixed by inversion and placed on ice for 10 - 120 minutes (protocol dependent) and centrifuged at 15,000 rpm for 10 - 30 minutes (protocol dependent).

2.8 Isolation and purification of nucleic acids

2.8.1 Isolation of bacteriophage lambda DNA

2.8.1.1 Small scale isolation of bacteriophage lambda DNA

Liquid lysates were prepared as described in Chapter 2 (2.13.2). 1.5 ml of lysate was added to a 2 ml centrifuge tube. Each lysate was centrifuged at 15,000 rpm for 5 minutes and the supernatant removed to a clean tube. 3 µl of DNase (20 mg/ml) and 3 µl RNase (20 mg/ml) were added to each tube. The tubes were mixed by inversion and incubated at room temperature for 30 minutes. To each tube 375 µl of PEG solution (26% PEG, 2.6 M NaCl) was added, to aid the precipitation of phage particles. The tubes were mixed by inversion, placed on ice for one hour and then centrifuged at 15,000 rpm for 5 minutes.

The supernatant was aspirated and the remaining pellet resuspended in 200 µl of TE $_{10,10}$ (10 mM Tris pH8, 10 mM EDTA). The tubes were maintained on ice throughout. 50 µl of phenol (TE saturated) were added to each and mixed vigorously. The tubes were centrifuged at 15,000 rpm for three minutes and 180 µl of each supernatant removed to a clean tube. To each tube 20 µl of 3 M sodium acetate and 200 µl of isopropanol were added, mixed by inversion and placed at 4°C for 2 hours. The tubes were centrifuged at 15,000 rpm for 15 minutes and the supernatant aspirated. 500 µl of ethanol (70%) was added to each tube. The tubes were centrifuged at 15,000 rpm for 2 minutes and the supernatant aspirated completely. The pellet was air dried for approximately 10 minutes and resuspended in 20 - 40 µl of TE (Tris 10 mM, EDTA 1 mM).

2.8.1.2 Large scale isolation of bacteriophage lambda DNA

Liquid lysates were prepared as described in Chapter 2 (2.13.2). 25 ml of lysate were added to a 50 ml polypropylene centrifuge tube. Each lysate was centrifuged at 8,000 rpm for 5 minutes (JA-20 rotor, Beckman J2-21) and the supernatant removed to a clean tube. 25 µl of DNase (20 mg/ml) and 3 µl RNase (20 mg/ml) were added to each tube. The tubes were mixed by inversion and incubated at room temperature for 30 minutes. To each tube 5 ml of PEG solution (26% PEG, 2.6M NaCl) was added; to aid the precipitation of phage particles. The tubes were mixed by inversion and placed on ice for one hour and then centrifuged at 15,000 rpm for 20 minutes (JA-20 rotor, Beckman J2-21). The supernatant was aspirated and the remaining pellet resuspended in 1.5 ml of phage buffer and transferred to a 2 ml centrifuge tube. 375 µl of PEG were added to each tube and the tubes mixed by inversion and placed on ice for 30 minutes. The tubes were then centrifuged at 15,000 rpm for 10 minutes and supernatant was aspirated completely. The remaining pellets were resuspended in 600 µl of TE $_{10,10}$ (10 mM Tris pH8, 10 mM EDTA). The tubes were maintained on ice. 600 µl of phenol (TE saturated) were added to each tube and mixed vigorously. The tubes were centrifuged at 15,000 rpm for three minutes and the aqueous phase removed to a clean tube. This phenol extraction was repeated a further two times and the aqueous phase once more removed to a clean tube. To each tube 60 µl of 3M sodium acetate and 600 µl of isopropanol were added, mixed by inversion and placed on ice for 2 hours. The tubes were centrifuged at 15,000 rpm for 15 minutes and the supernatant aspirated. 500 µl of ethanol (70%) was added to each tube. The tubes were centrifuged at 15,000 rpm for 2 minutes and the supernatant aspirated. The pellet was air dried for approximately 10 minutes and resuspended in 200 µl of TE (10 mM Tris, 1 mM EDTA).

2.8.2 Small scale isolation of plasmid DNA via alkaline lysis and PEG precipitation

1.5 ml of fresh overnight culture were placed in a 1.5 ml centrifuge tube and centrifuged at 15,000 rpm for 5 minutes. The supernatants were aspirated and the pellets resuspended in 200 µl of GET solution (50 mM glucose, 10 mM EDTA pH 8, 25 mM Tris pH 8). 300 µl of 200 mM NaOH / 1% SDS (freshly made) were added to each tube and mixed by gentle inversion. The solutions were then neutralised by the addition of 300 µl 3M potassium acetate and mixed by inversion. The samples were centrifuged at 15,000 rpm for 10 minutes and the supernatant removed to a clean tube. 2 µl of RNase (20 mg/ml) were added to each tube, mixed by inversion and incubated at room temperature for 30 minutes. 400 µl of chloroform:isoamyl alcohol (24:1) were added to each tube, mixed by inversion and then centrifuged at 15,000 rpm for 2 minutes and the upper phase removed to a clean tube. This chloroform extraction was repeated and the upper phase once more removed to a clean tube. To each tube an equal volume of isopropanol was added, mixed by inversion and centrifuged at 15,000 rpm for 10 minutes. The supernatant was aspirated and 500 µl of ethanol (70%) were added to each pellet. The tubes were centrifuged at 15,000 rpm for 2 minutes and the supernatant aspirated completely. The pellets were resuspended in 32 µl of sterile water. 42 µl of 13% PEG were then added to each tube, mixed by flicking and placed on ice for 20 minutes. The tubes were then centrifuged at 15,000 rpm (4°C) for 20 minutes and supernatants were aspirated off. 500 µl of ethanol (70%) were added to each pellet. The tubes were centrifuged at 15,000 rpm for 2 minutes and the supernatant aspirated completely. The pellets were air dried for approximately 10 minutes and resuspended in 20 µl of TE (Tris 10 mM, EDTA 1 mM) or sterile water.

2.8.3 Large scale isolation of plasmid DNA for transfection into CGR 8.8 embryonic stem cells

Large scale isolation of *Mobp* KO DNA from the TG-2 bacterial host was performed using the Qiagen Midi prep kit in accordance with the manufacturers instructions.

2.8.4 Isolation of DNA from CGR 8.8 cultured embryonic stem cells

1.5 ml centrifuge tubes containing 0.6 ml of freezing media (refer also to Chapter 2, 2.26), with CGR 8.8 embryonic stem (ES) cells were centrifuged at 15,000 rpm for 1 minutes and the supernatant aspirated completely. To the cell pellets, 340 µl of reagent B (400 mM Tris pH 8, 60 mM EDTA, 150 mM NaCl and 1% SDS) were added and the pellet resuspended by flicking. 100 µl of 3 M sodium perchlorate were added and the samples

mixed by vigorous inversion. The tubes were then placed at 65°C for 10 minutes and mixed by occasional inversion. 580 µl of chloroform:isoamyl alcohol (24:1) were added to each tube, mixed by vigorous inversion and the tubes were placed at room temperature for 20 minutes. The samples were centrifuged at 15,000 rpm for 5 minutes and 400 µl, of the upper phase, removed to clean tubes. To these, 800 µl of ethanol were added and the tubes mixed by inversion. Samples were placed at room temperature for 2 minutes and centrifuged at 15,000 rpm for 5 minutes. The supernatants were aspirated and 1 ml of ethanol (80%) was added to each pellet. The tubes were centrifuged at 15,000 rpm for 5 minutes and the supernatant aspirated completely. The pellets were air dried and subsequently resuspended in 60 µl of TE (Tris 10 mM, EDTA 1 mM).

2.8.5 Gel extraction of nucleic acids for cloning strategies

Restriction fragments were separated on agarose gels as described in Chapter 2 (2.14) and the appropriate DNA fragment isolated using the Qiaex II gel extraction kit in exact accordance with the manufacturers instructions.

2.9 Isolation of total RNA from mouse brain

2.9.1 Small scale isolation of RNA from mouse cerebellum

All glassware, pipette tips and tubes were treated by immersion in a 0.1% solution of diethyl-pyrocarbonate (DEPC), for at least 16 hours prior to autoclave sterilisation in order to prevent contamination with RNase. All reagents were prepared using DEPC treated H_2O (DEPC-H_2O). 20-100 mg of tissue from the cerebellum of mouse brain (C57BL/6, adult males) was placed in a 1.5 ml centrifuge tube containing 1 ml of Tri-reagent™ (Sigma). The tissue was homogenised by vigorous pipetting. The centrifuge tubes were placed at room temperature for 5 minutes. 0.2 ml of chloroform:isoamyl alcohol (24:1) was added to each tube, mixed by vigorous inversion and then placed at room temperature for 10 minutes. The tubes were then centrifuged at 12,000 *g* (4°C) for 15 minutes. The upper phase from each tube was removed to a clean tube. To each tube 500 µl of isopropanol were added, mixed by inversion and then placed at room temperature for 10 minutes. The tubes were then centrifuged at 12,000 *g* (4°C) for 15 minutes and supernatants were aspirated off. The pellets were washed with cold (4°C) 75% ethanol and then centrifuged at 9,000 *g* (4°C) for 5 minutes. The supernatants were aspirated off and the pellets were resuspended in 20 µl of DEPC-H_2O.

2.9.2 Large scale isolation of RNA from mouse cerebellum

0.55 g of cerebellar tissue was placed in a sterile 50 ml polypropylene centrifuge tube and 5 ml of Tri-reagent™ (Sigma). The tissue was homogenised using a Polytron™ homogeniser. The centrifuge tubes were placed at room temperature for 5 minutes. 1 ml of chloroform:isoamyl alcohol (24:1) was added to each tube, mixed by vigorous inversion and then placed at room temperature for 15 minutes. The tubes were then centrifuged at 10,000 rpm (JA-20 rotor, Beckman J2-21) at 4°C for 15 minutes. The upper phase from each tube was removed to a clean tube. To each tube 2.5 ml of isopropanol were added, mixed by inversion and then placed at room temperature for 10 minutes. The tubes were then centrifuged at 10,000 rpm (JA-20 rotor, Beckman J2-21) at 4°C for 5 minutes and supernatants were aspirated off. The pellets were resuspended in 1 ml off 75% ethanol and stored at -70°C.

2.10 Synthesis of first strand cDNA

2.10.1 Synthesis of first strand cDNA for the RT-PCR assay

Reactions were set up in 20 µl volumes and contained: 50% v/v RNA, 40U RNasin, 1 × reverse transcription buffer (Promega), 0.1 mg/ml BSA, 200 µM dNTPs, 50 pM {$(dT)_{23}$-dC/dG/dA} and 10 U AMV reverse transcriptase. The reactions were placed at 42°C for one hour and subsequently placed at 95°C for 5 minutes in order to heat inactivate the enzyme. Subsequently, PCRs were set up as described in Chapter 2 (2.17) in 10-25 µl volumes and using 0.5-10 µl of template, dependent upon the efficiency of the RNA isolation.

2.10.2 Synthesis of first strand cDNA for the oligo-capping procedure

Reactions were set up in 20 µl volumes and contained: 20% v/v RNA (section 2.25), 40 U RNasin, 1 × Superscript™ II reaction buffer (Gibco BRL), 0.1 mg/ml BSA, 1 mM µM dNTPs, 50 pM (random hexamers, Gibco BRL) and 200 U Superscript™ II (Gibco BRL). The reactions were placed at 23°C for 10 minutes, 42°C for 45 minutes and then placed at 95°C for 10 minutes. Subsequently, PCRs were set up as described in Chapter 2 (2.17) in 10-25 µl volumes using 2 µl of template.

2.11 Ligation of nucleic acids

2.11.1 Ligation of plasmid vector and insert DNA

Plasmid (vector) and insert DNAs were prepared as described in Chapter 2 (2.8.2) and Chapter 2 (2.8.5) respectively. Typically, 20 - 200 ng of vector and insert DNAs were used for ligation. Vector and insert DNA concentrations were estimated by agarose gel electrophoresis, along with molecular weight standards of known concentration. A 1:3 molar ratio of vector:insert DNAs was utilised throughout. Using this vector:insert ratio, 10 µl reactions were set up in the presence of the following: T4 DNA ligase 1 U (Weiss units) and 1 × ligase buffer (final concentration).

2.11.2 Ligation of PCR products using cloning vector pCR™ 2.1

PCR products were cloned using the TA cloning® kit (Invitrogen® version B 150626 25-0024) in exact accordance with manufacturers instructions, and DNA sequence determined as described in Chapter 2 (2.15) utilising sequencing primers M13/pUC forward and reverse (Table 2.3).

2.12 Transformation of DNAs into bacteria

2.12.1 Transformation of λPS derived plasmid DNAs into bacterial cells

1 ml of a fresh overnight culture of TG-2 was added to 100 ml of 2 × YT and incubated with agitation for 2 - 3 hours (or until $OD_{600\ nm}$ = 0.4 - 0.6). The cultures were then added to 40 ml polypropylene centrifuge tubes and centrifuged at 3,000 rpm for 5 minutes. The supernatant was then aspirated, the pellet resuspended in 25 ml of chilled (4°C) 50mM $CaCl_2$ and the tubes placed on ice for 20 minutes. The tubes were then centrifuged at 3,000 rpm for 5 minutes, the supernatant aspirated and the pellet resuspended in 5 ml of chilled (4°C) 50 mM $CaCl_2$. 100 µl of cell suspension was added to approximately 50 ng of each plasmid DNA. The tubes were incubated on ice for 30 minutes, placed at 42°C for 2 minutes and placed on ice. 1 ml of 2 × YT was added to each tube and incubated at 37°C for 30 minutes. 100-400 µl of cells were then plated on 90 mm LB plates with ampicillin at 50 µg/ml and incubated overnight at 37°C.

2.12.2 Transformation of plasmid DNA into bacterial cells

100 µl of a fresh overnight culture of TG-1 or TG-2 was added to 10 ml of 2 × YT and incubated with agitation for 2 - 3 hours (or until $OD_{600\ nm}$ = 0.4 - 0.6). The cultures were then added to 40 ml polypropylene centrifuge tubes and centrifuged at 2,000 rpm for 5 minutes. The supernatant was then aspirated, the pellet resuspended in 1 ml of chilled (4°C) transformation buffer (10 mM MES buffer pH 6.5, 100 mM rubidium chloride, 45 mM manganese chloride, 10 mM calcium chloride and 3 mM hexaminecobaltic chloride) and the tubes placed on ice for 15 minutes. The tubes were maintained on ice. 34 µl of N-N dimethyl formamide were then added, mixed by gentle inversion and placed on ice for 5 minutes. 34 µl of 5% v/v B-mercaptoethanol were then added, mixed by gentle inversion and placed on ice for 10 minutes. 34 µl of dimethyl formamide were then added, mixed by gentle inversion and placed on ice for 5 minutes. 100-200 µl of cell suspension was added to approximately 5-50 ng of each plasmid DNA. The tubes were incubated on ice for 30 minutes, placed at 42°C for 2 minutes and placed back on ice. 1 ml of 2 × YT was added to each tube and incubated at 37°C for 30 minutes. 100-400 µl of cells were then plated on 90 mm LB plates with ampicillin at 50 µg/ml and incubated overnight at 37°C.

2.13 Preparation of bacterial cell lysates

2.13.1 Preparation of bacterial cell lysates on solid media

A fresh overnight culture was added to melted NZCYM top agar (≤ 50°C) to provide a final concentration of 75 µl/ml v/v, agitated to mix and poured onto culture plates containing NZCYM bottom agar. These plates were allowed to cool to room temperature and set. A single plaque forming unit (*pfu*) of the desired bacteriophage was excised from top agar, streaked gently across the plate and incubated overnight at 37°C

2.13.2 Preparation of bacterial cell lysates in liquid media

2.13.2.1 Lysates for small scale isolation of bacteriophage lambda DNA

200 µl of overnight culture were diluted in 33 ml of phage buffer. To each 20 ml universal glass bottle, 2 ml of diluted cells were added. Approximately 1 cm^2 of top agar from each zone of clearing was added to each bottle. These were incubated for 30 minutes at room temperature. 8 ml of NZCYM liquid medium were then added to each bottle and the bottles incubated in shaking incubator at 37°C, until lysis occurred (six to seven hours).

2.13.2.2 Lysates for large scale isolation of bacteriophage lambda DNA

300 µl of overnight culture were diluted in 50 ml of phage buffer. To each 200 ml conical flask, 10 ml of diluted cells were added. Approximately 2 cm^2 of top agar from each zone of clearing was added to each bottle. These were incubated for 30 minutes at room temperature. 40 ml of NZCYM liquid medium were then added to each bottle and the bottles incubated in shaking incubator at 37°C, until lysis occurred (six to seven hours).

2.14 Electrophoresis

2.14.1 Agarose gel electrophoresis

Agarose gels of differing concentrations (0.5-2% w/v) were prepared in 1 × TBE, using multi purpose agarose MP (Boehringer Mannheim) and incorporated ethidium bromide at 4 µg/ml. Electrophoresis was performed at 50-100 volts for 1-3 hours at room temperature. DNA was visualised under ultra violet light .

2.14.2 Acrylamide gel electrophoresis and autoradiography

6% (w/v) acrylamide (19:1 acrylamide:bis-acrylamide; Anachem, supplied at 40%) / 10% glycerol non-denaturing gels (Gel: 20 × 56 × 0.04 cm, 0.5 cm lane width) were prepared in 0.5 or 1 × TBE. Electrophoresis was performed at 30 watts, 1500 volts for 5-6 hours at 4°C. In the cases of ESTs W33210, W36179 and W35449; 37.5:1 acrylamide:bis-acrylamide 6% gels were used. Denaturing gels (Gel: 20 × 56 × 0.04 cm, 0.5 cm lane width), at a final concentration of 6% (w/v) acrylamide (19:1 acrylamide:bis-acrylamide), 7 M urea were prepared in 1 × TBE. Electrophoresis was performed at 50 watts, 2500 volts for 1-2 hours at room temperature. All acrylamide gels were subsequently dried on 3 MM Whatman paper and X-ray film (Fuji XR) exposed for 3 hours to overnight at -70°C, with intensifying screens.

2.15 Generation of sequence data

2.15.1 Generation of ESTs and sequencing of cloned *Mobp* gene fragments

Details of the sequencing procedure utilised in this study have been reported previously

(Stewart et al., 1997). Briefly, single-pass cDNA sequences were obtained using an ABI 373 Stretch automatic sequencer. Dye-terminator cycle sequencing with *Taq* polymerase was utilised throughout. Double-stranded plasmid DNA was prepared from 3 ml overnight cultures, grown in Terrific broth supplemented with 100 mg ml^{-1} ampicillin. DNA was purified by either alkaline lysis followed by PEG precipitation (Perkin Elmer user bulletin 18) or with AGTC centriflex™ miniprep kits (Advanced Genetic Technologies Corporation). Typically, 1/20 of the miniprep DNA was analysed on an 0.8% agarose gel and 250 ng - 1 μg of DNA was used in each sequencing reaction. Clones were selected for sequencing on the basis of an *Mlu* I digest. *Mlu* I releases the entire insert from pSPORT (D'Allessio *et al.*, 1990). Fragments were separated by electrophoresis in 0.8% ordinary agarose gels as described (Section 2.14; chapter 2). Empty clones and those containing non-pSPORT derived plasmid fragments were rejected (5% approx.).

The cycle sequencing protocol was exactly as described in the ABI handbook, protocol P/N 402078 revision A, August 1995. The main primer used, in the generation of ESTs, was an extended version of the standard T7/T3α sequencing primer, (Table 2.3). Oligonucleotide primers utilised in the sequencing of the murine gene *Mobp* and associated intron/exon boundaries are defined in Table 2.3. Oligonucleotide primers were used at 3.2 pmol/20 μl reaction. Up to 650 bp of sequence was obtained, per sequencing reaction, using the methods above.

Oligonucleotide identification	Oligonucleotide sequence 5'-3'	Orientation (sense/antisense)
ASM 101	ATG AGT CAA AGT GGC AAA G	antisense
ASM 102	AGA TAA TAA GGT TCT ATT CAC	sense
ASM 103	CAC AGG AGC CCT GAA TGG	sense
ASM 104	AGG ACT GIG TGA ATC AGG CTC	sense
ASM 105	GAA GGA GGA GCA GGT CAA CC	sense
ASM 106	AGG TGA GGA GGG AGC TCT TC	sense
ASM 107	AAC TGT AGG TAA CTT GIG TC	sense
ASM 108	TAT GTC TGT GGC AAG GAG AGG	sense
ASM 109	CAT CTC TAC ACT TCC TGT GG	sense
ASM 110	AAA GCA CAC AGA TAT GGC CAC	antisense
ASM 111	ATC TTC TAC CAT AGT CAC AG	antisense
ASM 112	TAG GAT GCA CCC ATC CTC G	antisense
ASM 113	CAC TTC ATA CAT CCA CAT G	sense
ASM 114 *	ATT AAG CAA AGC ATT GGA GCG	antisense
ASM 115	GCC GAA GAG GCA TCT CTG TCC	sense
ASM 116	ATC TGG AAA AAC AGA AGC CAC G	antisense
ASM 117	TGA AGC CTC AGG TTC CCA CG	antisense
ASM 118	ACT ACA CAG ATC TAC CTC CG	antisense
ASM 119	CAA AGA GIG ACT GAG AGA ACG	antisense
ASM 120	GCT CAC TTA AAC CTT TCT GCG	sense
ASM 121	TCA GAG CGC CGC CAG CCA AGC	sense
ASM 122*	GAC CTC CCT GTC CTA CAT GG	sense
ASM 123	CTC CTG TCT TCC CTT GTC TGG	sense
ASM 124	CTG TAC TGG GGA GAG GAA	sense
ASM 125	CAC CAT GGC GTC CTC ACA GCC	sense
ASM 126	TTT GGT TCA GTT TTA TGT CC	sense
ASM 127	CAT AGG AAA GCT AAG CGA TGC	sense
ASM 128	TGT GIG AAT CAG GCT CAG GC	sense
ASM 129	AAA TCA ACA TAT CAG TTT TGC	sense
ASM 130	ACA GCT ATG ACC ATG ATT ACG CC	universal
ASM 131	TGT AAA ACG ACG GCC AGT	universal
ASM 132	AGC GGA TAA CAA TTT CAC ACA GGA	universal
ASM 133	AAC ATC AGT CAA GCC CTG AGC	antisense
ASM 134	TAG CTC TGC AAT CAC CTT TGC	antisense
ASM 135	CAA AGC CCT GTG TGA GGC TGG	antisense
ASM 136	CAT CAC ACG CGC ACG CGT ACG	antisense
ASM 137		
ASM 138		
ASM 139		
ASM 141	CAG TAA ATG CAG TAT GAC G	antisense
ASM 149	ATT TTA TTA CTT GTG ACA AGG	antisense
ASM 150		

Table 2.4

Oligonucleotide identification	Oligonucleotide sequence 5'-3'	Orientation (sense/antisense)
ASM 151	CCT CAT AAT TTA TAG GAC CC	antisense
ASM 152	ATC ACC TCT GTG GTG AAG TGT CC	antisense
ASM 153	ACA GAT GTG CGC ACG GTT GC	antisense
ASM 154	TGC TGG GAA CTG AAC TCA GG	antisense
ASM 155	GCT CCC ACA CCA TGG GGA CTG	antisense
ASM 156		
ASM 157		
ASM 158		
ASM 159		
ASM 160	AAT GTT ATA AAC TCC CAT CGG	antisense
ASM 161	ATA ACC AAT ACA GGA CTG TGG	antisense
ASM		
ASM 201 PCR	CCA GGC TCT CCA AGA ACC	sense
ASM 201 PCR	TCA CGC TTG GAG TTG AGG	antisense
ASM 202	CAG TGG ATG CTG AAG TGC	antisense
ASM 203	TAG TCT TCT GGC AGG CAC AGC	antisense
ASM 204	AGC AAC CGC TCT TGC AGA TGC	antisense
ASM 215 F	TCC TGA GTG CTC GTT TGT CTC C	sense
ASM 215 R	CCC ATC AGC TTC AAG CAA TCT C	antisense
ASM 221	ATC TCT TGC CTT TTG TCC	sense
ASM 24		antisense
ASM 241	ATC CTA AGA TTG CTT CTC GC	antisense
ASM B667F	TCT TGT ACA GAA ATG GAA TTG GC	
ASM B667R	TCA AGC TAC CAC GCA TGG	
r-oligo	GGA UCC UAA ACA AUU AAC CCU CAA A	sense
TAG 135	GAG ACG GAT CCT AAA CAA TTA ACC CCT AAA	antisense
MMSV-1F PCR	ATG AGT CAA AAA GTG GCC AAG GAG	sense
MMSV-1R PCR	ACA ATT TCT AGT GTC TTT GTG TTC TGC	antisense
FTG-1F	TCC TGA GTG CTC GTT TGT CTC C	sense
FTG-1R	CCC ATC AGC TTC AAG CAA TCT C	antisense
T7/T3α	ACA GCT ATG ACC ATG ATT AGC C	universal

Table 2.4 Oligonucleotides utilised in the study of the murine *Mobp* gene reported in Section 2

2.16 Sources of DNA samples utilised in genetic mapping studies

Mouse DNAs for the chromosomal allocation experiments were from one of three panels of animals (N2, N4, N6/7). These panels derived from a breeding programme to generate a set of consomic lines with one *Mus spretus* (SEG/Pas) chromosome each on a C57BL/6 background (J-L Guénet, pers. comm.; refer also to Section 1, 1.1). N2, N4 and N6/N7 correspond to DNAs obtained from the first, the third and the sixth combined with the seventh backcross generations respectively. Each generation has been typed for four to six anchor loci per chromosome. The composition of each panel is as follows: DNAs from N6/7, 11 DNAs containing one complete *M.spretus* (SEG/Pas) chromosome on a C57BL/6 background; N4, 19 DNAs, containing well documented fragments of *M.spretus* (SEG/Pas) chromosomes on a C57BL/6 background; N2, 19 DNAs selected from the primary backcross generation. For ESTs shown to be located in that small part of the genome not covered by *M.spretus* DNA in the N4 generation set (chromosome 7 [chr 7], 15-51 cM; chr 8, 4.6-35 cM; chr 9, 31 cM-telomere; chr 10, 46-64 cM; chr 11, 2-37 cM and chr X, centromere-3 cM), chromosomal allocation was finalised using DNAs of animals from the N2 backcross generation. Subset panels of DNAs from the European Collaborative Interspecific Backcross (EUCIB) were used to localise each EST. All DNAs were stored at 20 ng/µl.

2.17 Design of PCR primer pairs and PCR conditions

2.17.1 Polymerase chain reactions used in the mapping of genes expressed in the mouse ventral midbrain

Fragments ranging in size from 150 to 300 bp were generated by PCR amplification of sequence within the putative 3'UTR of ESTs, using primers listed in table 1.3.1 (section 1). Primers were designed with the aid of Gene Jockey II (Biosoft, Cambridge) and synthesised by Cruachem Ltd and Life Sciences, Gibco BRL. For the PCR, primers were at a final concentration of 0.5 pmol/µl, with 20 mM deoxyribonucleotides, 2 ng/µl DNA (mouse genomic) template, 1 mM Mg^{2+}, 1 × buffer (Mg^{2+} free) supplied with Taq polymerase (Promega), 5% deionised formamide, 1 mCi $\alpha^{32}P$ dCTP, and 0.2 units of Taq polymerase in a final volume of 10 µl. Samples were overlaid with oil (Sigma^{Tm}) and amplified in a Perkin Elmer Cetus MK1 thermocycler. After initial denaturation (94°C for 2 min), 25-30 (Table 1.3.1) cycles were performed (denaturation 94°C for 15 sec;

annealing 52-58°C (Table 1.3.1) for 30 sec; extension 72°C for 30 sec), followed by a final elongation step of 2 minutes at 72°C. PCR reactions with each primer pair that did not involve incorporation of a radioactive deoxyribonucleotide (dNTP) utilised dNTPs at a final concentration of 500 μM.

2.17.2 Polymerase chain reaction used in the mapping of *Mobp*

A 205 bp fragment was generated by PCR amplification of sequence within the 3'UTR of a cDNA fragment of the mouse *Mobp* gene, using primers ASM 215 forward and ASM 215 reverse (Table 2.3). PCR cycling conditions for primer pair ASM 215 forward and reverse are defined in Table 2.4.

2.17.3 Polymerase chain reaction used in the oligo-capping procedure

PCR conditions were as follows: gene specific oligonucleotides [Experiment OC I; ASM 201 and 202 and Experiment OC II; ASM 203 and 204] utilised in the experiment are defined in Table 2.4 and were at a final concentration of 0.5 pmol/μl, with 500 μM deoxyribonucleotides, 1 mM Mg^{2+}, 1 × buffer (Mg^{2+} free) supplied with Taq polymerase (Promega) and 0.2 units of Taq polymerase in a final volume of 10 μl. After initial denaturation (94°C for 2 min), 25 cycles were performed (denaturation 94°C for 15 sec; annealing 60°C for 30 sec; extension 72°C for 30 sec), followed by a final elongation step of 5 minutes at 72°C. Experiments OC I and OC II are defined in Section 2 (subsection 2.6).

2.18 Single-strand conformational polymorphism assay

The SSCP assay was performed based on the procedure of Orita *et al.* (1989), with significant alterations. All PCRs for SSCP assays involved the incorporation of a radioactive deoxyribonucleotide (αP^{32}dCTP) and were performed as described in section 2.17 and table 2.3 A 2 μl aliquot of a PCR reaction was diluted with 30 μl of 0.1% SDS, 10 mM EDTA. 2 μl of diluted PCR reaction was then added to 2 μl formamide loading buffer, heat denatured (90°C, 2 min) and 2 μl loaded onto a 6% acrylamide (19:1 acrylamide:bis-acrylamide) / 10% glycerol non-denaturing gel (Gel: 20 × 56 × 0.04 cm,

0.5 cm lane width) in 0.5 × TBE. Electrophoresis was performed as described in Chapter 2 (2.14).

2.19 Software used for data analysis

The sequence data was edited using Sequence Editor (version 1.0.3, ABI) and analysed using Gene Jockey II (Biosoft, Cambridge). Sequence comparisons were performed against all entries in the Genbank databases (performed January 5th-Sept 1st 1997) using the NIH BLAST facility (Altschul *et al.,* 1990) and the EMBL BLITZ facility (protein searches; Sturrock and Collins 1992). Comparisons of the sequences translated in all six reading frames were performed using the TBLASTX program. Sequence analysis and primer design was performed with the aid of Gene Jockey II (Biosoft, Cambridge). Linkage analyses for chromosomal allocation with the N2 consomic generation was performed using Genelink software (Montagutelli, 1988). Backcross panel subsets were selected using the Mbx database for the European Collaborative Interspecific Backcross (EUCIB).

2.20 Radioactive labelling of DNA

2.20.1 Oligonucleotide probes

Synthetic oligonucleotides (oligos) of 18-23 nucleotides in length were radioactively labelled by the addition of a γ-^{32}P rATP molecule to their 5'-hydroxyl termini. Phosphorylation reactions were set up in 10 µl volumes comprising: 5.5 pmol of oligonucleotide, 1 µl γ-^{32}ATP (3,000 Ci/mmol, 10 mCi/ml), 10 U (Weiss units) polynucleotide kinase (PNK) and 1 × PNK reaction buffer (final concentration). Reactions were incubated at 37°C for 1 hour and probes were then purified by passage through G-25 Nick™ (oligos < 20 nt) or G50 (oligos > 20 nt) Sephadex® (Pharmacia Biotech) columns in accordance with the manufacturers' instructions and the incorporation of radioactivity assessed by scintillation counting. All oligo probes utilised in these studies demonstrated radioactive incorporation in excess of 10^8 cpm/µg.

2.20.2 Double-stranded DNA probes

DNA fragments for probes were generated by PCR (MMSV-1F/R, ASM 201F/R, ASM215F/R and ASM 245/6 [Table 2.4]) or by band isolation from agarose gels using the Qiaex II Gel Extract kit, in accordance with manufacturers' instructions. DNA probes were labelled using a protocol based that developed by Feinberg and Vogelstein (Feinberg, 1983), which utilises a mixture of random hexamers in the priming of DNA synthesis. α^{32}dATP (3,000 Ci/mmol, 10 mCi/ml) is incorporated in this synthesis reaction. The reaction was set up using the Ready to go™ DNA labelling kit (Pharmacia Biotech) and was performed in accordance with the manufacturers instructions. Probes were then purified by passage through G50 Nick™ Sephadex® columns (Pharmacia Biotech) in accordance with the manufacturers' instructions and the incorporation of radioactivity assessed by scintillation counting. All double stranded (DS) DNA probes utilised in these studies demonstrated radioactive incorporation in excess of 10^8 cpm/μg.

2.21 Transfer of nucleic acids to nylon membranes via Southern blotting

Agarose gels (0.5-1.5%) were prepared for Southern transfer by successive immersion in 0.5 M NaOH, 1.5 M NaCl (20 minutes); 0.5 M Tris pH 7.5, 1.5 M NaCl (20 minutes) and 20 × SSC for 20 minutes (1 × SSC is 0.15 M NaCl, 0.0015 M sodium citrate). Southern transfer of DNA onto Magna™ nylon membrane (MSI) was performed overnight by capillary action, at room temperature and in the presence of 20 × SSC. The membrane was subsequently rinsed in 6 × SSC and placed at 80°C for 2 hours.

2.22 Screening of genomic libraries

Bacteriophage, plated at a density of ~10^5*pfu*/15 cm (diameter) plate, were transferred, in duplicate, onto Magna™ nylon membrane (MSI). The DNA was fixed to the membranes by successive blotting in: 0.5 M NaOH, 1.5 M NaCl (5 min); 0.5 M Tris pH 7.5, 1.5 M NaCl (5 min) and 2 × SSC for 10 min (1 × SSC is 0.15 M NaCl, 0.0015 M sodium citrate). A 1.6 kb cDNA fragment was used as a probe to screen the bacteriophage plaques, within the λ 2001 and λ PS genomic DNA libraries, for *Mobp* gene sequences. This cDNA fragment was generated by RT-PCR utilising a sense primer (MMSV-1F PCR, defined in Table 2.4) derived from the 5' region of the rat rOP1 cDNA (Yamamoto *et al.*, 1995)

nucleotides 65-88 and an antisense primer (MMSV-1R PCR, defined in Table 2.4) corresponding to nucleotides 1-27 of the cDNA fragment mmsv-1 (Montague *et al.*, 1997a) and was radioactively labelled (~10^8 cpm/μg) as described in section 2.20.

2.23 Nucleic acid hybridisation

2.23.1 Hybridisation of oligonucleotide probes to DNA bound to nylon membranes

The hybridisation solutions for screening with oligonucleotide probes contained: 5 × Denhardt's solution, 5 × SSC, 100 μg/ml Salmon sperm DNA (Boehringer Mannheim), 50 mM sodium phosphate and 1 mM sodium pyrophosphate and 20% v/v formamide. Membranes were incubated at 42°C and subsequently washed, at room temperature, with 2 × SSC, 0.1% SDS. Probes were utilised at 10^6 cpm/ml (hybridisation solution).

2.23.2 Hybridisation of DS DNA probes to DNA bound to nylon membranes

The hybridisation solutions for screening with DS DNA probes contained: 20% w/v dextran sulphate, 5% v/v sarkosyl NL 30 and 10 × SCP buffer (20 × SCP is 2 M NaCl, 0.6 M Na_2HPO_4, 0.02 M EDTA at pH 6.2). Membranes were incubated with DNA probes for 14 - 16 hours at 60°C and washed at 60°C with 0.3 × SSC, 0.5% SDS.

2.24 Identification of overlapping genomic fragments by DNA fingerprinting

Overlapping genomic fragments were identified utilising a DNA fingerprinting assay based on that of Coulson (Coulson *et al.*, 1986), with the following alterations: ~100 ng of DNA from each clone was incubated in a 4 μl reaction volume containing; 0.8 U *Hin*dIII, 0.8 U *Eco*RI, 0.8 U AMV reverse transcriptase (Promega), 0.1 μg RNase, 0.01 μCi αP^{32} dATP (3,000 Ci/mmol), 2 × One Phor-All buffer and in the absence of any dideoxy-nucleotide. The reactions were placed at 37°C for 45 minutes and subsequently placed at 68°C for 30 minutes in order to inactivate the reverse transcriptase. The reactions were then placed on ice. 2 U *Sau*3AI and 1 × React 4 (Promega) added (final concentration) in

a 2 µl volume were added. The reactions were placed at 37°C for 2 hours and stopped by the addition of 4 µl of formamide loading dye. The samples were then placed at 90°C for 5 minutes and subsequently on ice for 2 minutes. The samples were then loaded onto a 6% denaturing acrylamide gel (19:1 acrylamide:bis-acrylamide) and electrophoresed under the conditions described in Chapter 2 (2.14). The procedure is outlined in Figure 2.1

2.25 Detection of transcriptional start sites via the oligo-capping procedure

The transcription start points (tsp) of *Mobp*, were determined using the oligo-capping method (Maruyama and Sugano, 1994) with the following alterations: Poly $(A)^+$ RNA from adult male C57BL6 mice was kindly supplied by Dr T.A.Glencorse and A.B.Roberts (University of Glasgow) and gene specific oligonucleotides (ASM 201, 202, 203 and 204) utilised in the experiment are defined in Table 2.3. The *Bam* HI and r-oligos utilised were as described by Maruyama and Sugano (1994) and are also defined in Table 2.3. The procedure is outlined in Figure 2.2.

2.25.1 Incubation of messenger RNAs with bacterial alkaline phophatase

5-10 µg of mRNA in a 20 µl volume (DEPC-H_2O) were incubated with 10 U of bacterial alkaline phophatase (BAP) in a reaction volume of 100 µl, containing 1 × BAP reaction buffer and 1U RNasin. The samples were placed at 42°C for 1 hour and then an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added. The samples were mixed by vigorous inversion, placed on ice for 5 minutes and centrifuged at 15,000 rpm (4°C) for 5 minutes. The upper phase from each tube was removed to a clean tube. This phenol:chloroform extraction was performed three times. Then 1 µl of glycogen (20 µg/µl), 10 µl 3 M sodium acetate and 250 µl of ethanol were added to the upper (aqueous) phase. The samples were placed at -20°C overnight.

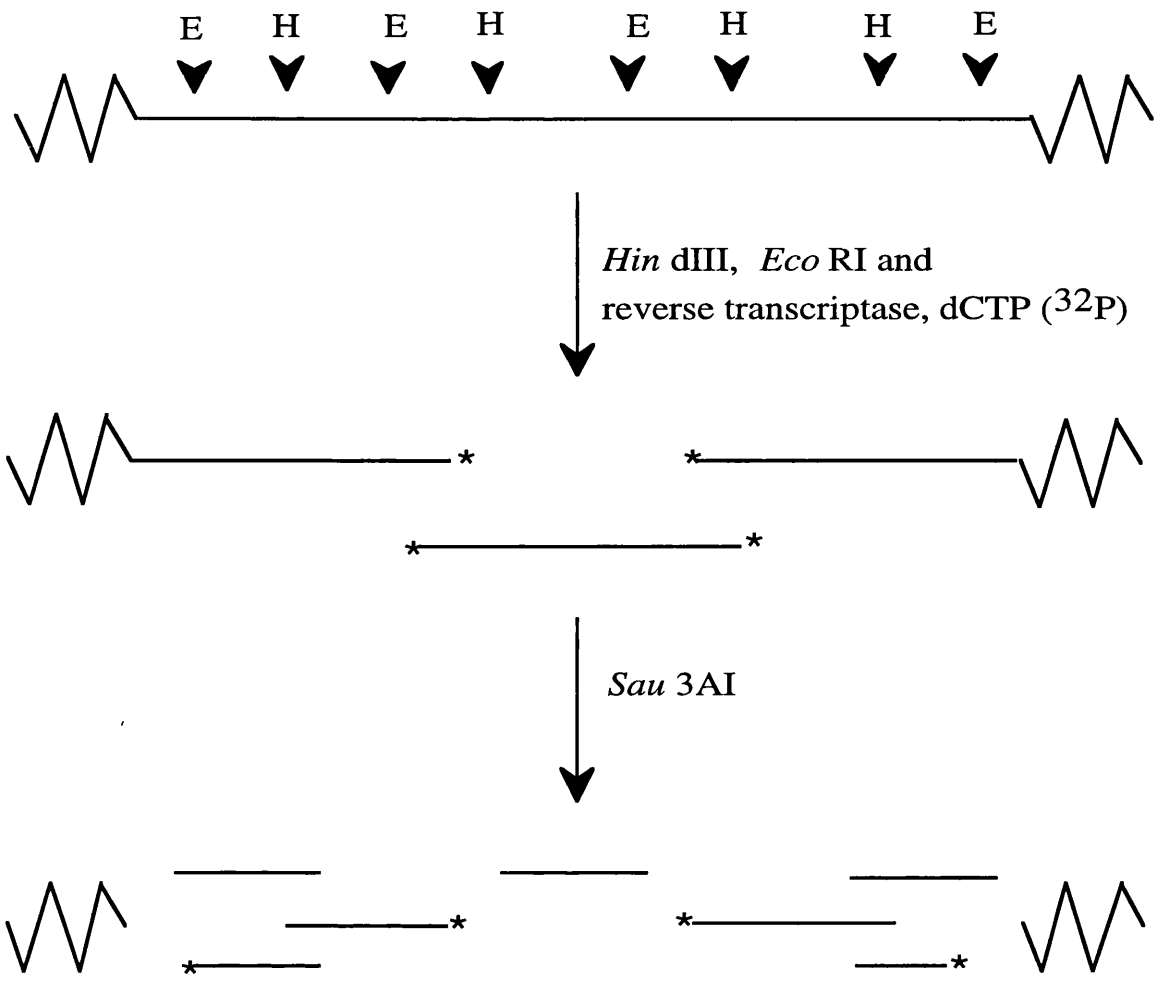


Figure 2.1 Schematic diagram of the DNA fingerprinting procedure. E, *Eco* RI; H, *Hin* dIII sites. *, radioactively labeled DNA fragments and ▼, restriction enzyme sites.

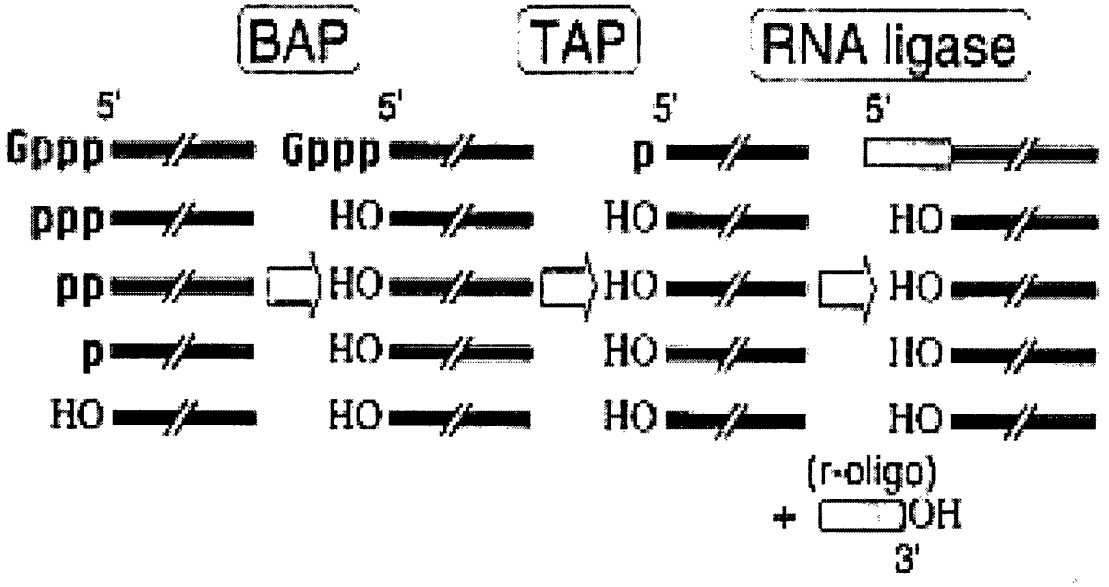


Figure 2.2 Oligo-capping procedure of mRNAs. RNAs are represented as solid bars and r-oligos as shaded bars. Poly (A)+ RNA consists of RNA molecules with various types of 5' ends as shown at the left margin. Gppp:CAP structure; p:phosphate; OH: hydroxyl.

2.25.2 Incubation of messenger RNAs with tobacco acid pyrophophatase

Samples from section 2.25.1 were centrifuged at 15,000 rpm (room temperature) for 10 minutes. The pellets were washed with cold (4°C) 75% ethanol and then centrifuged at 12,000 rpm (room temperature) for 5 minutes. The supernatants were aspirated and the pellets were resuspended in 270 µl of DEPC-H_2O. This RNA was then incubated with 10 U of tobacco acid pyrophophatase (TAP) in a reaction volume of 300 µl, containing 1 × TAP reaction buffer and 1 U RNasin. The samples were placed at 37°C for 1 hour and then an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). The samples were mixed by vigorous inversion, placed on ice for 5 minutes and centrifuged at 15,000 rpm (4°C) for 5 minutes. The upper phase from each tube was removed to a clean tube. This phenol:chloroform extraction was performed three times in total. Then 1 µl of glycogen (20 µg/µl), 30 µl 3 M sodium acetate and 750 µl of ethanol were added to the upper (aqueous) phase. The samples were then placed at -20°C overnight, centrifuged at 12,000 *g* (4°C) for 10 minutes and supernatants were aspirated. The pellets were washed with cold (4°C) 75% ethanol and then centrifuged at 9,000 *g* (4°C) for 5 minutes. The supernatants were aspirated and the pellets were resuspended in 20 µl of DEPC -H_2O.

2.25.3 Ligation of the synthetic RNA-oligonucleotide

The synthetic RNA-oligonucleotide (r-oligo) (Table 2.3) was ligated to the 5' termini of mRNA species (4 µl from section 2.25.2), in the presence of 1 × T4 DNA ligase reaction buffer, 200 U of T4 DNA ligase, 25% PEG, 100 U RNasin, 150 pmol r-oligo. The reaction was performed in a total volume of 100 µl and placed at 16°C overnight. 200 µl of DEPC-H_2O and 300 µl of phenol:chloroform:isoamyl alcohol (25:24:1) were added. The samples were mixed by vigorous inversion, placed on ice for 5 minutes and centrifuged at 15,000 rpm (4°C) for 5 minutes. The upper phase from each tube was removed to a clean tube. This phenol:chloroform extraction was performed three times in total. Then 1 µl of glycogen (20 µg/µl), 30 µl 3 M sodium acetate and 750 µl of ethanol were added to the upper (aqueous) phase. The samples were then placed at -20°C for 1 hour, centrifuged at 12,000 rpm (4°C) for 10 minutes and supernatants were aspirated. The pellets were washed with cold (4°C) 75% ethanol and then centrifuged at 12,000 rpm (4°C) for 5 minutes. The supernatants were aspirated and the pellets were resuspended in 100 µl of DEPC-H_2O. 100 µl 5 M ammonium acetate and 500 µl ethanol were added, mixed by inversion and placed on ice for 1 hour. The samples were then centrifuged at

12,000 rpm (4°C) for 10 minutes and supernatants were aspirated. The pellets were washed with cold (4°C) 75% ethanol and then centrifuged at 12,000 rpm (4°C) for 2 minutes. The supernatants were aspirated and the pellets were resuspended in 12 µl of DEPC -H_2O.

2.26 Tissue culture

2.26.1 Embryonic stem cells

The embryonic stem (ES) cell line CGR 8.8 was a kind gift of Dr Bill Skarnes (Medical Research Council, Centre for Genome Research, Edinburgh).

2.26.2 Preparation of tissue culture flasks and plates

All tissue culture flasks and plates were gelatinised as follows: The growth surface of each receptacle was overlaid with a 0.1% gelatin solution and placed at room temperature for at least 15 minutes prior to use. The excess gelatin solution was aspirated immediately prior to use.

2.26.3 Cell culture

The medium used for cell culture in this study contained: 1 × minimal essential medium (MEM), 0.25% sodium bicarbonate, 1 × non-essential amino acids, 1 mM sodium pyruvate, 2 mM Glutamax™ (Gibco BRL, Life Technologies), 0.77% v/v β-mercaptoethanol, 10% S-serum (Foetal calf serum, Sigma), 100 µg/ml penicillin/streptomycin and 2×10^4 U/ml Leukaemia Inhibitory Factor (LIF). 1 ml of medium was added to a cryovial containing CGR 8.8 cells, thawed rapidly at 37°C. The cells were dispersed by gentle pipetting and removed to a gelatinised 5 cm^2 tissue culture flask. The volume was increased to 5 ml by further addition of medium. The flask was then placed at 37°C in the presence of 5% CO_2 overnight. The medium was then aspirated and replaced with 7.5 ml of fresh medium and placed at 37°C in the presence of 5% CO_2 for 24 hours. The medium was then aspirated and the cells washed twice with 7 ml of 1 × PBS. 1 ml of TVP (0.025% Trypsin, 1% v/v chick serum, 1 mM EDTA and 1 × PBS) was then added to the flask and the flask agitated until the cells were observed to release from the flask (normally 3-4 minutes). 5 ml of medium were added to the flask and the cells

transferred to a 50 cm^2 flask. This cycle of feeding and trypsinising was repeated twice over the ensuing four day period; the cells being transferred to 75 cm^2 and 150 cm^2 flasks in succession.

2.26.4 Transfection of CGR 8.8 embryonic stem cells with DNA

When the cells were 40-50% confluent, in the 150 cm^2 flask, the medium was aspirated and the cells washed twice with 50 ml of 1 × PBS. 2 ml of TVP was then added to the flask and the flask agitated as above. 8 ml of medium was then added to the flask and the cells transferred to a 15 ml polypropylene centrifuge tube. The cells were then centrifuged 1,000 rpm for 1 minute (Jouan, CR 312), the supernatant aspirated and the cells gently resuspended in 10 ml of 1 × PBS. This centrifugation and PBS wash was repeated twice. The supernatant was aspirated, the cells were gently resuspended in 0.8 ml of 1 × PBS, and 50 µl of linearised DNA (p*Mobp*-N/TK; defined in Section 2, 2.6 and illustrated in Figure 2.6.4) was added. The cells and DNA were transferred to a 0.4 cm electroporation cuvette (Biorad, Gene Pulser). Electroporation was performed at 0.8 kv, 600 Ω; with a time constant of 0.1. The transfected cells were placed at room temperature for 20 minutes and then added to 30 ml of medium contained in a 50 ml polypropylene centrifuge tube and mixed by gentle pipetting. 2 ml aliquots were added to each of 15 tissue culture plates (10 cm diameter) and the volume completed by the addition, to each, of 8 ml of medium. The plates were then placed at 37°C, in an atmosphere containing 5% CO_2 overnight. The medium was then aspirated and replaced by the addition of 10 ml of medium containing G418 (0.2 mg/ml) and Gancyclovir (0.55 µg/ml). N.B. one plate was identified as a control and was fed a medium lacking Gancyclovir. The medium was replaced with 10 ml of fresh medium containing G418 (0.2 mg/ml) and Gancyclovir (0.55 µg/ml) on a daily basis, over a period of 14 days. By day eight distinct colonies, resistant to the drugs used in this study, could be observed with the naked eye.

2.26.5 Picking drug resistant colonies

The medium was aspirated and the cells washed twice with 10 ml of 1 × PBS. The plates were then overlaid with 10 ml of 1 × PBS. Using a pipette tip (2-200 µl), colonies were gently agitated and removed from the plates in 100 µl volumes of 1 × PBS and each placed in 24 well (16 mm) tissue culture plate. The cells were dispersed by gentle pipetting. 100 µl of TVP were then added to each well and the plate gently agitated. After a period, not

in excess of 5 minutes, 2 ml of medium was added to each well. The plates were then placed at 37°C, in an atmosphere containing 5% CO_2 for 48 hours. The medium was then aspirated and 1 ml of fresh medium added. This was repeated on a daily basis until the a confluent layer of cells was obtained.

2.26.6 Harvesting of selected colonies

The selected colonies were allowed to grow to confluence. Colonies, having reached this stage, were treated as follows: 1 ml of fresh medium was added to each colony 1-2 hours prior to preparation for freezing. The medium was aspirated and the colony washed twice with 2 ml of 1 × PBS and the PBS subsequently removed. 100 µl of TVP were then added to each well and the plate gently agitated. After a period, not in excess of 5 minutes, 1 ml of freezing medium (9:1 [vol:vol] medium:DMSO) was added to each well. 0.4 ml from each well was placed in a 1.2 ml cryovial for storage under liquid nitrogen. The remaining 0.6 ml was placed in a 1.5 ml centrifuge tube for DNA isolation as described in Chapter 2 (2.8.4).

SECTION 1

1.1 Introduction

1.1.1 Choice of the single-strand conformation polymorphism (SSCP) assay for genetic mapping of mouse ESTs

The importance of efficient and reliable detection of single base substitutions is widely accepted within biomedical research. In recent years SSCP has emerged from the range of available options as a convenient technique due to its sensitivity and ease of use. I have investigated the use of SSCP in generating genetic mapping markers from mouse expressed sequence tags (ESTs) which are short sequences generated by single-pass end sequencing of cDNAs. In an endeavour to establish a rapid genetic mapping strategy for genes expressed in the CNS of mouse I have been concerned to maximise the frequency with which SSCPs, based on the limited EST sequence available, are informative i.e. show detectable differences between *Mus spretus* and C57BL/6. SSCP assays are compared below with the range of most frequently used techniques.

1.1.1.1 Review of available technologies

Since the establishment of restriction fragment length polymorphism (RFLP) analysis (Botstein *et al.*, 1980), the range of available options for detection of single base substitutions / mutations has increased dramatically. The range of the most frequently used techniques now includes: RFLP analysis (Botstein *et al.*, 1980; Jiang *et al.*, 1989; Takahashi and Ko, 1993); allele specific oligonucleotide (ASO) hybridisation (Wallace *et al.*, 1981; Conner *et al.*, 1983; Saiki *et al.*, 1988); heteroduplex assays (White *et al.*, 1992); mismatch cleavage assays (RNase and chemical cleavage) (Myers *et al.*, 1985; Winter *et al.*, 1985; Cotton *et al.*, 1988); sequencing of PCR products (Murray, 1989); denaturing gradient gel electrophoresis (DGGE) (Fischer and Lerman, 1980; Myers *et al.*, 1985; Lerman and Silverman, 1987; Myers, 1987; Noll and Collins, 1987; Sheffield *et al.*, 1989) and single strand conformation polymorphism (SSCP) analysis (Orita *et al.*, 1989a; Orita *et al.*, 1989b; Suzuki *et al.*, 1990; Okamato *et al.*, 1991; Spinardi *et al.*, 1991; Beier *et al.*, 1992; Glavac and Dean, 1993; Hayashi and Yandell, 1993; Kupryjanczyk *et al.*, 1993; McCallion *et al.*, 1996).

RFLP has proved itself to be exceptionally useful in situations where no sequence information is available (Jiang *et al.*, 1989; Silver, 1995). RFLPs are detected at high frequency between *Mus* species (Avner *et al.*, 1988). However, the cost of utilising the battery of enzymes (Silver, 1995) which may be required to screen many different loci in

search of polymorphisms can be prohibitive. ASO hybridisation is a sensitive technique which detects genetic variation over short lengths (1-20 nt) in PCR products (Wallace *et al.*, 1980; Conner *et al.*, 1983; Saiki *et al.*, 1989; Hayashi and Yandell 1993; Silver, 1995) but is limited by the short length over which it is effective. Heteroduplex assays (White *et al.*, 1992) are based on the difference in electrophoretic mobility of matched and unmatched heteroduplexed PCR products. However, few published data were available on the sensitivity of the technique (Hayashi and Yandell 1993). Mismatch cleavage is a sensitive technique (Winter *et al.*, 1985) which involves the identification and cleavage of DNA, at the point of heteroduplex mismatch, by chemical or enzymatic means. However, cleavage utilising RNase has demonstrated insensitivity to certain mismatches (Meyers *et al.*, 1985) and the more sensitive chemical cleavage techniques require expertise in the performance of the biochemical steps involved (Cotton *et al.*, 1988). Sequencing of PCR products is exceptionally sensitive but has historically proved to be labour intensive and expensive when used to genotype large numbers of DNAs are clearly prohibitive. The development of automated sequencing technologies means that this need not necessarily be the case. DGGE is based on the fact that the mobility of a partially melted DNA duplex is virtually zero and that the site of mismatch tends to melt early (Fischer and Lerman, 1980; Meyers *et al.*, 1985; Noll and Collins, 1987). DGGE is very sensitive but requires special apparatus to control temperature, long primers, frequently entailing the addition of a 30-40 nt GC clamp (Sheffield *et al.*, 1989), and is labour intensive (Hayashi and Yandell 1993). However, the theoretical basis for the detection of polymorphisms with DGGE has been reported (Meyers *et al.*, 1987). No such model exists for the 3-dimensional (3-D) structure of single strands under given conditions (e.g. in SSCP analysis). Maxam and Gilbert (1975) suggested that intramolecular interactions beyond Watson and Crick pairing are important determinants of folding. The 3-D structure of single stranded DNA molecules is more complex than a simple stem loop or secondary structure as determined by X-ray crystallography [Rich, 1976 #810]. SSCP does, however, readily lend itself to empirical optimisation (Hayashi and Yandell 1993), the formation of higher order structures being dependent on length, sequence, and physical factors (e.g. temperature and ionic strength, Hayashi and Yandell, 1993)

In native gels conformations of single strand DNA molecules are stabilised by intrastrand interactions (Orita *et al.*, 1989[a,b]). Therefore, conformation and mobility are dependent on sequence composition; a mobility shift reflects a conformational change. It has also been suggested that more than one metastable conformation is available to a single stranded DNA molecule (Orita *et al.*, 1989[a,b]; Beier *et al.*, 1992; Glavac and Dean 1993;

Hayashi and Yandell 1993). It noteworthy, however, that the patterns provided by SSCP assays are seldom ambiguous (Beier *et al.*, 1992). The SSCP assay has developed into a highly sensitive, convenient technology (Orita *et al.*, 1989[a,b]; Suzuki *et al.*, 1990). It does not require restriction enzyme cleavage or hybridisation and lends itself to PCR amplification (Orita *et al.*, 1989[b]) and strand separation on thin polyacrylamide gels (Orita *et al.*, 1989[a,b]; Suzuki *et al.*, 1990; Beier *et al.*, 1992; Beier, 1993; Glavac and Dean 1993; Hayashi and Yandell 1993; Beier *et al.*, 1995; McCallion *et al.*, 1996). Dissociated PCR fragments up to 600 nt in length can be separated without loss of resolution (Ludeke *et al.*, unpublished observations; cited in Hayashi and Yandell, 1993). Detection levels as high as 100% have been reported (Michaud *et al.*, 1991).

1.1.1.2 Selection of SSCP as the method of choice for mapping expressed sequences

Genetic mapping requires the detection of allelic differences between parental DNA sequences. In a recent report by Silver (1995), the author cites five criteria for an ideal polymorphism detection protocol. These are: (1) detection of any and all basepair variants in a DNA region as multiple alleles; (2) no requirement for sequence information from each allele; (3) no requirement for synthesis of ASOs; (4) no requirement for special skills or equipment; (5) rapidity and ease of result reproduction.

The conformation and mobility of a single stranded DNA molecule in a native gel are dependent on sequence composition (Orita *et al.*, 1989[a,b]). In SSCP analyses a sequence polymorphism between single stranded DNA molecules frequently results conformational change is reflected in the subsequent detection of a polymorphism. SSCP may be combined with PCR amplification and consequently requires sequence from only one allele and does not require the generation of ASOs. SSCP assays may be performed using standard molecular biology technologies (PCR amplification in combination with sequencing type gel electrophoresis) and as such does not require any skills or equipment not normally found in molecular biology laboratories. SSCP has been demonstrated to be efficient and reproducible (Orita *et al.*, 1989[a,b]; Suzuki *et al.*, 1990; Beier *et al.*, 1992; Beier, 1993; Hunter *et al.*, 1993; Beier *et al.*, 1995; McCallion *et al.*, 1996; McCallion *et al.*, in prep.). Silver concludes that all of these criteria are satisfied to a "good degree" in the form of SSCP analysis.

1.1.2 Generation and use of interspecific backcross and consomic panels in the construction of genetic maps

1.1.2.1 Genetic maps

A genetic map is a representation of the distribution of a set of loci within a genome (Silver, 1995). Loci included in a mapping project may be related according to any number of parameters e.g. functional or structural homologies, a predetermined chromosomal assignment or expressed sequences deriving from a common tissue source or developmental timepoint. Mapping of any locus may be accomplished at any one of a number of levels of resolution; from chromosomal assignment to ever increasing precision in subchromosomal localisation and linkage to known loci and ultimately to mapping loci to a DNA sequence. Three types of genetic map may be derived; linkage, chromosomal and physical. These maps are distinguished by the methods used for their generation and the metric used for establishing distances within them (Silver, 1995).

The linkage (recombination) map was the first to be developed subsequent to the rediscovery of Mendel's work early in this century. The construction of a genetic linkage map is a fundamental step in the structural and functional characterisation of a mammalian genome (Copeland and Jenkins, 1991). This map form may be generated for loci existing in two or more allelic (heritable) forms. Thus, monomorphic loci may not be mapped in this fashion. Linkage maps are generated by comparing the number of offspring receiving a parental or recombinant allele combination from a parent carrying two different alleles at two or more loci. Analysis of this type of data permits the establishment of whether or not two loci are genetically linked. Chromosomal assignment is normally accomplished when a new locus is found to be linked to a previously assigned locus (Copeland and Jenkins, 1991; Silver, 1995). The linkage map is the only map based on classical breeding analysis. Distances are measured in centimorgans (cM; 1 cM = 1% crossover rate).

Chromosome (cytogenetic) maps are based on karyotype information. All mouse chromosomes are cytogenetically defined according to their size and banding pattern. Map positions are defined using band names. Chromosome maps may now be generated using somatic cell hybrid lines, containing portions of the mouse karyotype in a background genome derived from another species. Correlation of the presence or expression of a mouse gene with the presence of a mouse chromosome or chromosomal region permits chromosomal allocation. Other approaches utilising locus specific probes for *in situ* hybridisation (ISH) or correlation of cytogenetic abnormalities with mutant phenotypes are less accurate (Avner *et al.*, 1988; Silver, 1995) and require experience in

analysing mouse chromosomes. All mouse chromosomes are telocentric and are consequently not as readily distinguished from each other, at the cytogenetic level as human (Avner *et al.*, 1988; Silver, 1995). The distinction between chromosome and linkage maps is now perhaps more historical than real, linkage groups coalesce to chromosomes.

The physical map is based on direct analysis of DNA and is measured in base pairs, bp; kilobase pairs, kbp and megabase pairs, Mbp. Arbitrary distinction is made between short range (< 30 kbp) and long range (> 30 kbp) physical maps. Detection of the same large restriction fragment by probes unique for two different loci establishes physical linkage. In the future long range physical maps consisting of overlapping clones (Yeast artificial chromosome, YACs; P1 and Cosmid clones) will be established, covering each chromosome in the mouse genome.

This study will primarily restrict itself to the generation of linkage maps but will acknowledge the value of correlating map positions on linkage and physical maps (Levitt, 1991; Polymeropoulos *et al.*, 1992, 1993; Silver, 1995). The benefit of correlating map positions of loci in both linkage and physical map forms, to those undertaking physical mapping projects, is quite clear. The ability to correlate an interval (physical or linkage), associated with a pathology, with expressed genes provides strong candidate genes for further study (Burke *et al.*, 1987; Levitt, 1991; Polymeropoulos *et al.*, 1993).

1.1.2.2 Generation of backcross mapping panels

Most breeding experiments begin with an outcross; mating two animals or strains considered unrelated to each other. For many experiments the unrelated strains are two different inbred strains. Inbred strains are considered to be homozygous across their entire genome. Thus an outcross can be symbolised as $A_{i\text{-}j}/A_{i\text{-}j} \times a_{i\text{-}j}/a_{i\text{-}j}$ (where A and a are two alleles of the same locus and the number of loci present is in the range i-j), and the offspring resulting from such a cross (first filial generation), symbolised by F1.

All F1 animals are identical and heterozygote ($A_{i\text{-}j}/a_{i\text{-}j}$) throughout their genomes. The F1 animals obtained from this cross can then be mated with animals from a parental strain that are homozygous ($A_{i\text{-}j}/A_{i\text{-}j}$ or $a_{i\text{-}j}/a_{i\text{-}j}$) throughout their respective genomes. Such a cross is termed a backcross. The two generation outcross-backcross combination is one of the major breeding protocols used in linkage analysis (Avner *et al.*, 1988; Copeland and Jenkins, 1991; Breen *et al.*, 1994; Silver, 1995; Section 1.1.2). Mendel's first law predicts

that offspring from a backcross to a/a will be distributed in roughly equal proportions between the two genotypes, at any one locus; 50% A/a and 50% a/a. Until recently genetic mapping in mouse has been reliant on the analysis of 2 and 3-point crosses established between genetically non-identical laboratory strains or recombinant inbred strains. However, these analyses are limited by the degree of allelic difference between laboratory strains (Copeland and Jenkins, 1991). These limitations have recently been overcome by the use of interspecific crosses, exploiting the genetic diversity inherent among wild species (Avner *et al.*, 1988). One of the most distantly related *Mus* species that may interbreed with laboratory strains, and still provide fertile F1 hybrid animals, is *Mus spretus* (Avner *et al.*, 1988). Backcrosses between *Mus spretus* and common laboratory strains have become the method of choice in the construction of multilocus linkage maps of the mouse genome (Copeland and Jenkins, 1991). C57BL/6 mice are frequently chosen as the laboratory strain in a backcross breeding program because they are so well genetically characterised (Copeland and Jenkins, 1991). The European Collaborative Interspecific Backcross (EUCIB; Breen *et al.*, 1994) was generated in this way; utilising *Mus spretus* (SPR and SEG/Pas) and *Mus domesticus* (C57BL/6) inbred strains. It has established the largest and most precise purpose bred genetic mapping resource, for the mouse genome, in the world (Breen *et al.*, 1994). The EUCIB resource was utilised in the genetic mapping of ESTs expressed in the ventral midbrain of mouse, reported in these studies (Section 1, 1.3-1.4).

This kind of approach is limited by a number of factors. Firstly, laboratory strain × *Mus spretus* F1 males are sterile (Avner *et al.*, 1988; Copeland and Jenkins, 1991). As a consequence only recombination data from F1 females can be obtained, and recombination in the X-Y pseudoautosomal region cannot be established (Copeland and Jenkins, 1991). Secondly, the approach is limited by the amount of DNA available from each animal in the cross. This will limit the number of probes or markers that can be localised in a DNA panel. However, the advent of PCR based mapping strategies (for example: Orita *et al.*, 1989[a,b]; Sheffield *et al.*, 1989; White *et al.*, 1992), means that DNA quantity is no longer a major limiting factor.

1.1.2.3 Generation and utility of consomic strains

A consomic strain is one in which one entire chromosome from one strain is transferred onto the genetic background of another strain. This is in contrast to congenic strains in which only a chromosome region is transferred. The transfer is achieved by iterative backcrossing with the strain providing the genetic background (Silver, 1995; Guénet, pers.

comm.). Consomic strains are normally completed after a minimum of ten backcross generations (Silver, 1995; Guénet, pers. comm.). In most studies the chromosome to be made consomic is the Y.

However, Guénet (Institut Pasteur) has recently initiated a backcross breeding project to transfer each individual chromosome from the *Mus spretus* (SEG/Pas) genome onto a C57BL/6 background. Each generation is screened using a battery of PCR microsatellite markers (Guénet, pers. comm.). Genotyping of DNAs from each generation, with a battery of molecular markers, has permitted the use of intermediate generations in chromosomal allocation (Guénet, pers. comm.; refer also to discussion in Section 1, 1.3). In this study (Section 1, 1.3) the first backcross generation (N2), the third backcross generation (N4) and the sixth and seventh backcross generations (N7/8) provided informative results, permitting the allocation of ESTs to mouse chromosomes. This breeding programme also establishes a stable and unlimited resource for the immediate chromosomal allocation of new markers (Guénet, pers. comm.). It was envisaged that the use of a single panel of DNAs from consomic strains in the allocation of novel markers to mouse chromosomes would dramatically reduce the labour involved in chromosome allocation utilising whole genome scanning. It is noteworthy that radiation hybrid panels for mapping the mouse genome were unavailable at the inception of these studies. This study (Section 1) reports the use of both consomic and EUCIB DNA panels in combination with a PCR based SSCP assay in the establishment of a rapid genetic mapping strategy for mouse ESTs.

1.2 Establishment of an SSCP assay

1.2.2 Analysis of parameters affecting SSCP sensitivity

A wide range of techniques, for the detection of allelic differences between, DNA sequences is available, but single-strand conformation polymorphism (SSCP) analysis of PCR products (Orita *et al.*, 1989[a]; Orita *et al.*, 1989[b]; Suzuki *et al.*, 1990; Beier *et al.*, 1992; Beier, 1993; Beier *et al.*, 1995; McCallion *et al.*, 1996) has come to be widely used because of its rapidity, sensitivity and ease of use (refer also to introduction of Section 1, 1.1.1). In order to make optimal use of SSCP in the mapping of large numbers of mouse ESTs, for which limited sequence exists, it is essential to achieve high sensitivity levels for polymorphism detection with the minimum of optimisation work for each EST. A number of parameters, including temperature, gel concentration, median pore size and the concentration of additives such as glycerol, have been shown to affect the detection of known mutations using SSCP (Glavac and Dean 1993; Hayashi and Yandell 1993; Spinardi *et al.*, 1993). Most studies have, thus far, reported mutation detection in coding regions (Orita *et al.*, 1989[a]; Orita *et al.*, 1989[b]; Suzuki *et al.*, 1990; Okumato *et al.*, 1991; Beier *et al.*, 1992; Glavac and Dean 1993; Hayashi and Yandell 1993; Kupryjanczyk *et al.*, 1993; Spinardi *et al.*, 1993). In contrast, the ESTs utilised in this study (Stewart *et al.*, 1996) derive primarily from the 3'UTRs of transcribed sequences, which are frequently more AT rich than coding regions (Stewart *et al.*, 1996). Though the role of sequence composition in the detection of SSCPs had been acknowledged (Orita *et al.* 1989[a,b]; Glavac and Dean 1993; Hayashi and Yandell 1993), it was not known how, if at all, the sequence bias within 3'UTRs would affect the detection of SSCPs. Neither the utility of nor parameters for the detection of SSCPs in 3' directed DNA fragments had been reported at the time that this study was initiated.

It was not practicable to adjust all of the previously assessed parameters for each of a large number of ESTs. This series of experiments demonstrates that a simple laboratory protocol using two gel reaction mixes providing different median pore sizes and otherwise standard running conditions allows allelic differences between *M.spretus* and C57BL/6 EST sequences to be detected very efficiently. This study compared the utility of gel reaction mixes with two different acrylamide:bis-acrylamide ratios (19:1 and 37.5:1), in SSCP detection. All results reported are consistent upon repetition.

1.2.2.1 Comparison of cross linking densities between acrylamides SSCP assays

In order to examine the effect, on the frequency of SSCP detection, of altering median pore size within a native acrylamide gel, two gel matrices were used (19:1, acrylamide:bis-acrylamide and 37.5:1, acrylamide:bis-acrylamide). Thirteen PCR products (150-300 bp) from putative 3' untranslated (3'UTR) sequences of thirteen different mouse brain EST sequences were chosen at random for this study. EST sequences were obtained by single-pass sequencing of cDNA clones from a mouse brain-region specific cDNA library produced by oligo-dT priming (Davies *et al.*, 1994). DNA sequence from the 3'UTR demonstrates a high frequency of base mismatches between the mouse subspecies most frequently used in genetic mapping (Takahashi and Ko, 1993). Only eight EST-based PCR products gave distinct SSCPs between *M.spretus* and C57BL/6 in the 19:1, acrylamide:bis-acrylamide (19:1), 6% polyacrylamide gels that are in widespread use (Table 1.2.1). However, no fewer than four out of the five that failed to give distinct SSCPs did provide resolvable SSCPs in 37.5:1 acrylamide:bis-acrylamide (37.5:1), 6% gels (Figure 1.2.1; Table 1.2.1). Thus the frequency of detecting informative SSCPs can be raised significantly by testing systematically with 37.5:1 as well as 19:1. For the eight distinct polymorphisms detected on 19:1 gels, running in 37.5:1 acrylamide:bis-acrylamide did not improve the clarity of the result, and in two instances (W33249 and W33240) did not provide SSCPs. The SSCPs demonstrated, between *M.spretus* and C57BL/6, by PCR amplified sequences from within these ESTs are illustrated in Appendix I. N.B. The utility of 60:1 gels was also examined but demonstrated no of further enhancement in resolution or frequency of detection of SSCPs, in 3'UTR, between the mouse species examined.

1.2.2.2 Comparison of ionic strength of electrophoretic buffer in SSCP assays

It is also interesting to note that the SSCP resulting from one pair of PCR primers (EST W33210) was dramatically further enhanced in both 19:1 and 37.5:1, 6% gels, by running with 1 × TBE buffer compared with running in 0.5 × TBE buffer (Figure 1.2.2). This contrasts with previous reports that use of 0.5 × TBE frequently improves resolution (Orita *et al.*, 1989). However, I still recommend the use of 0.5 × TBE as a standard running buffer; for all other amplified EST sequences used in this study 0.5 × TBE provides an equal or enhanced resolution compared with 1 × TBE. All other conditions used (Figure

EST identification	Oligonucleotide sequences 5' - 3'	PCR product size (bp)	SSCP 6%, 19:1	SSCP 6%, 37.5:1
W33210	AAC GCA CTG CTG CCA TCT AGT G CGT AAG CCC TTA GAA TGC CAG AC	249	-	+
W36179	GGT CCG TTT CCA CAA TTC AGG ACT ACT AAC TAG CAG CGA CCA GAC CTG	184	-	+
W36203	TTA TCG CTC TGC ATG TGT GGA G GGT CCT GTG CAT TCA ATG TGA AC	243	+	+
W33249	CAT CCG AGT GAC AAC AGA CAA AAG CTT CCC AAA GTC AAA TGA GTC TC	259	+	-
W36181	GCC TTC AGG ATC TTC TCA CAT GAG CCA TTT GTG GGA GCG TTG TC	199	+	+
W36200	CCA TGA AGA AAT AGA TAG CAA GGT G AAG CCT GGG AAA GAT ACT GGA C	152	+	+
W35430	CCA ATT ATT TGT ATC AAT CAG GAC C CGG GAA AGA GAG GTA TGG CTA G	259	+	+
W39947	CTG AAA CAT TAA ATG GAG GTC CCC CCA GTG AGA GAA CAA ATC CTG ACC	160	-	+
W35449	CAT GGT CTT TCC TTC CAT CAG G GAA GCA GAG AGC TAT AAG CCT TCT G	221	-	+
W33240	TTC TCG CTG TCT ATG ACT TAG CAG G TTT GAC CAG GAA TCT GAA ATT AGG AG	291	+	-
W36223	CTG AAA CAT TAA ATG GAG GTC CCC CCA GTG AGA GAA CAA ATC CTG ACC	200	-	-
Z36363	TCC TGA GTG CTC GTT TGT CTC CC CCC ATC AGC TTC AAG CAA TCTC	205	+	-
AA008736	TCG AGA TCA TTG TCT GAA AGT ACA TC TCC CGT TTG GAA AAG TTG G	272	+	+

Table 1.2.1 Detection of allelic differences by SSCP within PCR products from *M.spretus* C57BL/6 DNAs based on EST sequence data. +, detectable SSCP; -, no detectable SSCP. Appendix I illustrates the most distinct SSCPs obtained, for each of the amplified sequences above, in this study.

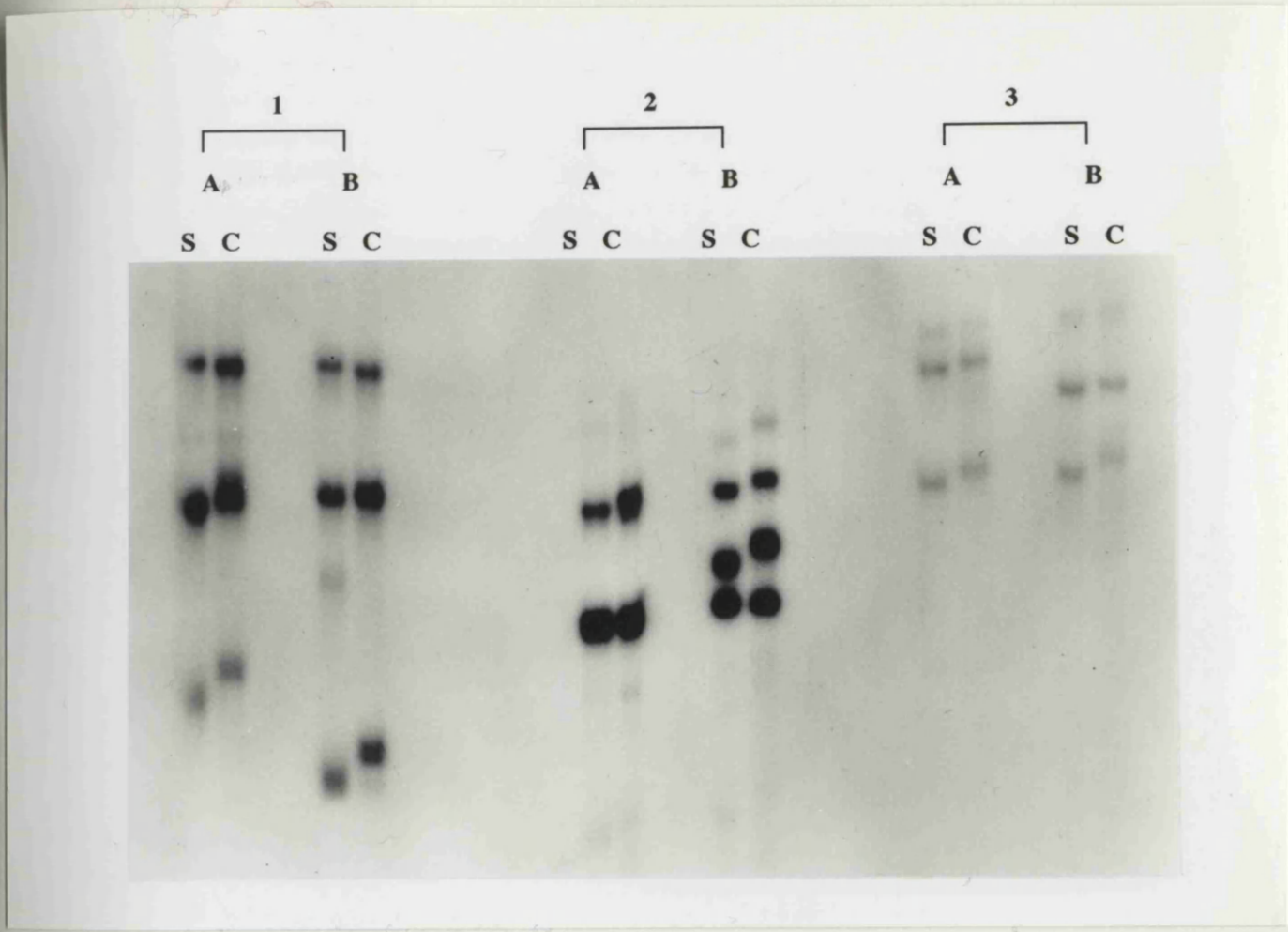


Figure 1.2.1 The effect of different acrylamide:bis-acrylamide ratios on electrophoretic mobility in the SSCP assay of sequences amplified using PCR primers designed to putative 3'UTRs of 3 mouse brain-derived ESTs 1, 2 and 3 (W39947, W35449 and W36179 respectively). Templates are defined as follows: S, *Mus spretus* (substrain SEG/Pas) and C, C57BL/6. Running conditions are the following: A, 0.5 × TBE, 4°C, 10% glycerol and 19:1 acrylamide:bis-acrylamide; B, 0.5 × TBE, 4°C, 10% glycerol and 37.5:1 acrylamide:bis-acrylamide.

Figure 1.2.2 The effect of different acrylamide:bis-acrylamide ratios and buffer concentrations on electrophoretic mobility in the SSCP assay of sequences amplified using PCR primers designed to putative 3'UTRs of a mouse brain-derived EST (W33210). Templates are defined as follows: S, *Mus spretus* (substrain SEG/Pas) and C, C57BL/6. Running conditions are the following: A, 0.5 × TBE, 4°C, 10% glycerol and 19:1 acrylamide:bis-acrylamide; B, 0.5 × TBE, 4°C, 10% glycerol and 37.5:1 acrylamide:bis-acrylamide; C, 1 × TBE, 4°C, 10% glycerol and 19:1 acrylamide:bis-acrylamide; D, 1 × TBE, 4°C, 10% glycerol and 37.5:1 acrylamide:bis-acrylamide.

1.2.1) are standard SSCP conditions based on the combined experience of many groups (Orita *et al.*, 1989; Suzuki *et al.*, 1990; Beier *et al.*, 1992; Glavac and Dean 1993; Hayashi and Yandell 1993; Spinardi *et al.*, 1993).

1.2.3 Discussion

This study reaffirms the suggestion that the ratio of acrylamide:bis-acrylamide plays an equal if not more important role than gel concentration in the resolution of SSCPs (Glavac and Dean 1993; Hayashi and Yandell 1993). These data would further indicate that simply using two levels of crosslinker percentage with otherwise standardised SSCP conditions (Orita *et al.*, 1989[a,b]; Suzuki *et al.*, 1990; Beier *et al.*, 1992; Beier, 1993; Glavac and Dean 1993; Hayashi and Yandell 1993; Beier *et al.*, 1995) allows allelic differences within short DNA stretches of *M.spretus* and C57BL/6 mouse lines to be detected efficiently. Over 92% of the ESTs gave an informative SSCP with these mouse lines from the first PCR product, compared with 62% for a standard SSCP assay utilising a single median pore size. This enables the rapid genetic mapping of ESTs using, for example, the available EUCIB resource (Breen *et al.*, 1994). This protocol (described in full in Chapter 2) is now routinely utilised in a collaborative project, with the Guénet laboratory (Institut Pasteur), for the large-scale genetic mapping of mouse ESTs. A pilot study, establishing the primary use of this protocol in the genetic mapping of ESTs expressed in the mouse ventral midbrain, is reported in Section 1 (1.3).

The experimental stages in the establishment of a PCR-SSCP assay are outlined below.

- Optimisation of PCR conditions
- PCR: amplification of parental DNAs with the incorporation of a radioactive nucleotide (refer also to Section 1, 1.3)
- Electrophoresis: separation of DNAs by in native polyacrylamide gels (19:1 and 37.5:1 acrylamide:bis-acrylamide, 6% polyacrylamide gels)
- Selection of gel matrix providing the more distinct SSCP of the two matrices examined
- Use of selected conditions, for individual ESTs, in genetic mapping studies (Section 1, 1.3 and 1.4)

1.3 Establishment of a genetic mapping strategy for genes expressed in the central nervous system (CNS) of mouse

1.3.1 Establishment of an SSCP assay

An initial series of PCR reactions without incorporation of a radioactive deoxyribonucleotide were performed in order to establish optimal amplification conditions for primers on mouse genomic DNA (SEG/Pas). All primer pairs were observed to provide single products of the expected size as determined by electrophoresis on a 2.5% agarose gel under the PCR conditions given in Table 1.3.1. It should also be noted that primer pair W36162 provided a second distinct band of higher molecular weight. This PCR product was also sequenced (U58912, see Table 1.3.1) and demonstrated no significant sequence homology to the expected product (W36162). However, it did provide a distinct SSCP permitting its chromosomal allocation (Table 1.3.2). PCR reactions with incorporation of radioactive α^{32}P dCTP were then performed on both the parental strains with which the breeding programme was initiated (C57BL/6 and *M. spretus* SEG/Pas) and *M. spretus* SPR. The latter was also included because it was used as a parent in the breeding programme to generate the EUCIB resource which is used subsequently. The products were electrophoresed in a non-denaturing gel for SSCP assessment (Chapter 2) and subsequently visualised by autoradiography. SSCP analysis identified clear polymorphisms discriminating between these parental strains, for >90% of ESTs generated within the putative 3'UTR. This procedure is referred to as PCR*-SSCP analysis throughout. It should be noted that the *Mus spretus* mice utilised by the EUCIB originate from 2 subpopulations: SEG/Pas and SPR (Breen *et al.*, 1994). Of the 15 primer pairs for which results are reported in this study, 14 provided SSCP band patterns distinguishing between both the *M.spretus* substrains and the C57BL/6 strain. However, in the case of the PCR product for EST W33210 the band patterns provided by the two *Mus spretus* strains were clearly distinct but the band pattern provided by the amplified SPR sequence could not be distinguished from that of the C57BL/6 strain.

1.3.1 Chromosomal assignment

Radioactively labelled PCR*-SSCP analyses were performed, for each EST, on panels of DNAs deriving from several backcross generations of a breeding programme that generates consomic lines (Guénet, pers. comm; refer also to Section 1, 1.1.1) as listed in Table 1.3.1 For those ESTs tested on the N2 DNAs (Figure 1.3.1), the resultant typings were analysed using Genelink software. These analyses provided a linkage group around

the corresponding ESTs. For those ESTs tested on the N4 (Figure 1.3.2) or N6/N7 DNAs the resultant typings were compared with the tables of known data (Appendix II and Appendix III respectively), indicating the expected segregation patterns for any SEG/Pas chromosomal fragment flanked by the anchor markers utilised in the breeding programme. These analyses indicated on which chromosome each EST was located. Genotyping data, permitting the chromosomal assignment of these sequences, are listed in Appendix IV.

1.3.2 Determination of genetic map location

1.3.2.1 Position between anchor markers (EUCIB)

The chromosome allocations were confirmed in each case by studying the genetic typing results of EUCIB subset panels (8-20 animals per chromosome), selected from the Mbx database (http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html) for each putative chromosome or regional allocation. The recombination events between anchor markers for the animals in these EUCIB panels had been well characterised and subdivided the relevant chromosome into 4 or 5 distinct regions. PCR*-SSCP analyses were performed on the DNAs from these panels. The segregation pattern observed for each EST confirmed the chromosomal allocation and implicated a region between two anchor markers or between an anchor marker and the centromere (Table 1.3.2) as a likely location. Genotyping data, permitting the confirmation of chromosomal assignment of these sequences, are listed in Appendix VI.

Figure 1.3.1 An example of the genetic linkage analysis, providing chromosomal allocation, of a novel EST using intermediate generation of the breeding programme to generate a consomic panel. Shown are PCR*-SSCP analyses of a 164 bp fragment within the putative 3'UTR of W36213, using as template the parental strain DNAs (*M.spretus* SPR and C57BL/6, lanes 1 and 2 respectively), DNA minus control (Lane 3) and 19 DNAs, from the N2 generation of the breeding program to generate a set of consomic lines with one *Mus spretus* (SEG/Pas) chromosome each on a C57BL/6 background (Lanes 4-22). The clear mobility difference between the parental strains readily allows distinction between the homozygote (e.g. lane 6) and heterozygote (e.g. lane 11) patterns.

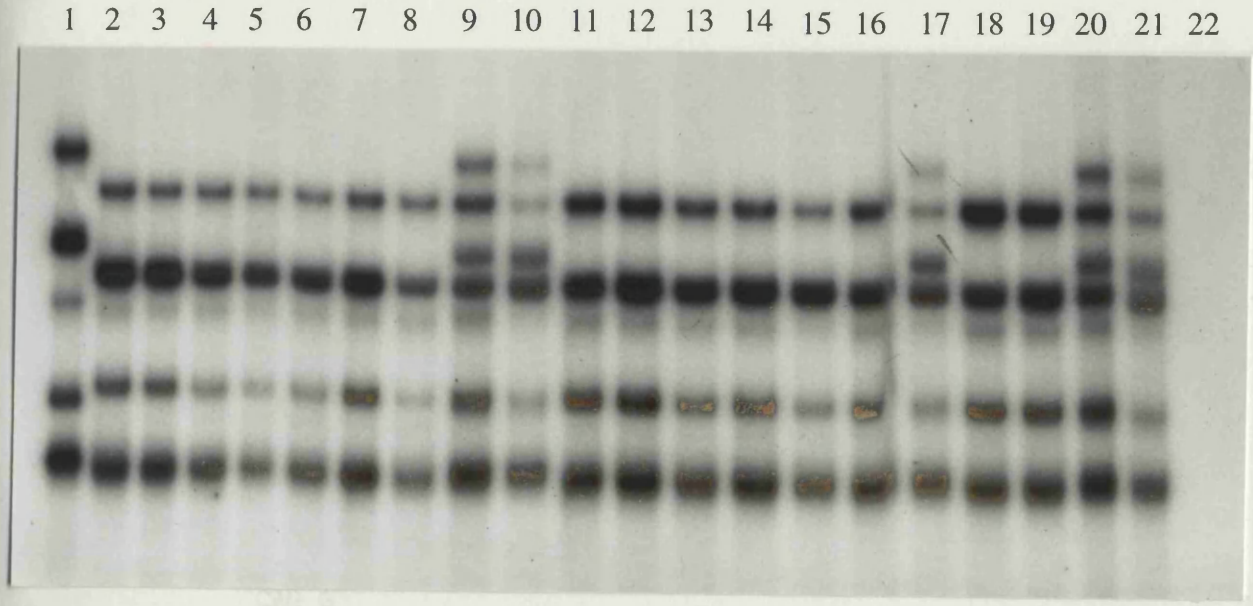


Figure 1.3.2 An example of the genetic linkage analysis, providing chromosomal allocation, of a novel EST using intermediate generation of the breeding programme to generate a consomic panel. Shown are PCR*-SSCP analyses of a 249 bp fragment within the putative 3'UTR of W39933, using as template the parental strain DNAs (*M.spretus* SPR and C57BL/6, lanes 1 and 2 respectively), DNA minus control (Lane 22) and 19 DNAs, from the N4 generation of the breeding program to generate a set of consomic lines with one *Mus spretus* (SEG/Pas) chromosome each on a C57BL/6 background (Lanes 3-21). The clear mobility difference between the parental strains readily allows distinction between the homozygote (e.g. lane 6) and heterozygote (e.g. lane 9) patterns.

Clone identification	Oligonucleotide sequences 5' - 3'	DNA panel used for chromosome allocation	PCR product size(bp)	PCR T_A	PCR cycle number
W36259	CAC AAA CAT CCA CCC TCC ATC TTA ACT TGG TTA GCC	N6/N7	179	55°C	25
W36162	CCA GGT GTA TTG TAT CCA GAG C GTC TCA TAG CGA ACA ATA CAG C	N4	233	52°C	30
W33240	TTC TCG CTG TCT ATG ACT TAG CAG G TTT GAC CAG GAA TCT GAA ATT AGG AG	N4	291	55°C	25
W36135	AAA ACC TGT GTG AGG CGA GC CAT AGA AAG CAC AGA CAA GTG G	N4	275	58°C	25
W39933	ACT GCT TTT CTC GGA CTA CTG C CTT ATC TGA GCA GTG GGA GTG G	N4	180	55°C	25
W36139	AAA ATG ACC GAG GCG TTA GG ACA CTG GCT GTT GAA GAA GC	N4	100	55°C	25
**Th*	AGA GGC TTT CCC ATG TGT GTG G GGA GAC CTT TCC TTC CTT TAT TGA G	N4	172	57°C	30
W33210	AAC GCA CTG CTG CCA TCT AGT G CGT AAG CCC TTA GAA TGC CAG AC	N4	249	57°C	30
W39952	TGA AAT AAG AGG ACT ACT CAA GG AAA CAG TCT TGG GTC ATA GG	N4	143	55°C	25
W39947	CTG AAA CAT TAA ATG GAG GTC CCC CCA GTG AGA GAA CAA ATC CTG ACC	N4	200	53°C	25
W39956	ATT CTA ATG TGG TGA CCT CC ACA GGA GTC TCT GCT TCC	N4	160	58°C	25

Table 1.3.1

Clone identification	Primer sequences 5' - 3'	DNA panel used for chromosome allocation	PCR product size (bp)	PCR T_A	PCR cycle number
W36213	TCA CTG TAA CAT TGG GGG GTG G TTT TGT CTA TGT CCA AGA GG	N2	278	55°C	25
W35449	CAT GGT CTT TCC TTC CAT CAG G GAA GCA GAG AGC TAT AAG CCT TCT G	N4	221	55°C	25
W36221	CAC TAA CTA ACC TCA TTC TTG GG ACA GTT ACT CAG CAC AGT GAA GAC G	N6/N7	221	55°C	25
W36179	GGT CCG TTT CCA CAA TTC AGG ACT ACT AAC TAG CAG CGA CCA GAC CTG	N2	184	55°C	25
U58912	CCA GGT GTA TTG TAT CCA GAG C GTC TCA TAG CGA ACA ATA CAG C	N4	360	52°C	30

Table 1.3.1 continued Chromosomal allocation of 14 ventral midbrain ESTs and 1 anonymous DNA sequence. *, 3'UTR sequence from the mouse tyrosine hydroxylase gene was used as a control for the mapping procedure. Data for EST PCR primer pairs are shown giving their corresponding EST accession number (clone ID), sequence, predicted product size annealing temperature (T_A) and PCR cycle number for these assays.

Clone Identification	Chromosome	Region	Distance from proximal‡ or distal* anchor marker (cM)	Confidence interval (CI)
W36259	1	*Col3a1-D1Mit12*	1.5‡	2.5
W36162	2	*D2Mit11-D2Nds3*	12.19‡	5.5
W33240	2	*D2Mit1-D2Mit11*	7.1*	5.7
W36135	4	*Tryp1-D4Mit52*	6.7‡	3.4
W39933	5	*D5Nds8-D5Nds2*	5.65‡	4.5
W36139	6	*D6Nds4-D6Nds5*	3.25*	3.7
Th	7	*D7Mit40-D7Mit15*	r,c	r
W33210	9	*D9Mit42-D9Nds10*	r	r
W39952	10	*D10Mit20-D10Mit7*	2.1‡	1.7
W39947	10	*D10Mit16-D10Mit20*	13*	6.4
W39956	13	*D13Mit17-D13Mit61*	8.75‡	4.3
W36213	17	*D17Mit39-D17Mit123*	7.8‡	3.5
W35449	18	*D18Mit8-D18Mit25*	3.4‡	1.7
W36221	18	*D18Mit20-D18Mit24*	0‡	<1.8
W36179	19	*D19Nds1-D19Mit1*	6.5‡	4.0
U58912	X	*PLP-DXHar2*	r	r

Table 1.3.2 Localisation of 14 ventral midbrain ESTs and 1 anonymous DNA sequence

Note. Data for EST loci are shown giving their positions as a distance relative to an adjacent anchor marker. All confidence intervals (CI) are at the 95% level. The recombination frequency was calculated as (rf × 100), where rf = number of recombinant events between the proximal marker and the EST × (number of DNAs, recombinant within the region of interest, tested)$^{-1}$. The standard error was calculated as (±SE = (rf(1-rf)/total)$^{0.5}$. All markers generated in this study have been submitted to the Genome Database (GDB) for approval of loci names. At this time loci are named according to their EST accession number.

[c]Positive control allocation, *Position is relative to distal marker, ‡Position is relative to proximal marker, [r]Regional allocation only

1.3.2.2 Distance between anchor markers (EUCIB)

A second EUCIB subset panel was then selected from the Mbx database (http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html), for each EST. These panels included DNAs recombinant within the implicated regions of the genome. More than 750 animals had already been typed for all regions flanked by anchor markers, providing panels of more than 50 DNAs demonstrating recombination events within each region (Breen *et al.*, 1994). PCR*-SSCP analyses were performed on 22 DNAs from each of these panels of recombinants (Figure 1.3.3). At least 19 DNAs provided interpretable results in each instance. The distance between an EST locus and the respective proximal marker was taken as the fraction of interpretable results demonstrating recombination events between the proximal marker and the EST locus, multiplied by the distance between the flanking markers. W33210 could not be further localised using the EUCIB DNAs. The recombinant DNAs in the region of interest *D9Mit42-D9Nds10* derived from *M.spretus* SPR/Pas and C57BL/6 parents alone. These parental DNAs do not demonstrate an observable SSCP for the amplified sequence within this EST and as such did not permit further SSCP genetic typing. It is not surprising to find single-stranded amplification products, from SPR and SEG/Pas *Mus spretus* subpopulations, that demonstrate differing mobilities. The report by Breen *et al.* (1994) notes the occurrence of sequence variants between these subpopulations. Genotyping data, permitting the genetic localisation of sequences reported in this study, are listed in Appendix VI.

1.3.3 Database sequence comparisons

Sequence comparisons were performed against all entries in the Genbank databases using the NIH BLAST (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast [23rd August 1997]) facility (Altschul *et al.*, 1990). Comparisons of the sequences translated in all six reading frames were performed using the TBLASTX (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast [23rd August 1997]) program. High scoring pairs, those demonstrating a high degree (>70%) of identity with data base sequences ESTs in this study, are indicated in Table 1.3.5. Those searches yielding no such homology to ESTs in this study (W36259 and W36179) are indicated by a backslash (/) symbol.

Seven of the ESTs (W36162, W39933, W33210, W39948, W39956, W36213, and W36221) demonstrate homology (>78%) with previously identified gene sequences in mouse, rat or man (Table 1.3.3). Amongst these, W36162, W39948, W36213, and W36221 demonstrate homology with genes whose expression patterns have been characterised and are known to be expressed in the brain (genes encoding: Arrestin-E,

connexin 43, endothelin receptor type-B [endothelin-B^R] and the kinesin heavy chain respectively).

Arrestin-E is a member of a family of proteins which act as agonists to the G-protein linked receptors involved in the phototransduction cascade permitting colour vision (Craft *et al.*, 1994). Connexin 43 is a gap junction protein of the mammalian heart and brain (el Aoumari *et al.*, 1990). Expression in the mammalian brain is restricted to glial fibrillary acidic protein positive ($GFAP^+$) cells (Astrocytes). It is not surprising to find genes with glial expression patterns within such a library (Adams *et al.*, 1991; Davies *et al.*, 1994). Endothelin-B has been reported to be expressed in the neuronal cells of the hippocampus and cerebellum (Zeng *et al.*, 1997). It is also expressed in glial cells, specifically astrocytes and microglia; (Rogers *et al.*, 1997) liver and the placenta (Zeng *et al.*, 1997). Kinesin is known to be a cytoskeletal motor protein which is involved in cell division and axonal transport (Noda *et al.*, 1995). It is noteworthy that EST W33210 demonstrates 95% identity to rat glycoproteins gp55 and gp65, and 98% identity to the unidentified ESTs D50463 and mr26f07.r1. The sequence D50463 derives from a signal trap experiment to detect secreted proteins (Shirozu *et al.*, 1996) which correlates with the fact that glycoproteins gp55 and gp65 are major components of synaptic membrane in the forebrain (Langnaese *et al.*, 1997). Glycoprotein gp65 is brain specific and gp55 is reported present in most tissues examined (Langnaese *et al.*, 1997). Furthermore W39956 demonstrates 97% identity to the 84 kD mouse heat shock protein.

Four ESTs (W33240, W36135, W36139 and W39953) demonstrate homology (>70%) with unidentified rat and mouse cDNA sequences (Table 1.3.5). Three further sequences (W36259, W36179 and U58912) demonstrate no homology to any non-self sequence in these searches.

It is further noteworthy that the gene encoding deoxycytidine kinase (*DCK*) has been mapped in human to chromosome 4 q13.3-q21.1 (Stegmann, 1993) which overlaps the corresponding region of synteny to the region allocated to EST W39933 (between *D5Nds8* and *D5Nds2*; Table 1.3.5).

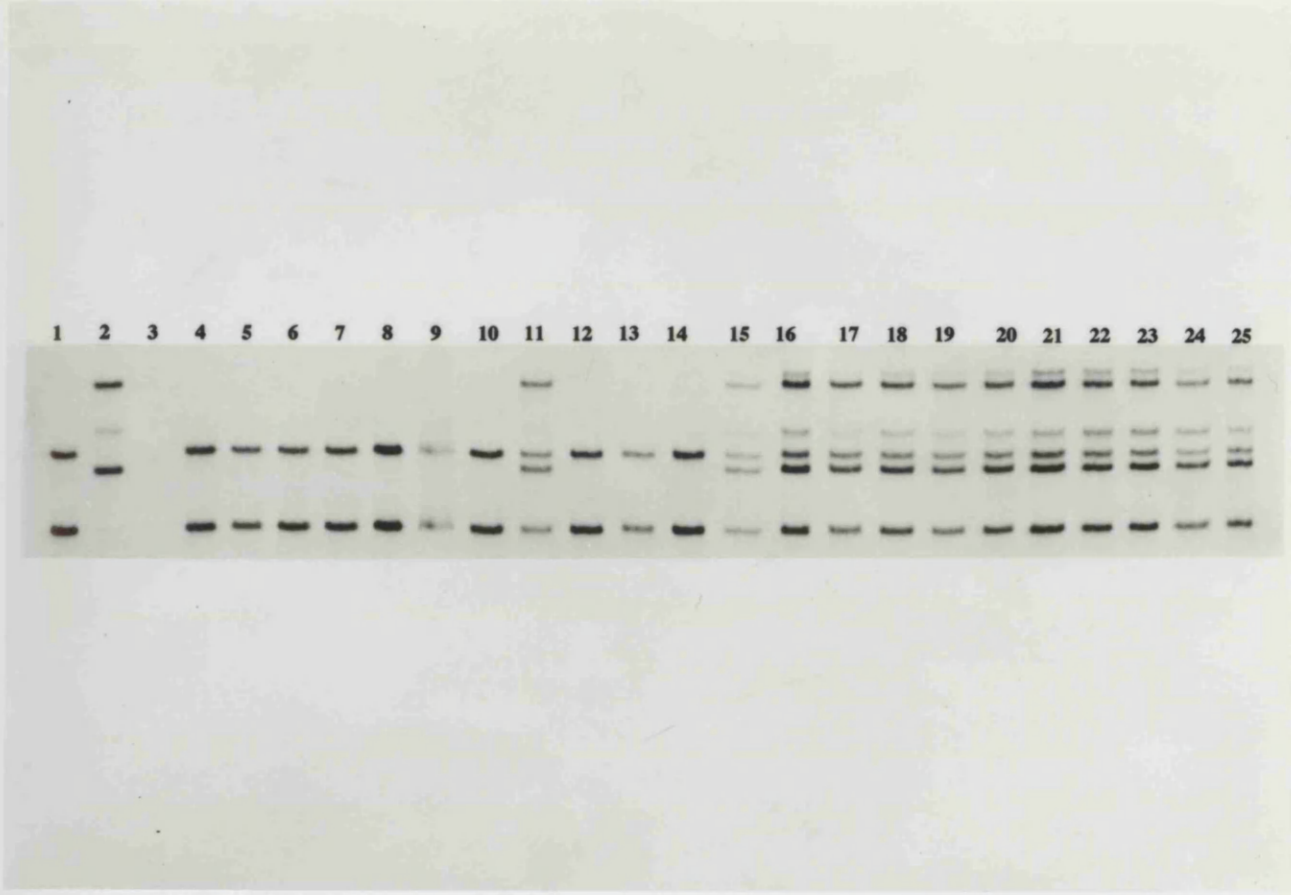


Figure 1.3.3 An example of the genetic localisation of a novel EST a subset panel of DNAs from the EUCIB resource. Shown are PCR*-SSCP analyses of a 179 bp fragment within the putative 3'UTR of W36259, using as template the parental strain DNAs (C57BL/6 and *M.spretus* SPR, lanes 1 and 2 respectively), DNA minus control (Lane 3) and 22 DNAs, recombinant between the markers *Col3α1* and *D1Mit12*, from the backcross [(C57BL/6 x SPR)F_1 × C57BL/6] (Lanes 4-25). The clear mobility difference between the parental strains readily allows distinction between the homozygote (e.g. lane 4) and heterozygote (e.g. lane 11) patterns.

Accession number	Blastn and Tblastx search results	High scoring pairs (Identities with data base sequences)
W36259	/	/
W36162	Blastn: Locus RRU03630 388 bp mRNA Definition *Rattus rattus* arrestin-E mRNA, partial cds. Accession U03630 NID g458208 Source black rat. *Rattus rattus* Tblastx: mp96h09.r1 Soares 2NbMT *Mus musculus* cDNA clone 577121 5' similar to SW:ARRE_RAT P37200 arrestin-E	86% (over 270nt) Minus / plus orientation 94%
W33240	Tblastx: mo28e05.r1 mouse embryo 13 5dpc *Mus musculus* cDNA clone 554912 5'	94%
W36135	Tblastx: EST108869 *Rattus* sp. cDNA 5' end	70% (over 60 residues)
W39933	Blastn: Locus MMDEOKI 2812 bp RNA Definition *Mus.musculus* mRNA for Deoxycytidine kinase. Accession X77731 NID g456676 SOURCE house mouse; *Mus musculus* Tblastx: mi36f09.r1 Soares mouse embryo NbME13.5 14.5 *Mus musculus* cDNA clone 465641 5' similar to gb:X77731 *M.musculus* mRNA for Deoxycytidine kinase	>95% (over 170nt) Plus / Plus 100% (over 50 residues)

Table 1.3.3

Accession number	Blastn and Tblastx search results	High scoring pairs (Identities with data base sequences)
W36139	Blastn: Locus HUMD 5A 09 M3 302 bp mRNA Definition Human HepG2 3' region (gene signature) Accession D17249 NID g598852 Source *Homo sapiens* Tblastx: mr28a09.r1 Soares mouse 3NbMS *Mus musculus* cDNA clone 598744 5'	85% (over 80 nt) Plus / Plus 95% (over 90 residues)
W33210	Blastn: Locus D50463 1900 bp mRNA Definition Mouse SDR1 mRNA Accession D50463 NID g1747303 Source *Mus musculus* Locus RNGP55 2031 bp and RNGP65 200? bp RNA Definition *R.norvegicus* mRNA for glycoproteins 55 and 65. Accession X99337 and X99338 (?) NID g1806275 and g Source Norway rat; *Rattus norvegicus* Tblastx: mr26f07.r1 Soares mouse 3NbMS *Mus musculus* cDNA clone 598597 5'	98% (over 150 nt) Plus / Plus >95% (over 300 nt) Plus / Plus 98% (over 100 residues)
W39953	Tblastx: ve72e09.r1 Soares mouse mammary gland NbMMG *Mus musculus* cDNA clone 831784 5'	98% (over 90 residues)

Table 1.3.3

Accession number	Blastn and Tblastx search results	High scoring pairs (Identities with data base sequences)
W39948	Blastn:	
	Locus MMCOX43 1926 bp DNA	96% (over 380 nt)
	Definition *M.musculus* gene for connexin 43.	
	Accession X62836	
	NID g50522	
	KEYWORDS connexin43; gap junction protein; transmembrane protein.	
	Source house mouse; *Mus musculus*	
	Tblastx:	
	mn27e07.r1 Beddington mouse embryonic region *Mus musculus* cDNA clone 539172 5' similar to gb:X61576 M.musculus mRNA for connexin 43	97% (over 100 residues)
W35449	Tblastx:	
	mq82h11.r1 stratagene mouse melanoma *Mus musculus* cDNA clone 585285 5'	96% (over 78 residues)
	vk50c05.r1 stratagene mouse T cell 937311 *Mus musculus* cDNA clone 958088 5'	91% (over 71 residues)
W39956	Blastn:	
	Locus S46109 806 bp mRNA	98% (over 248 nt)
	Definition HSP90=heat shock protein	
	Accession S46109	
	NID g257730	
	SOURCE *Mus* sp. heart.	
	Tblastx:	
	my22a12.r1 Barstead mouse pooled organs MPLRB4 *Mus musculus* cDNA clone 696574 5' similar to gb:M18186 Mouse 84 kD heat shock protein mRNA	97% (over 46 residues)

Table 1.3.3

Accession number	Blastn and Tblastx search results	High scoring pairs (Identities with data base sequences)
W36213	Blastn: Locus HUMHETBR7 2972 bp DNA Definition Human gene for endothelin-B receptor (hET-BR), exon 7. Accession D13168 NID g285924 SOURCE *Homo sapiens*	78% (over 219 nt)
	Tblastx: vf75g04.r1 Soares mouse mammary gland NbMMG *Mus musculus* cDNA clone 849654 5' similar to gb:S57283 endothelin-B receptor (Human)	82% (over 84 residues)
W36221	Blastn: Locus MMU86090 3791 bp mRNA Definition *Mus musculus* pancreatic beta-cell kinesin heavy chain mRNA, Accession U86090 NID g2062606 SOURCE house mouse; *Mus musculus*	98% (over 410 nt)
	Tblastx: mz91d01.r1 Soares mouse lymph node NbMLN *Mus musculus* cDNA clone 720769 5' similar to gb:X65873 kinesin heavy chain (Human)	95% (over 98 residues)
W36179	/	/
U58912	/	/

Table 1.3.3 Data generated by sequence comparisons performed against all entries in the Genbank databases using the NIH BLAST (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-blast, 23rd August 1997) [comparison at the nucleotide level] and TBLASTX (http://www.ncbi.nlm.nih. gov/cgi-bin/ BLAST/nph-blast, 23rd August 1997) [comparison at the amino acid level] facilities. Searches yielding no homology (>70%) to ESTs in this study are indicated by a backslash (/) symbol.

1.3.4 Association with known neurological mutations in mouse

Three of the ESTs mapped in this study are located close to one or more neurological mutations in unidentified genes. These are found on: chromosome 5 (chr5), *tilted* (W39933); chr10, *hesitant* (W39948); chr13, *twirler*, *ataxia* and *sphingomyelinosis* (W36221). However, all three sequences demonstrate >95% identity to DNA sequences from known genes. EST W39933 is linked to the *tilted* locus. However, the EST W39933 corresponds to the gene encoding deoxycytidine kinase. Loss of function mutation in this gene would result in the failure of DNA replication of a dividing cell and subsequent cell death. This must preclude its involvement in a transmissable mutant phenotype. EST W39948 is linked to the *hesitant* locus. The *hesitant* mouse demonstrates ataxia and reduced male fertility (Kapfhamer *et al.*, 1996). The EST W39948 corresponds to the gene encoding connexin 43, which is a gap junction protein of the brain and heart (el Aoumari *et al.*, 1990). It is conceivable that proteins involved in cell-cell communication in the brain may also play a role in neuropathy. Genotyping of the marker for W39948 on the backcross panel utilised in mapping the *hesitant* locus may prove worthwhile in excluding the gene, corresponding to this EST, from a role in the *hesitant* pathology or present it as a candidate gene for further study. EST W36221 is linked to the loci for the mutants *twirler*, *ataxia* and *sphingomyelinosis*. This EST corresponds to the gene encoding kinesin. The reported function of this protein, as a cytoskeletal motor protein in axonal transport and cell division (Noda *et al.*, 1995), presents it as a good candidate for involvement in neuropathy. Thus the marker for W36221 may also be genotyped on the respective backcross panels utilised in mapping the *twirler*, *ataxia* and *sphingomyelinosis* loci or any neuropathology linked to this region in the future.

1.3.5 Discussion

I have utilised SSCP in the mapping of 14 ventral midbrain expressed ESTs. This straightforward and sensitive technique permits the resolution of polymorphisms in short amplified single stranded sequences differing by a single base (Orita *et al.*, 1989[a,b]; Suzuki *et al.*, 1990; Okamato *et al.*, 1991; Beier *et al.*, 1992; Beier, 1993; Glavac and Dean 1993; Hayashi and Yandell 1993; Kupryjanczyk *et al.*, 1993; Spinardi *et al.*, 1993; Beier *et al.*, 1995; McCallion *et al.*, 1996). Using single pass sequencing of 3' directed cDNAs allows ready access to the highly polymorphic 3'UTR (Ko *et al.*, 1994; Wilcox *et al.*, 1991). Each EST has been mapped genetically using the combined resources of genetically typed DNAs deriving from several backcross generations of a breeding programme that generates consomic lines (refer also to Section 1, 1.1.1; Guénet, pers.

comm.) and EUCIB DNAs. The map distances and locations have been assigned using the linkage analysis programmes provided by the Mbx database (http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html) and Genelink software. The EUCIB facility is the most precise and largest purpose bred mouse mapping resource presently available (Breen *et al.*, 1994). Discordances are evident between map positions assigned to anchor loci on EUCIB and Jackson Laboratory maps (Encyclopedia of the Mouse Genome, release 3, http://www.informatics.jax.org.). Such discordances predict that new loci mapped on the EUCIB panel can only tentatively be matched with known loci placed on the map according to the Jackson Encyclopedia of the mouse genome and vice versa. Subsequent correlation of mapping data, generated using the EUCIB and Jackson resources, requires approval by the respective chromosome committees.

The consomic line breeding programme was initiated to provide a stable and less limiting source of DNA for the immediate chromosomal allocation of new markers within the mouse genome. The progressive manner in which each generation is typed, for anchor markers, reduces the loss of DNA in screening final generations (section 1.1; Guénet, pers. comm.). For those considering the development of such consomic lines in other organisms it should be noted that our results show that typed DNA from early backcross generations can be used for chromosomal allocation.

This study also raises questions of relevance to researchers utilising SSCP and EUCIB DNA in mapping. The differing motility of single-stranded amplification products from SPR and SEG/Pas *Mus spretus* strains is not unexpected. Sequence variants have previously been noted between these *Mus spretus* subpopulations (Breen *et al.*, 1994). Clearly, instances will continue to arise where a marker distinguishes between C57BL/6 and one of the *M.spretus* parental subpopulations but not the other. It is not possible at present to determine the frequency with which this may occur. Researchers intending to use SSCP with the EUCIB DNAs should be aware of this and the resulting possibility of a reduction in the available precision in mapping. The extent to which precision may be lost is dependent on the number of progeny available, from the informative subspecies, for a region of interest.

The sequence homology between certain ESTs (W36162, W39933, W33210, W39948, W39956, W36213, and W36221) combined with a knowledge of the source tissue provides a greater level of information for those looking for expressed sequences corresponding to candidate genes in a region associated with a known pathology. Without further investigation no great significance can be laid upon the correlation of any EST locations with regions associated with neuropathies. However, the correlation is

noteworthy and it is expected that further investigation of these and other brain expressed ESTs will, in the future, provide gene candidates for known neuropathies. I have established a collaborative project with the laboratory of Jean-Louis Guénet (Institut Pasteur) with the aim of increasing the numbers of ESTs genetically mapped in the mouse genome. The preliminary results of this work are reported in Section 1, 1.4.

1.4 Collaborative genetic mapping of ESTs expressed in the ventral midbrain of mouse

1.4.1 Establishment of collaborative studies

The establishment of an integrated map, which includes information regarding the genetic and physical location of a sequence as well as its temporal and spatial expression patterns requires the mapping of large numbers of ESTs (Takahashi and Ko, 1993; refer also to Chapter 1). Such a map will permit elucidation of the organisation of mammalian genomes, particularly gene clustering and gene number (Levitt, 1991; refer also to Chapter 1). I have established a collaborative project with the Guénet laboratory (Institut Pasteur) with the aim of genetically mapping expressed sequences in the mouse genome (refer also to discussion in Section 1, 1.3), on a larger scale than the study reported in Section 1, 1.3. This study reports the preliminary data arising out of this collaborative work. In excess of 100 primer pairs have been designed to sequence within the putative 3'UTR of these cDNA fragments. I report the genetic mapping of the first 32 of these ESTs. These ESTs (Stewart *et al.*, 1996) derive from the same cDNA library resource utilised in Section 1, 1.2 and 1.3 (Savioz and Davies, 1995).

A second pilot study has also been established, with the Serikawa laboratory (Kyoto University, Japan). The purpose of which was to assess the utility of PCR oligonucleotides, designed to sequence within mouse ESTs, as genetic markers within the rat genome. Hence preliminary data is also reported, indicating the chromosome location of 7 EST loci, genetically mapped in the mouse genome, within the rat genome. These studies were performed by Dr Yasushi Kondo (Kyoto University, Japan) and utilised rat × mouse hybrid DNA panels, constructed within the Serikawa laboratory (Serikawa, pers. comm.). Mapping genes in mouse and rat will permit further definition of the boundaries of syntenic regions between the species (Guénet, pers. comm; Serikawa, pers. comm.). ESTs localising near previously identified synteny boundaries will be selected for fine mapping in both species, refining the regions encompassing the boundaries in question.

1.4 .2 Extant data from the collaborative mapping of expressed sequences within the mouse genome

All data arising, thus far, from this study are listed in Table 1.4.1. Allocation of ESTs to a particular chromosome and their subsequent localisation were performed as described in Section 1, 1.3. It is noteworthy that two of the ESTs demonstrated length polymorphism

between the mouse strains *M. spretus* (SEG/Pas) and C57BL/6. These amplified products could be resolved by electrophoresis in ordinary agarose gels. Consequently the SSCP assay was not required to allocate these sequences to mouse chromosomes.

EST	Mouse chr.	X-over between EST and closest marker	Confidence interval	Probability	Polymorphism detection method	Rat chr
GP 1.176	ND	ND	ND	ND	Agarose	ND
GP 1.108	1	0/23 with *Col3a1*	0.00-12.21	$2.40E^{-7}$	SSCP	ND
GP 2.228	16	0/27 with *D5Mit379*	0.00-10.5	$1.50E^{-8}$	SSCP	ND
GP 3.106	ND	ND	ND	ND	ND	ND
GP 4.145	11	0/31 with*D11Nds19*	0.00-9.21	$9.30E^{-10}$	Agarose	ND
GP 4.196	8	0/36 with *D8Mit19*	0.00-7.98	$2.90E^{-11}$	SSCP	ND
GP 4.265	1	1/39 with *Col3a1*	0.07-13.48	$1.50E^{-10}$	SSCP	ND
GP 5.9	16	0/27 with *Prm1*	0.00-10.5	$1.50E^{-8}$	SSCP	ND
GP 5.170	14	0/31 with *D14Nds1*	0.00-9.21	$9.30E^{-10}$	SSCP	ND
GP 6.15	9	5/36 with *D9Mit19* (proximal)	4.67-29.5	$1.30E^{-5}$	SSCP	ND
GP 6.236	3	0/28 with *D3Mit19*	0.00-10.15	$7.50E^{-10}$	SSCP	ND
GP 7.15	11	0/30 with *D11Nds9*	0.00-9.5	$1.90E^{-9}$	SSCP	ND
GP 7.28	4	0/22 with *Mos*	0.00-12.73	$4.80E^{-7}$	SSCP	ND
SPMP22.6	11	3/16 with *Duf-1*	4.05-45.65	0.02127	SSCP	ND
GP 2.22	4/11	5/20 with *D11Pas1*	8.66-49.1	0.04139	SSCP	ND
GP 1.25	7	0/19 with *serca1*	0.00-14.39	$3.80E^{-6}$	SSCP	ND
GP 2.30	1	3/15 with *Mylf* and *Vil*	4.33-48.09	0.03516	SSCP	14
GP 3.36	2	2/36 with *D2Mit1* (proximal)	0.68-18.66	$1.90E^{-8}$	SSCP	ND

Table 1.4.1

EST	Mouse chr	X-over between EST and closest marker	Confidence interval	Probability	Polymorphism detection method	Rat chr
GP 3.42	16	1/15 with *Prm1* (proximal)	0.17-31.95	0.00098	SSCP	11
GP 2.48	4	1/16 with *Mos*	0.16-30.23	0.00052	SSCP	5
GP 3.54	17/3	ND	ND	ND	SSCP	ND
GP 2.61	4	2/26 with *Mos* (proximal)	0.95-25.13	$1.00E^{-5}$	SSCP	ND
GP 3.64	8	4/29 with *Junb* (proximal)	3.89-31.66	0.0001	SSCP	ND
GP 3.71	5	2/14 with *D5Pas2*	1.78-42.81	0.1294	SSCP	ND
GP 2.107	7	3/21 with *Hras* (proximal)	3.05-36.34	0.0149	SSCP	ND
GP 2.120	7	ND	ND	ND	SSCP	ND
GP 2.129	9	1/17 with *Mod1*	0.15-28.69	0.00027	SSCP	8
GP 2.137	2	1/21 with *D2Mit*1	0.12-23.82	$2.1E^{-5}$	SSCP	ND
GP 2.138	14	ND	ND	ND	SSCP	15
GP 2.234	5	ND	ND	ND	SSCP	14
GP 2.245	8/6	ND	ND	ND	SSCP	ND
GP 2.271	2/5	ND	ND	ND	SSCP	3

Table 1.4.1 Localisation of ESTs expressed in the ventral midbrain of mouse

Data for EST PCR primer pairs are shown giving their corresponding EST identification number (clone ID). *Note.* Data for EST loci are shown giving their positions as a recombination fraction relative to their respective adjacent anchor marker. All locations are centromere distal of the nearest marker except where otherwise stated. All confidence intervals (CI) are at the 95% level. Where available, EST chromosome locations established within the rat genome are also listed. chr, chromosome; ND, indicates that data from a particular stage in the experimentation has not yet been obtained (not done). Where two or more chromosomes are provided the location of the EST remains to be confirmed.

1.4 .3 Discussion

This report extends the work performed in Section 1, 1.3. It demonstrates the use of an already established strategy (McCallion *et al.*, 1996; McCallion *et al.*, in prep.) in the genetic mapping of mouse ESTs on a larger scale. It has been estimated that mapping 1000 expressed sequences will permit the identification of up to 20 genes involved in previously identified pathologies (Guénet, pers. comm.) based solely on the correlation of EST and disease loci. As discussed in Section 1, 1.3 the correlation of any EST locus with a region associated with a neuropathy is noteworthy. Further investigation of these and other ESTs will, in the future, provide gene candidates for known pathologies. The availability of gene targeting strategies in mouse permits subsequent evaluation of gene function (refer also to Chapter 1 and Section 2, 2.6).

1.5 Qualitative assay of spatial and temporal expression of ESTs selected from the genetic mapping study

1.5.1 Qualitative-RT PCR assay of the temporal and spatial expression pattern of selected ESTs

The degree of novelty evident within the EST resource utilised in this study strengthens suggestion that it provides a valuable resource in the search for genes with relevance to neuropathology (Levitt, 1991; Stewart *et al.*, 1996). However, it is acknowledged that selection of particular genes, for further study, based upon sequence information alone would be premature (Stewart *et al.*, 1996). It is further suggested (Stewart *et al.*, 1996) that little can be said about the interest level of novel sequences without further characterisation of their expression patterns or sequence characterisation of their coding regions. The report by Savioz and Davies (1995) demonstrates the use of heterologous probes from precise regions of the bovine brain in an endeavour to identify genes with higher levels of expression in the substantia nigra (SN) than in the total ventral midbrain or cerebellum. This type approach can be somewhat laborious (Davies *et al.*, 1994) and, due to the physical constraint of how many probes may be manipulated at any one time, only provides information concerning a very limited number of tissues. This chapter describes a pilot study initiated with the aim of establishing a qualitative screen for the detection of gene transcripts corresponding to ESTs produced by the ongoing EST sequencing project in the laboratory of R.W. Davies (Davies *et al.*, 1994; Stewart *et al.*, 1996).

Many studies have demonstrated the use of PCR in the quantitation of RNA species (Delidow *et al.*, 1989; Wang *et al.*, 1989; Ferre, 1992; Foley *et al.*, 1993; Grassi *et al.*, 1994; Hamoui *et al.*, 1994). These studies attempt to circumvent a variety of sources of quantitative error, inherent in the PCR, including: inconsistencies in the reaction efficiencies of individual reverse transcription (RT) (Delidow *et al.*, 1989; Grassi *et al.*, 1994) and PCR (Delidow *et al.*, 1989; Wang *et al.*, 1989; Ferre, 1992; Grassi *et al.*, 1994) steps; inconsistencies in the reaction concentrations of template molecules (Delidow *et al.*, 1989); the exponential nature of PCR (Wang *et al.*, 1989; Ferre, 1992; Foley *et al.*, 1993) and the reproducibility of results (Ferre, 1992; Foley *et al.*, 1993). It has also been acknowledged (Foley *et al.*, 1993; Hamoui *et al.*, 1994) that establishment of a quantitative PCR strategy, even for a single primer pair, requires substantial time and effort. This approach was considered too labour intensive and expensive to implement for use as a large scale expression screen. Consequently, the aim of this pilot study was simply to provide information on the presence or absence of any desired RNA species in a tissue of

interest at a particular postnatal timepoint. The value and effective use of PCR in such studies has been acknowledged previously (Delidow *et al.*, 1989). The PCRs were maintained at a moderate cycle number (30 cycles) in an attempt to reduce PCR anomalies. All results are consistent upon repetition.

1.5.2 Qualitative-RT PCR assay of expression of selected ESTs in tissues of the central and peripheral nervous systems and in non-brain tissues

In an endeavour to discover if any of the selected ESTs demonstrated brain specific expression, PCRs were performed on total cDNA preparations from a number of CNS, PNS and non-brain tissues. Template cDNAs from brain (CNS; postnatal day 16), sciatic nerve (P16), spleen, kidney, testis and liver were PCR amplified using primer pairs designed to putative 3'UTR sequence of selected ESTs (listed in Table 1.5.1; defined in Table 1.3.1). The results of these PCRs are listed in Table 1.5.1. Detection levels are arbitrarily defined as strong, moderate or weak.

Four of the eleven arbitrarily selected EST primer pairs (W33240, W33210, W36213, W36221) detect their corresponding cDNA species, at moderate to strong detection levels, in all tissues examined (Table 1.3.1). cDNA corresponding to the EST primer pair W39956 is also detected in all tissues examined but is only weakly detected in the brain and in kidney (Table 1.3.1). EST primer pairs W36139 and W39953 detect their corresponding cDNA species, at moderate to strong detection levels, in all tissues examined except kidney (W36139) and spleen (W39953) (Table 1.3.1). cDNAs corresponding to the EST primer pairs W36259 and W39933 are detected (at weak to moderate levels) in brain, kidney, testis and spleen (W36259 only) but are not detected in sciatic nerve or liver. EST primer pair W35449 detects its corresponding cDNA species weakly in brain tissue, sciatic nerve and liver. It is noteworthy that EST W33210 demonstrates the same degree of sequence identity with the two genes encoding the glycoproteins gp65 and gp55 respectively; the latter of which demonstrates almost ubiquitous expression (Shirozu *et al.*, 1997). Expression of the gene encoding glycoprotein 65 is restricted to the brain (Shirozu *et al.*, 1997). W36213 corresponds to the endothelin-B receptor type which has been reported to be detected in brain, liver and placenta (Noda *et al.*, 1995). There is no evidence to suggest that its expression should be restricted to these tissues. PCR primers, designed to amplify sequence from the gene encoding cyclophillin were used as a control for invariant transcript levels (Montague *et al.*, 1997). Transcript from this gene is readily detected in all tissues under examination (Table 1.3.1). Transcript corresponding to the EST W36135 is not detectable in any tissue examined in this study. Figure 1.5.1 A-D illustrates the result of RT-PCR assays for four

of the EST sequences in this study.

Clone Identification	Brain	Sciatic nerve	Spleen	Kidney	Testis	Liver	Sequence
W36259	++	-	++	++	+	-	unknown
W33240	+++	+++	+++	+++	+++	+++	cDNA (mouse)
W36135	-	-	-	-	-	-	cDNA (rat)
W39933	+	-	-	+	+	-	DCK
W36139	++	+++	+++	-	+++	+++	cDNA (man)
W33210	+++	+++	+++	+++	+++	+++	gp55 & gp65
W39953	+++	+++	-	+++	++	++	cDNA (mouse)
W39956	+	+++	++	+	++	++	HSP90
W36213	++	+++	++	++	++	+++	hET-BR
W35449	+	+	-	-	-	+	cDNA (rat)
W36221	++	+++	++	+++	+++	+++	Kinesin
Cyclophilin	+++	+++	+++	+++	+++	+++	/

Table 1.5.1 RT-PCR detection of RNA species, corresponding to EST sequences, in brain (P16) and non-brain (P15) tissues. Detection levels are arbitrarily identified as: +++, strong; ++, moderate; +, weak and -, not detected. Corresponding sequences (Sequence) are identified as: DCK, Deoxycytidine kinase (mouse); gp55 and gp65, glycoproteins 55 (rat) and 65 (rat) respectively; HSP90, 90 kD heat shock protein (mouse); hET-BR, endothelin receptor-type B (man); Kinesin, kinesin heavy chain (mouse). PCR primers designed to amplify sequence from the gene encoding cyclophilin, were utilised as a positive control for expression at all examined tissues.

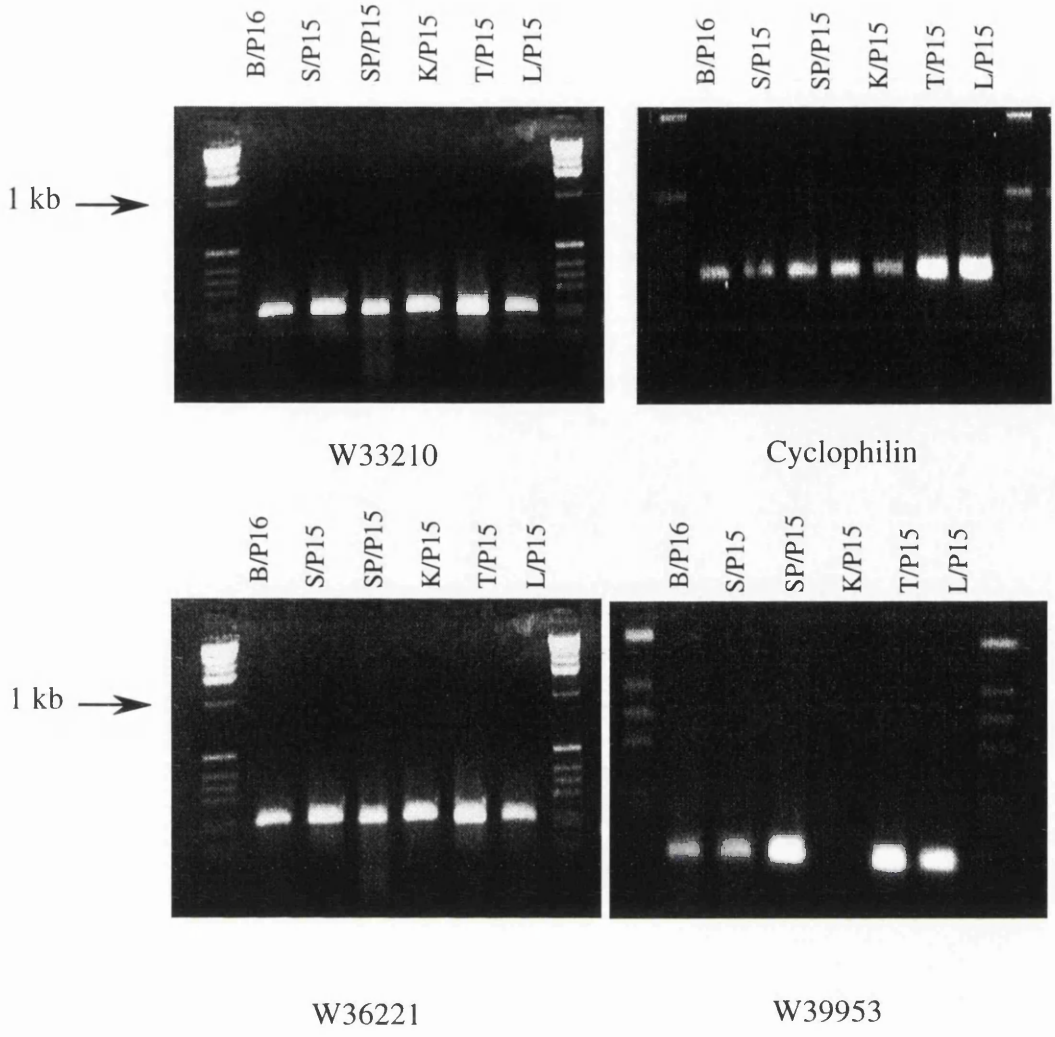


Figure 1.5.1 Examples of results from RT-PCR assays performed using RNA derived from brain and non-brain tissues (as described in Section 1.5.1) with oligonucleotides designed to amplify sequence within the EST sequences: W33210 (Figure 1.5.1 A), Cyclophilin (Figure 1.5.1 B), W36221 (Figure 1.5.1 C) and W39953 (Figure 1.5.1 D). B: brain; S: Spinal cord; SP: Spleen; K: Kidney; T: Testis; L: Liver. P[n]: Post natal day

1.5.3 Qualitative RT PCR assay of postnatal expression in the CNS

In a subsequent endeavour to asses the temporal expression pattern of these ESTs, within CNS tissues; PCRs were performed on cDNAs from spinal cord and brain at a number of postnatal timepoints. Template cDNAs from spinal cord (P1, P4) and brain (P4, P5, P16, P43) were PCR amplified using primers utilised in 1.5.1 (listed in Table 1.5.2; defined in Table 1.3.1). The results of these PCRs are listed in Table 1.5.2.

That nine of the eleven EST sequences examined were detectable in the adult (P43) is not surprising; the library from which the ESTs were generated was prepared using 8 week old mice (Davies *et al.*, 1994). Of the six ESTs (W36259, W33240, W36135, W36139, W39953, W35449) demonstrating no homology to any previously identified gene sequence, transcripts corresponding to two (W33240 and W39953) were detected in the spinal cord (P1, P5) but remained undetected in the brain until P16 (Table 1.5.2). The amplified cDNA sequence corresponding to EST W35449 was only detected, in this assay, at timepoints P16 and P43 (Table 1.5.2). Transcripts, corresponding to ESTs W36259 and W36139, were detected at moderate to strong and weak levels respectively, at all timepoints examined (Table 1.5.2). Transcript sequence corresponding to EST W36135 was not detected at any timepoint in this assay, nor was it detected in any tissue. This is not surprising; the library from which these ESTs were generated (Savioz and Davies, 1995) is enriched for low expression genes in the ventral midbrain and the brain cDNA utilised was prepared from whole brain RNA (Montague *et al.*, 1997). The cDNA species corresponding to EST W36135 may not have been represented amongst the template species utilised in the PCR, due to its very low representation amongst whole brain RNA. However, it may more simply reflect a failure of the PCR amplification step utilising the corresponding primer pair. PCR primers, designed to amplify sequence from the gene encoding cyclophillin were used as a control for invariant transcript levels (Montague *et al.*, 1997).

Clone Identification	Spinal cord (P1)	Spinal cord (P5)	Brain (P4)	Brain (P5)	Brain (P16)	Brain (P43)	Sequence
W36259	+++	+++	++	++	+++	+++	unknown
W33240	+++	++	-	-	+	++	cDNA (mouse)
W36135	-	-	-	-	-	-	cDNA (rat)
W39933	-	-	-	-	-	+++	DCK
W36139	+	+	+	+	+	+	cDNA (man)
W33210	++	+++	-	-	+	-	gp55 & gp65
W39953	+	+	-	-	+++	+++	cDNA (mouse)
W39956	-	-	-	-	+	-	HSP90
W36213	+++	++	++	+	+	+	hET-BR
W35449	-	-	-	-	+	+++	cDNA (rat)
W36221	++	++	++	++	+++	+++	Kinesin
Cyclophilin	+++	+++	+++	+++	+++	+++	/

Table 1.5.2 RT-PCR detection of RNA species, corresponding to EST sequences, in CNS tissues of mice at six postnatal timepoints (P, Postnatal day). Detection levels are arbitrarily identified as: +++, strong; ++, moderate; +, weak and -, not detected. Corresponding sequences (Sequence) are identified as: DCK, Deoxycytidine kinase (mouse); gp55 and gp65, glycoproteins 55 (rat) and 65 (rat) respectively; HSP90, 90 kD heat shock protein (mouse); hET-BR, endothelin receptor-type B (man); Kinesin, kinesin heavy chain (mouse). PCR primers designed to amplify sequence from the gene encoding cyclophilin, were utilised as a positive control for expression at all examined postnatal timepoints.

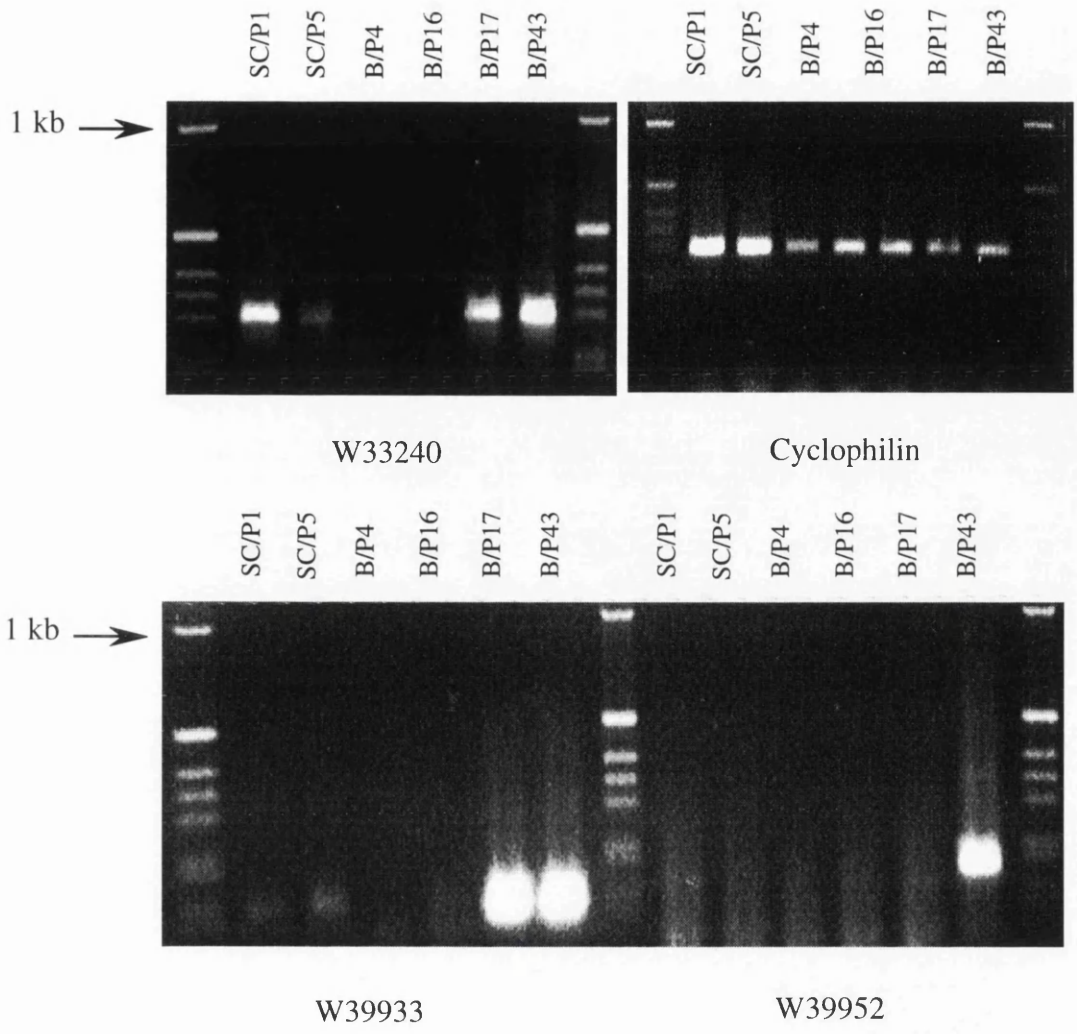


Figure 1.5.2 Examples of results from RT-PCR assays performed using RNA derived from CNS tissues (P1, P4, P5, P16, P17 and P43; as described in Section 1.5.2) with oligonucleotides designed to amplify sequence within the EST sequences: W33240 (Figure 1.5.1 A), Cyclophilin (Figure 1.5.1 B), W39933 (Figure 1.5.1 C) and W39952 (Figure 1.5.1 D). B: Brain; SC: Spinal cord. P/[n]: Post natal day.

1.5.4 Discussion

I have utilised RT-PCR in the detection of cDNA species corresponding to ESTs generated by the ongoing sequencing project (Stewart *et al.*, 1996; Stewart *et al.*, unpublished) and subsequently used in the establishment of a genetic mapping strategy for mouse ESTs (McCallion *et al.*, 1996; McCallion *et al.*, in press). This pilot study provides an initial and arbitrary screen. It establishes a sensitive, yet qualitative approach which provides information regarding the presence or absence of an RNA species in a tissue at a specific timepoint. Transcript corresponding to each EST has been sought in non-brain (spleen, kidney, testis and liver), PNS (sciatic nerve) and CNS (spinal cord; P1, P5 and brain; P4, P5, P16, P43). Transcript from the gene encoding cyclophilin is utilised throughout as a spatially and temporally invariant positive control (Montague *et al.*, 1997).

Identification of transcript species present in all tissues examined is not surprising (Davies et al., 1994; Savioz and Davies, 1995). The data reported by Savioz and Davies (1995) suggest that only 1.16% of transcripts sought will correspond to genes with expression enhanced in the substantia nigra (SN) but not in other brain regions. No indication is given, however, with respect to what proportion may be expected to demonstrate brain restricted expression. The study further identifies a number of novel sequences that remain undetected in some tissues (W36259 and W35449) and in the brain at certain postnatal timepoints (W36135, W33240 and W39953).

The data generated by this study would suggest that tentative inferences may be made with respect to the library source and the ESTs generated from it. All but two transcript sequences sought were detected in the adult brain. This correlates with the use of adult mice in the library preparation. It does not, however, suggest that lack of detection equates to lack of transcript presence in the tissue. Detection of genes that are expressed, at low levels, in a tissue specific manner in the brain may require a more refined approach. For the information provided by such a screen to be dramatically enhanced, the utility of RNA from regionally dissected tissue e.g. brain; ventral midbrain, hippocampus, cerebellum, thalamus, pons etc. must be examined.

The number of EST sequences utilised in this pilot study is far too small to gain any insight into how well the expected composition (Savioz and Davies, 1995) of the library correlates with the observed. However, this study establishes that such a qualitative assay can be utilised to provide an arbitrary means of selection of sequences for further study i.e. *in situ* hybridisation (ISH), isolation and sequencing of full length cDNAs, protein prediction or functional study (gene disruption). Detection of a subset of transcripts

expressed only in the brain, or more precisely only in specific brain regions, will provide one such platform for selection of sequences for further study.

I have also established a collaborative project with the laboratory of Dr M. Schalling (Karolinska Institute, Sweden) for the performance of ISH with probes designed to sequence derived from selected ESTs. This will provide an immediate assay, of spatial expression in the mouse brain, for any selected sequence arising out of this or related studies (genetic mapping) using this EST resource.

1.6 Discussion

These studies establish an efficient strategy for the genetic mapping of mouse ESTs. They also establish a qualitative RT-PCR screen of the temporal and spatial expression patterns of selected ESTs. The combined use of these strategies provides a firm foundation for the identification of gene candidates for involvement in known neuropathies, and in particular for neuropathies affecting the ventral midbrain e.g. Parkinson's disease.

1.6.1 Correlation with disease

These studies have also established a number of noteworthy correlations with disease loci (refer to Section 1, 1.3). The ability to correlate map positions of expressed sequences with disease loci is widely accepted as a crucial step in the identification of a gene associated with a pathological process (Levitt, 1991; Wilcox *et al.*, 1991; Polymeropoulos *et al.*, 1992, 1993; Takahashi and Ko, 1993; refer also to Section 1, 1.1). As described in Chapter 1, this form of candidate gene approach has already been validated (Gibson *et al.*, 1995) by its use in the successful identification of *shaker1* locus in mouse, and the subsequent demonstration (Well *et al.*, 1995) that the homologous gene in man is associated with Usher syndrome Type 1B. Though the numbers of ESTs mapped in this study are small, it is expected that increasing the throughput of the mapping process will result in the provision of a significant number of gene sequences as candidates for involvement in known pathologies (refer also to Chapter 1; Section 1, 1.3 and Section 2, 2.1-2.2).

1.6.2 Future studies

1.6.2.1 Increasing the number of ESTs genetically mapped in the mouse

A total of 46 ESTs have been genetically mapped in this series of studies. A further 80 primer pairs have been designed to sequence within the putative 3'UTRs of ESTs. The collaborative group, established during the process of these studies, have undertaken to genetically map in excess of 100 ESTs within the next six months. This author acknowledges that the availability of radiation hybrid DNA panels, for mapping the mouse genome, will in the future permit high throughput genetic mapping in mouse to become a reality. When such panels become available they will be utilised in association with available automated systems (robot workstations and multiwell plate systems) to increase the rate at which new ESTs are mapped. In the future, where possible, expressed sequences mapped in mouse will also be mapped in man and rat, aiding the development

of a comparative gene map and improving access to mutations within all three species.

1.6.2.2 Increasing the information obtained from the qualitative RT-PCR assay of the temporal and spatial expression patterns of selected ESTs

The establishment of a qualitative PCR screen, providing information regarding the presence or absence of an RNA species in a tissue at a specific timepoint, represents a distinct contribution to the establishment of an integrated gene map (defined and discussed in Chapter 1). This contribution will be further enhanced by the use of region specific tissues from the nervous system (refer also to discussion in Section 1, 1.5). Consequently, temporal and spatial expression data will be available for each EST, mapped by this collaborative unit, in the future.

1.6.3 Conclusion

These studies (Section 1, 1.2-1.5) represent a considerable contribution to the establishment of a gene map of the mouse genome. I established a technically feasible and efficient way to assay large numbers of expressed sequences with respect to their position in the mouse genome, and their temporal and spatial expression patterns. The PCR-SSCP assay, developed in this study, is currently being utilised for the genetic mapping of ESTs in mouse (Guénet laboratory; Institut Pasteur, Paris) and rat (Serikawa laboratory; Kyoto, Japan).

Section 2

2.1 Introduction

2.1.1 Identification and study of a gene encoding a novel major myelin component

The single gene *Mobp* encodes the novel major myelin component myelin-associated oligodendrocytic basic protein (MOBP) (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). As described in Chapter 1 a cDNA fragment, corresponding to a portion of 3' UTR within *Mobp*, was first identified among cDNAs derived from substantia nigra and striatum (Davies *et al.*, 1994; Savioz and Davies, 1995). Like *Mbp*, the *Mobp* gene encodes a family of related oligodendrocyte specific basic proteins of 69, 71, 81, 99 and 170 amino acids (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a). These protein isoforms correspond to predicted M_r values of 8.2, 8.4, 9.7, 11.7 and 19 kD respectively (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). Antibodies raised against the common N-terminal (aa 1-68), shared by all previously identified MOBP isoforms, identified bands on western blots larger than those predicted from their respective cDNA sequences (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). The reason for these discrepancies in M_r values has not yet been elucidated but is thought likely to be a consequence of post-translational modification of the protein isoforms (Holz *et al.*, 1996; Montague *et al.*, 1997a). Though all isoforms thus far identified (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a) share in common amino acid residues 1-68, they differ in the length and polarity of their respective C-terminal regions (Holz *et al.*, 1996; Montague *et al.*, 1997a). The predicted MOBP protein isoforms are, like MBP, basic, hydrophilic and positively charged (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). Like MBP, MOBP lacks a signal peptide and is predicted to be an extrinsic membrane associated protein (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a). MBP binds the negatively charged acidic phospholipids of the cytoplasmic membrane, an interaction known to stabilise the structure of myelin (Roach *et al.*, 1985; Mikoshiba *et al.*, 1991). MOBP has been shown to localise in the MDL of compact CNS myelin (Yamamoto *et al.*, 1994; Montague *et al.*, 1997a) in a manner similar to MBP. It has been suggested that MOBP may play a role in compacting or stabilising the myelin sheath (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a). A recent study (Montague *et al.*, 1997b), which evaluated MOBP expression in the myelin mutant *shiverer*, has demonstrated that the 20 kD MOBP isoform is significantly reduced in the brain of *shiverer* mice when compared with wild-type animals. This suggests that any role that MOBP plays in the compaction of myelin may be secondary to or dependant on MBP. This suggestion is strengthened by the results of

recent study (Montague *et al.*, 1997a) which demonstrated that MOBP was barely detectable in the *jimpy* mutant mouse, which exhibits severe hypomyelination. However, these studies do not exclude MOBP from a role in stabilising myelin via interactions with MBP, cytoskeletal or other membrane molecules. Another recent study (Montague *et al.*, 1997a) has also demonstrated that MOBP colocalises with the microtubular network of oligodendrocytes.

In total the reports by Yamamoto (1994) and Holz (1996) have identified 6 *Mobp* splice variants, in the rat. Published data (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a) indicate that the *Mobp* transcriptional unit is subject to a complex series of alternative splicing events (refer also to discussion in Section 2, 2.3). *In situ* hybridisation studies (Montague *et al.*, 1997) have demonstrated that individual splice variants, of the *Mobp* mRNA, can be detected in the ventral columns of the spinal cord around postnatal day 1 (P1). Further evidence supporting the proposed role for the *Mobp* gene product in myelination is found in the fact that the temporal occurrence of the *Mobp* transcript correlates closely with the assembly of myelin sheaths (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a). Furthermore, the caudal to rostral progression of signal, corresponding to the *Mobp* transcript, with respect to time, in the postnatal brain, also parallels the spatial and temporal presence of *Mbp* transcript (Kirstensson *et al.*, 1986), and the progress of myelination.

Two of the reported splice variants (*Mobp* 81A and *Mobp* 81B; Holz *et al.*, 1996) encode the same MOBP isoform. These transcripts differ only in their respective 3'UTRs. It is noteworthy that these transcripts localise to different parts of the oligodendrocyte (Holz *et al.*, 1996; Montague *et al.*, 1997) and it has been suggested that, as suggested for *Mbp* RNAs (Brophy *et al.*, 1993), their respective 3'UTRs may contain transport signals (Holz *et al.*, 1996). The *Mobp* transcript may be transported to the oligodendrocyte processes for translation on free ribosomes (Holz *et al.*, 1996; Montague *et al.*, 1997), in a manner similar to *Mbp* RNAs (Colman *et al.*, 1982). It has also been suggested that the differential localisation of transcripts may also be influenced by oligodendrocyte maturity (Montague *et al.*, 1997). The report by Montague *et al.*, (1997) also demonstrates that, as with other myelin genes (*Plp*, *Mbp*, *Cnp* and *Mag*), *Mobp* splice variants are detectable differentially during development. It is noteworthy that the splice variant encoding MOBP 170 can be detected at embryonic day 12 (E12) (Montague *et al.*, 1997), whereas the other identified splice

variants are only detectable in the postnatal brain (Montague *et al.*, 1997). This suggests that MOBP 170 may have a function alternative or additional to involvement in myelination (Montague *et al.*, 1997).

Myelin-associated oligodendrocytic basic protein (MOBP) is a major CNS myelin component. Therefore mutations in the gene encoding it might well underlie an inherited CNS neuropathy, and/or a known mouse neurological mutation. In order to evaluate the possible role of the *Mobp* gene product in known neuropathies I have undertaken the series of studies reported in Section 2 (2.2-2.6). Firstly, I have genetically mapped the *Mobp* gene in the mouse (McCallion *et al.*, 1996). Secondly, in order to elucidate the complex splicing events required to generate the known splice variants of the *Mobp* gene transcript and to provide a platform for the functional analyses of the *Mobp* gene product, I have examined the genomic organisation of the mouse gene. I have also examined the correlation between the *Mobp* gene structure and the production of the identified splice variants (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a). Lastly, I have used the information obtained from these studies to pursue a gene disruption strategy in an endeavour to begin to elucidate the function of this novel gene.

Multiple sclerosis is a major human demyelinating disease which preferentially affects CNS myelin. Multiple sclerosis is generally accepted to be an autoimmune disorder, but the autoantigen(s) responsible is not known (Martyn, 1991; Allen and Brankin, 1993; Steinman, 1996). Oligodendrocyte-specific myelin proteins are particularly interesting in this context (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a). The *Mobp* gene product is expressed exclusively in the CNS, and is therefore an interesting candidate gene for further study in relation to multiple sclerosis (refer also to discussion in Section 2, 2.7).

2.2 Genetic mapping of the murine gene (*Mobp*) encoding myelin-associated oligodendrocytic basic protein

2.2.1 Establishment of an SSCP assay

An initial series of PCRs, that did not involve incorporation of a radioactive deoxyribonucleotide (dNTP), were performed to establish optimal amplification conditions for primers ASM 215 F/R (defined; Table 2.4, Chapter 2) on mouse genomic DNA (*Mus spretus*; SEG/Pas). These primers were observed to provide a single product of the appropriate size (205 bp) as determined by electrophoresis in 2.5% agarose gels under the PCR conditions given in Chapter 2. The identity of the PCR product was confirmed by sequencing, as described in Chapter 2. A PCR, incorporating the radioactive deoxyribonucleotide αP^{32}dCTP was then performed on the parental (C57BL/6 and SEG/Pas) DNA templates. The products were separated by electrophoresis in a non-denaturing gels for SSCP assessment (as described in Chapter 2) and subsequently visualised by autoradiography. SSCP analysis identified a clear polymorphism discriminating between parental strains (Figure 2.2.1, lanes 1-3). It should be noted that the *Mus spretus* mice utilised by the EUCIB originate from 2 subpopulations: SEG/Pas and SPR (Breen *et al.*, 1994) and provide correspondingly distinct SSCP band patterns. In the case of this *Mobp* PCR product the two *Mus spretus* strains both differed clearly from C57BL/6, and from one another (Figure 2.2.1, lanes 1-3).

2.2.2 Chromosomal assignment

Radioactively labelled PCR and SSCP analysis (PCR-SSCP*) were performed on a panel of 21 DNAs. This panel included 19 DNAs from the N4 generation of the consomic breeding programme (described in Section 1, 1.1) and both the parental strains with which the programme was initiated (C57BL/6 and *Mus spretus* [SEG/Pas]). No heterozygotes were observed amongst the panel of N4 DNAs. This result implicated one of 6 regions known to be uninformative amongst this panel, since the corresponding regions of *Mus spretus* (SEG/Pas) DNA were not present in this N4 set: Chromosome (Chr 7) 7, 15-51 cM; Chr 8, region 4.6-35 cM; Chr 9, 31 cM-Telomere; Chr 10, 46-64 cM; Chr 11, 2-37 cM and Chr X centromere - 3 cM. Subsequently, PCR-SSCP* analysis was performed on a panel of DNAs including 19 DNAs from the N2 generation and both parental strains. The SSCP typing results for amplification of these DNAs with the PCR primers ASM 215 F/R (defined; Table 2.4, Chapter 2) can be seen in Figure 2.2.1. These data were then entered into a Genelink analysis programme (Montagutelli, 1988) containing all of the

available N2 generation typing data for all anchor loci utilised in the generation of the consomic lines. The segregation pattern implicated chromosome 9 alone.

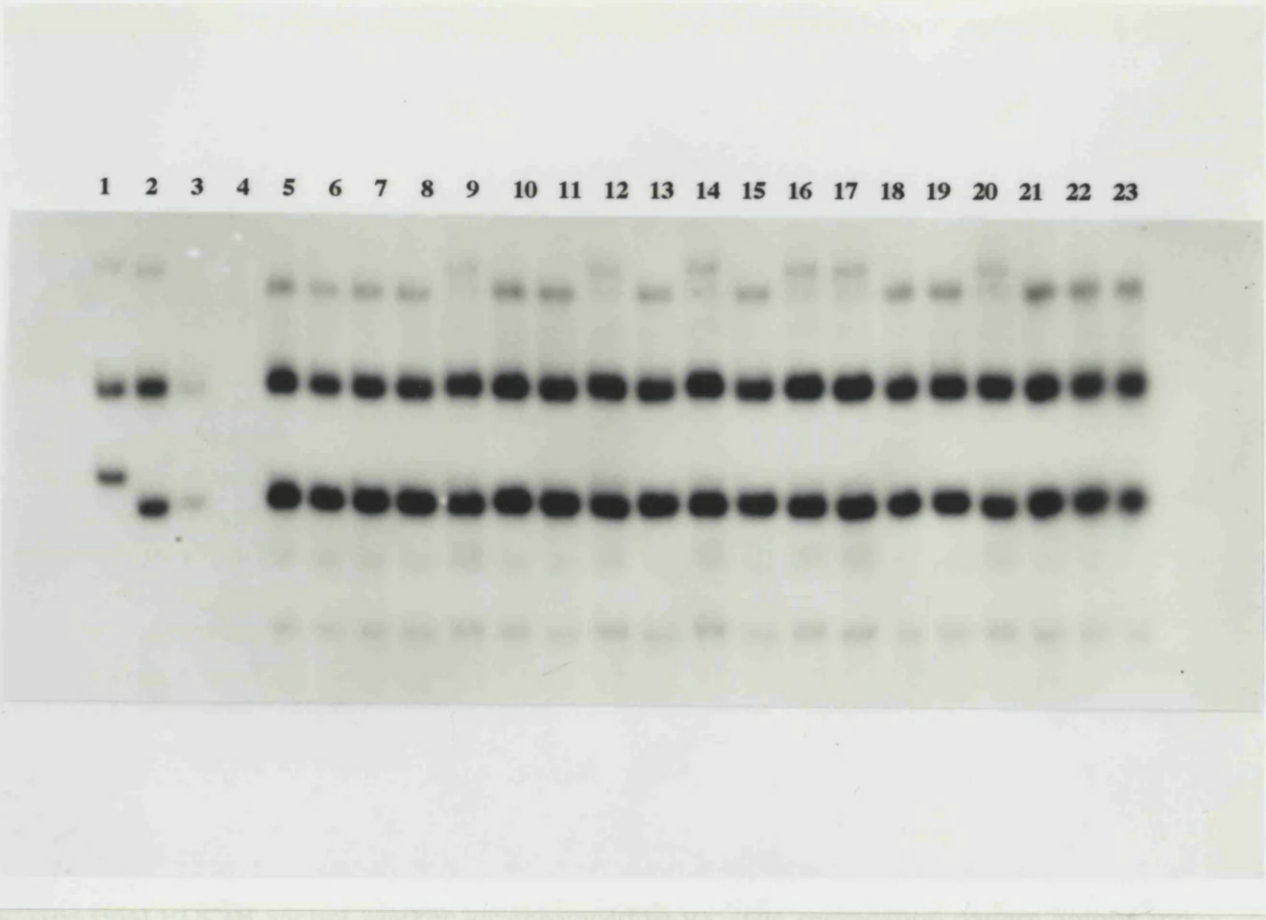


Figure 2.2.1 An example of the genetic linkage analysis, providing chromosomal allocation, of a novel EST using intermediate generation of the breeding programme to generate a consomic panel. Shown are PCR*-SSCP analyses of a 205 bp fragment within the 3'UTR of *Mobp*, using as template the parental strain DNAs (*M.spretus* SEG/Pas, SPR and C57BL/6, lanes 1, 2 and 3 respectively), DNA minus control (Lane 4) and 19 DNAs, from the N2 generation of the breeding program to generate a set of consomic lines with one *Mus spretus* (SEG/Pas) chromosome each on a C57BL/6 background (Lanes 5-23). The clear mobility difference between the parental strains readily allows distinction between the homozygote (e.g. lane 6) and heterozygote (e.g. lane 12) patterns.

2.2.3 Determination of genetic map location

A EUCIB subset panel of 8 animals was selected from the Mbx database (http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html). The recombination events between anchor markers for these animals had been well characterised and subdivided chromosome 9 into 4 regions. PCR-SSCP* analysis was performed on the DNAs from this panel. The segregation pattern (Figure 2.2.2) observed for *Mobp* (ASM 215 F/R; Table 2.3, Chapter 2), shown in Figure 2.2.2 and listed in Table 2.2.1, implicated the region between *D9Mit24* and the telomere.

2.2.3.1 Position between anchor loci

A second EUCIB subset panel was then selected from the Mbx database. This panel included 17 DNAs non-recombinant between *D9Mit24* and *D9Mit19* and 17 DNAs (defined by suffix *; Table 2.2.1) demonstrating recombination events within this region. These DNAs were selected to establish the position of *Mobp* relative to *D9Mit24* and *D9Mit19*, i.e. whether *Mobp* was situated between these markers or telomeric to *D9Mit19* (the final EUCIB anchor marker for chromosome 9). The segregation pattern among this limited DNA set demonstrated that in animals non-recombinant for the region *D9Mit24* to *D9Mit19* no recombination events were observed between *Mobp* and either marker but that among animals demonstrating recombination events within this region, several recombination events were observed between *Mobp* and *D9Mit24* (Table 2.2.1).

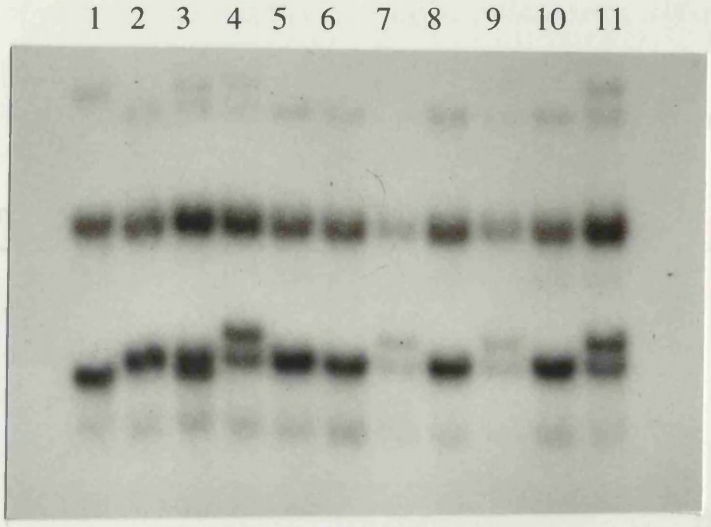


Figure 2.2.2 Location of *Mobp* to a subdivision of chromosome 9 using the EUCIB panel. Shown are PCR*-SSCP analyses of a 205 bp fragment within the putative 3'UTR of *Mobp*, using as template the parental strain DNAs (*M.spretus* SEG and C57BL/6) lanes 1 and 2 respectively and a subset panel of 9 DNAs from EUCIB resource (Lanes 3-11). The clear mobility difference between the parental strains readily allows distinction between the homozygote (e.g. lane 6) and heterozygote (e.g. lane 11) patterns. Refer also to Table 2.2.1. N.B. Lane 3 corresponds to PCR-SSCP assay performed using DNA from the EUCIB resource which used the *M. spretus* substrain SEG/Pas as the *Mus spretus* parent. Lanes 4-11 correspond to PCR-SSCP assays performed using DNAs from the EUCIB resource which used the *M. spretus* substrain SPR as the *Mus spretus* parent. The difference in electrophoretic mobility between single-stranded DNA molecules deriving from these two backcross sources is evidence of a strain difference between *Mus spretus* SEG/Pas and *Mus spretus* SPR (refer also to Discussion 2.2.6) within the amplified sequence.

		DNAs with known recombination events within Chromosome 9							
Marker	Marker Location	LS061	LB118	LB104	LS348	LB050	LB165	LB112	LB113
D9Mit42	8cM	1[a]	2	2	1	2	1	2	1
D9Nds10	21cM	2[a]	1	2	1	2	1	2	1
D9Nds8	28cM	2	1	1	2	2	1	2	1
D9Mit24	52cM	2	1	1	2	1	2	2	1
D9Mit19	66cM	2	1	1	2	1	2	1	2
Mobp	-	2	1	1	2	1	2	1	2

Table 2.2.1 Location of *Mobp* to a subdivision of chromosome 9. The typing data displayed above, for 5 EUCIB anchor markers on chromosome 9, clearly subdivide the chromosome into 4 regions. The PCR-SSCP alleles of the mouse gene *Mobp* demonstrate the same segregation pattern as marker *D9Mit19*, thus allocating *Mobp* to the region between *D9Mit24* and the telomere. [a]1 and 2 correspond to homozygote and heterozygote SSCP patterns respectively.

2.2.3.3 Localisation of *Mobp* within the interval *D9Mit24 - D9Mit19*

The segregation pattern observed indicated that *Mobp* is situated between *D9Mit24* and *D9Mit19*. The recombination events within this panel more specifically implicated the region between *Col7α1* and *D9Mit19*. The Mbx database (http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html) indicated that 42 DNAs had been typed for this region, using *Col7α1*, of which 33 demonstrated recombination events in the interval. This was considered insufficient for accurate localisation. Consequently animals recombinant in the region *D9Mit24* to *D9Mit19* were chosen.

952 animals had been typed for this region, providing 126 recombinants between *D9Mit24* and *D9Mit19*, including 33 known recombinants between *Col7α1* and *D9Mit19*. All 126 samples provided interpretable results. The segregation pattern observed indicated that the *Mobp* gene is situated between *D9Mit24* and *D9Mit19*. Recombination events between *D9Mit24* and *Mobp* indicated that *Mobp* was located at a distance from *D9Mit24*, equivalent to 84.1% (106/126) of that between *D9Mit24* and *D9Mit19*, while recombination events between *D9Mit19* and *Mobp* indicated that *Mobp* was located at a distance distance from *D9Mit19*, equivalent to 15.9% (20/126) of that between *D9Mit24* and *D9Mit19*. These data (Table 2.2.2) place *Mobp* 2.2 cM (±1cM) proximal to *D9Mit19* (Figure 2.2.3).

2.2.3.4 Genotyping of a new EUCIB anchor marker *D9Mit55*

In an endeavour to further refine the map position allocated to *Mobp*, the marker *D9Mit55* was also mapped relative to *D9Mit24* and *D9Mit19* using the same DNA panel. 107 samples provided interpretable results. Recombination events between *D9Mit24* and *D9Mit55* were 42% (45/107) of those observed between *D9Mit24* and *D9Mit19*, while recombination events between *D9Mit19* and *D9Mit55* were 58% (62/107) of recombination events between *D9Mit24* and *D9Mit19*. These data (Table 2.2.3) place *D9Mit55* 8.1 cM proximal to *D9Mit19*, and thus 5.9 cM proximal to *Mobp*. It should be noted that both *D9Mit55* and *Mobp* may now be used as anchor loci on the EUCIB panel (Figure 2.2.3) and have been their associated data have been submitted to the Mbx database (http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html).

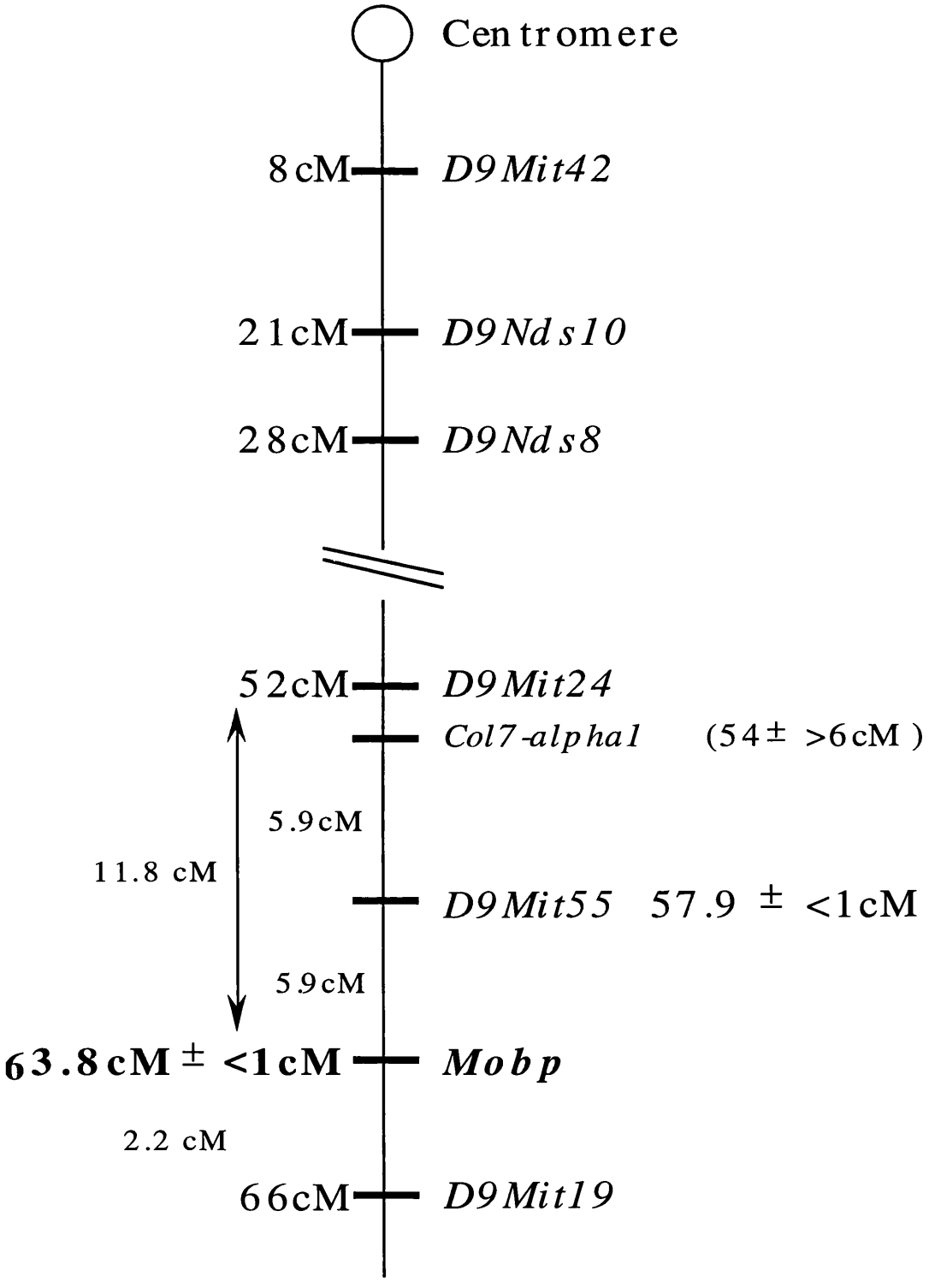


Figure 2.2.3 Genetic map of mouse chromosome 9 based on the Mbx database (http://www.hgmp.mrc.ac.uk/MBx/MBx Homepage.html) showing the location of the *Mobp* gene encoding myelin-associated oligodendrocytic basic protein. The distances given to the left of the map are based on recombination estimates of *Mobp* and *D9Mit55* loci typed using the EUCIB DNA panel (Table 2.2.2 and Table 2.2.3).

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LB112*	a1	Yes	LS193	2	Yes	LS813	/	/
LB113*	a2	Yes	LS196	2	Yes	LS815	2	No
LB147*	a/	/	LS197	1	Yes	LS817	/	/
LB159*	/	/	LS217	2	No	LS818	2	Yes
LB162	/	/	LS229	1	Yes	LS819	1	No
LB164	2	Yes	LS230	1	Yes	LS821	1	No
LB207*	1	Yes	LS258	2	Yes	LS832	1	Yes
LB211*	/		LS261	2	Yes	LS854	/	/
LB234*	2	Yes	LS265	2	No	LS868	1	Yes
LB236*	/	/	LS266	2	Yes	PB017	2	No
LB245*	1	No	LS276	/	/	PB024	1	No
LB290*	2	Yes	LS326	1	Yes	PB063	1	Yes
LB303*	1	Yes	LS331	2	Yes	PB064	1	Yes
LB309	1	Yes	LS371	2	Yes	PB074	1	Yes
LB350	1	Yes	LS404	1	Yes	PB075	2	No
LB389*	1	Yes	LS448	1	Yes	PB080	1	Yes
LB398	2	Yes	LS452	2	Yes	PB086	2	Yes
LB409*	2	Yes	LS458	2	Yes	PB113	/	/
LB424*	/	/	LS462	2	Yes	PB125	2	Yes
LB443*	/	/	LS470	1	Yes	PB130	2	Yes
LB506*	1	Yes	LS492	/	/	PB136	2	Yes
LB509*	/	/	LS552	1	Yes	PB144	1	Yes
LB511	/	/	LS554	2	Yes	PB162	2	No
LB514	1	No	LS559	1	Yes	PB171	2	Yes
LB521	2	Yes	LS569	2	Yes	PB194	/	/
LB528	1	Yes	LS577	1	No	PB195	1	Yes
LB529	2	Yes	LS579	2	No	PB201	1	Yes
LB542	1	No	LS583	2	Yes	PB202	2	Yes
LB545	1	No	LS586	2	Yes	PB223	2	Yes
LB591	2	Yes	LS589	1	Yes	PB233	1	Yes
LB592	2	No	LS704	1	Yes	PB238	2	Yes

Table 2.2.2

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LB598	2	Yes	LS714	2	Yes	PB245	2	Yes
LB613	2	Yes	LS726	1	Yes	PS003	/	/
LB615	1	No	LS728	2	Yes	PS009	1	Yes
LS025	2	Yes	LS730	1	Yes	PS013	1	Yes
LS094	1	Yes	LS731	1	Yes	PS017	/	/
LS095	/	/	LS758	1	Yes	PS049	1	Yes
LS122	2	Yes	LS762	1	Yes	PS050	2	Yes
LS127	1	Yes	LS767	2	Yes	PS051	1	Yes
LS151	2	Yes	LS776	1	Yes	PS056	1	Yes
LS154	1	Yes	LS777	1	Yes	PS061	2	Yes
LS178	2	Yes	LS782	1	Yes	PS062	1	No
LS191	2	Yes	LS808	1	Yes	PS066	2	Yes

Table 2.2.2 continued. Localisation of *Mobp* between EUCIB anchor markers *D9Mit24* and *D9Mit19*. The data displayed above are for the PCR-SSCP* typing of *Mobp* (ASM 215; defined in Table 2.3, Chapter 2) on DNAs demonstrating recombination events in the interval *D9Mit24* - *D9Mit19*. $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LB112	a1	Yes	LS193	1	No	LS813	1	No
LB113	a2	Yes	LS196	2	Yes	LS815	/	/
LB147	a/	/	LS197	2	No	LS817	2	No
LB159	2	No	LS217	2	No	LS818	1	No
LB162	2	/	LS229	1	Yes	LS819	/	/
LB164	2	Yes	LS230	2	No	LS821	/	/
LB207	/	/	LS258	/	/	LS832	2	No
LB211	/	/	LS261	/	/	LS854	/	/
LB234	2	Yes	LS265	/	/	LS868	/	/
LB236	/	/	LS266	/	/	PB017	2	No
LB245	/	/	LS276	/	/	PB024	1	No
LB290	1	No	LS326	2	No	PB063	1	Yes
LB303	2	No	LS331	1	No	PB064	1	Yes
LB309	1	Yes	LS371	1	No	PB074	1	Yes
LB350	1	Yes	LS404	1	Yes	PB075	2	Yes
LB389	1	Yes	LS448	2	No	PB080	2	No
LB398	2	No	LS452	1	No	PB086	/	/
LB409	2	Yes	LS458	2	Yes	PB113	/	/
LB424	/	/	LS462	/		PB125	1	No
LB443	/	/	LS470	/		PB130	/	/
LB506	1	Yes	LS492	/		PB136	1	No
LB509	2	No	LS552	1	Yes	PB144	2	No
LB511	1	No	LS554	/		PB162	2	No
LB514	1	No	LS559	/		PB171	2	Yes
LB521	2	Yes	LS569	2	Yes	PB194	/	/
LB528	2	No	LS577	/		PB195	1	Yes
LB529	1	No	LS579	/		PB201	1	Yes
LB542	/	/	LS583	1	No	PB202	1	No
LB545	/	/	LS586	2	Yes	PB223	2	Yes
LB591	2	Yes	LS589	2	No	PB233	1	Yes
LB592	2	No	LS704	2	No	PB238	1	No

Table 2.2.3

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LB598	1	No	LS714	2	Yes	PB245	2	Yes
LB613	2	Yes	LS726	2	No	PS003	2	No
LB615	/	/	LS728	/	/	PS009	2	No
LS025	/	/	LS730	1	Yes	PS013	1	Yes
LS094	1	Yes	LS731	1	Yes	PS017	1	No
LS095	/	/	LS758	/	/	PS049	/	/
LS122	/	/	LS762	/	/	PS050	1	No
LS127	1	Yes	LS767	2	Yes	PS051	1	Yes
LS151	2	Yes	LS776	1	Yes	PS056	1	Yes
LS154	1	Yes	LS777	1	Yes	PS061	1	No
LS178	2	Yes	LS782	2	No	PS062	1	No
LS191			LS808	2	No	PS066	2	Yes

Table 2.2.3 Localisation of *D9Mit55* between EUCIB anchor markers *D9Mit24* and *D9Mit19*. The data displayed above are for the PCR typing of a microsatellite sequence defined by the marker *D9Mit55* (ASM B667 F/R; defined in Table 2.3, Chapter 2) on DNAs demonstrating recombination events in the interval *D9Mit24* - *D9Mit19*. $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

2.2.4 Association with known neurological mutations in mouse and man

These data show that the *Mobp* gene encoding mouse myelin-associated oligodendrocytic basic protein is located centromere-distal on chromosome 9 at 63.8 cM ± <1 cM (Figure 2.2.3), in a region containing three uncharacterised mouse neurological mutations which are located in the corresponding region of the Mouse Genome Chromosome 9 consensus map (Encyclopedia of the Mouse Genome, release 3, http://www.informatics.jax.org.). These mutations are: *tippy* (Lane, 1984); *spinner* (Deol and Robins, 1962) and *ducky* (Snell, 1955), of which *ducky* is the only mutant for which myelin deficiency has been reported as a phenotypic component.

2.2.5 Examination of possible involvement in the neurological mutant ducky (*du*)

The mouse mutant *du* has been mapped centromere proximal to the marker *D9Mit55* (Frankel, pers. comm.). The position of *D9Mit55*, which is not one of the EUCIB anchor markers, was previously unknown with respect to *Mobp*. Thus, genotyping of this marker, using the same panel of DNAs as *Mobp*, permits direct comparison of the loci. These data (Tables 2.2.2 and 2.2.3 and Figure 2.2.3) clearly place *D9Mit55* and as a consequence the *du* locus, in a position proximal to *Mobp* and thus excludes the possibility of *Mobp* involvement in the *du* pathology.

2.2.6 Discussion

I have mapped the gene (*Mobp*) encoding the mouse homologue of *Mobp* to distal chromosome 9 (Figure 2.2.3). The data locate *Mobp* at 63.6 cM from the centromere (±1cM). Comparison of data listing known recombination events between marker (*D9Mit24*, *D9Mit55* and *D9Mit19*) loci, and those data generated by the genotyping of the *Mobp* marker (ASM 215 F/R) indicate that *Mobp* lies between *D9Mit55* and *D9Mit19*. The accuracy of the map position provided by the analysis of these data approaches the limit of precision afforded, by the EUCIB resource, to this region of the mouse genome at this time. The map distances and location of *Mobp* have been assigned using the linkage analysis programs provided by the Mbx database (http://www.hgmp.mrc.ac.uk/MBx/MBx Homepage.html) and Genelink software (Montagutelli, 1988). Also, *Mobp* has been placed on the map provided by the Mbx analysis of EUCIB DNAs. This mapping facility is the most precise and largest purpose bred mouse mapping resource presently available (Breen *et al.*, 1994). Discordances are evident between map positions assigned to anchor loci on EUCIB and Jackson Laboratory maps (Encyclopedia of the Mouse Genome,

release 3, http://www.informatics.jax.org.). Such discordances predict that new loci mapped on the EUCIB panel can only tentatively be matched with known loci placed on the map according to the Jackson Encyclopedia of the mouse genome and vice versa.

The consomic line breeding programme was initiated to provide a stable and less limiting source of DNA for the immediate chromosomal allocation of new markers within the mouse genome. The progressive manner in which each generation is typed, for anchor markers, reduces the loss of DNA in screening final generations (refer also to Section 1, 1.1.3; Guénet, pers. comm.). For those considering the development of such consomic lines in other organisms it should be noted that our results show that typed DNA from early backcross generations can be used for chromosomal allocation.

This study also raises questions of relevance to researchers utilising SSCP and EUCIB DNAs in mapping. As discussed in section 1.3, the differing motility of single-stranded amplification products from SPR and SEG/Pas *Mus spretus* subpopulations (Figure 2.2.2) is not unexpected; the report by Breen *et al.* (1994) notes the occurrence of sequence variants between these *Mus spretus* subpopulations. In the case of *Mobp,* all three parental strains differed in band mobility in the SSCP assay. Researchers intending to use SSCP with the EUCIB DNAs should be aware of this and the resulting possibility of a reduction in the available precision in mapping. The extent to which precision may be lost is dependent on the number of progeny available, from the informative subspecies, for a region of interest (refer also to the discussion of Section 1.3).

The map position assigned to *Mobp* by this study (Figure 2.2.3) and its situation between *D9Mit55* and *D9Mit19* suggests that it is closely linked to 3 known neurological mutations in the corresponding region on the Jackson map (Encyclopedia of the Mouse Genome, release 3, http://www.informatics.jax.org.). The published genetic mapping data available for these mutants was inadequate to allow exclusion of allelism with *Mobp* and further work is required for most. However, the possibility of *Mobp* involvement in *du* pathology has been excluded by this study. It is also noteworthy that a new insertion mutation with a transient dysmyelination phenotype maps to distal chromosome 9 (Orian *et al.*, 1994). The corresponding region of synteny to distal mouse chromosome 9 in the human genome includes chromosome 3p21-22. The only identified inherited human neurological disorder associated with a gene in the 3p21-22 region is gangliosidosis, which has been shown to correlate with mutations in the gene encoding beta-galactosidase (Shows *et al.*, 1979). However, the human gene encoding *Mobp* will be a candidate for involvement in CNS neuropathies found to be linked with this region in the future.

2.3 Analysis of *Mobp* splicing patterns by sequence comparison

2.3.1 Comparison of all published *Mobp* sequences derived from rat data

Recent reports by Yamamoto *et al.* (1994) and Holz *et al.* (1996) have identified six splice variants of the *Mobp* transcript, in rat, generated by the novel gene *Mobp* encoding myelin-associated oligodendrocytic basic protein. Yamamoto *et al.* (1994) identified 2 messenger RNA species (rOP1 and rOPRP1) encoding polypeptides of 71 and 170 amino acids respectively. Subsequently, Holz *et al.* (1996) reported a further four splice variants (MOBP 69, 81A, 81B and 99) encoding polypeptides of 69, 81 and 99 amino acids respectively. The latter study, by Holz *et al.* (1996), notes that the isolated transcripts have their 5' regions in common. This common region encodes amino acids 1-68 of all the proteins predicted to be translated from the *Mobp* splice variants. The aim of this study was to assess whether any sequence units, other than that reported (Holz *et al.*, 1996), were shared amongst the published transcripts (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). This required the performance of sequence alignments (Figure 2.3.1) comparing all the rat *Mobp* transcript sequences published in the reports of Yamamoto *et al.* (1994) and Holz *et al.* (1996).

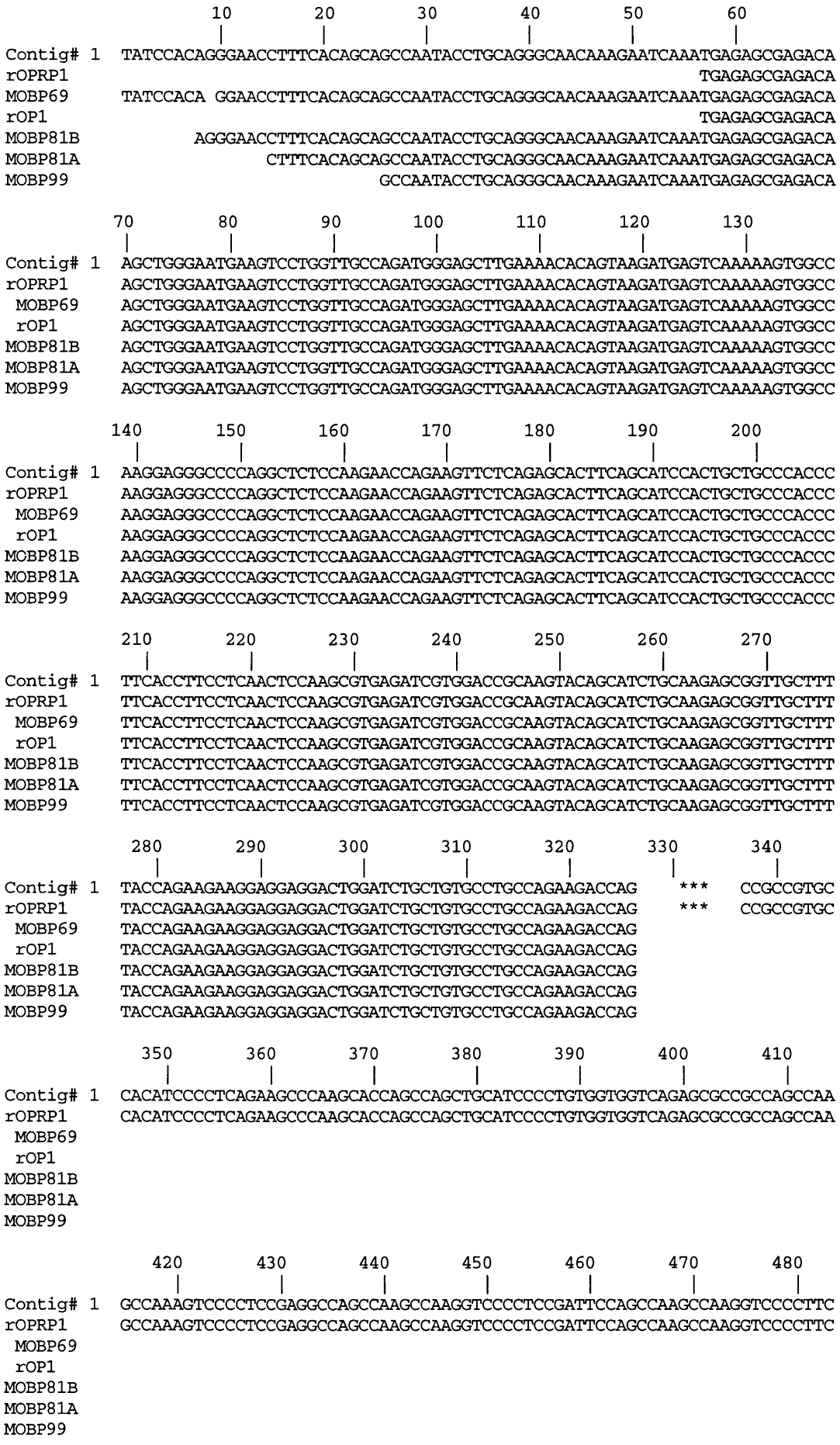
2.3.1.1 Generation of a sequence multi-alignment of published *Mobp* sequences derived from rat

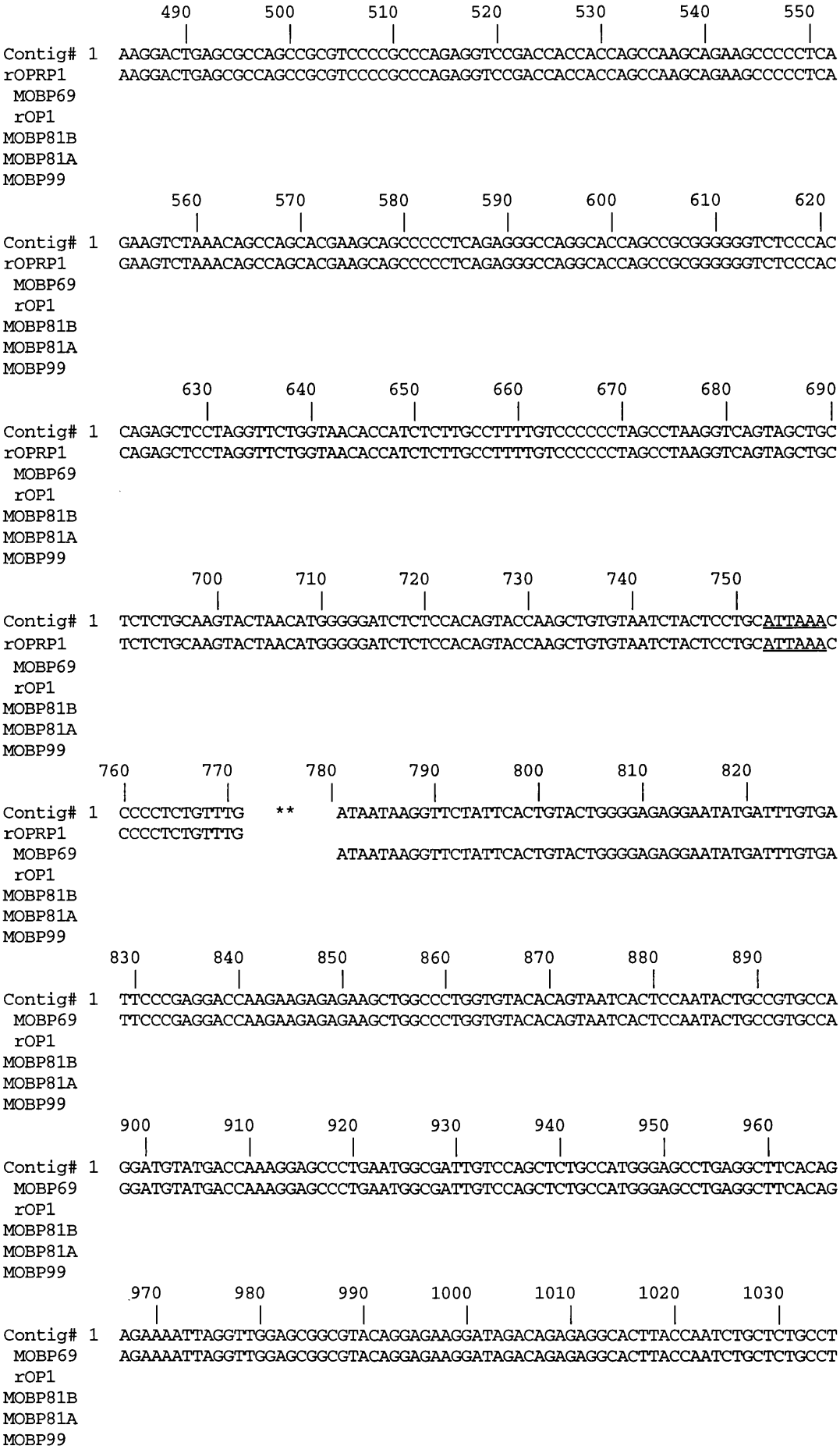
Sequences were imported to the Genejockey software (Cambridge, Biosoft), from the database according to their EMBL accession numbers (defined in table 2.3.1) and aligned by eye within the Genejockey sequence multialignment feature. Figure 2.3.1 illustrates the results of this analysis. Exonic sequence units are arbitrarily defined using the nucleotide numbers listed above the sequence and are identified in Table 2.3.2. Though the numbering of exonic sequence units (Q-Z) is purely arbitrary and cannot infer the exact order or number of exons comprising the *Mobp* gene, the order of some exonic segments may be deduced: R is downstream of Q; V is downstream of T and U; X is downstream of W; Y is downstream of V and Z is downstream of Y. Nothing, however, can be inferred about the linear relationship between H and I or indeed about the linear relationship between S and T-Z. Table 2.3.3 demonstrates that certain sequence units (V, X, Y, Z), downstream of B (encoding amino acids 1-68, shared by all MOBP isoforms), are shared by two or more of the published transcripts. Exonic segment V is transcribed in the generation of Mobp 71 (rOP1; Yamamoto *et al.*, 1994), Mobp 69, Mobp 81A and Mobp 81B; exonic segment X is transcribed in the generation of Mobp 69 and Mobp 81B

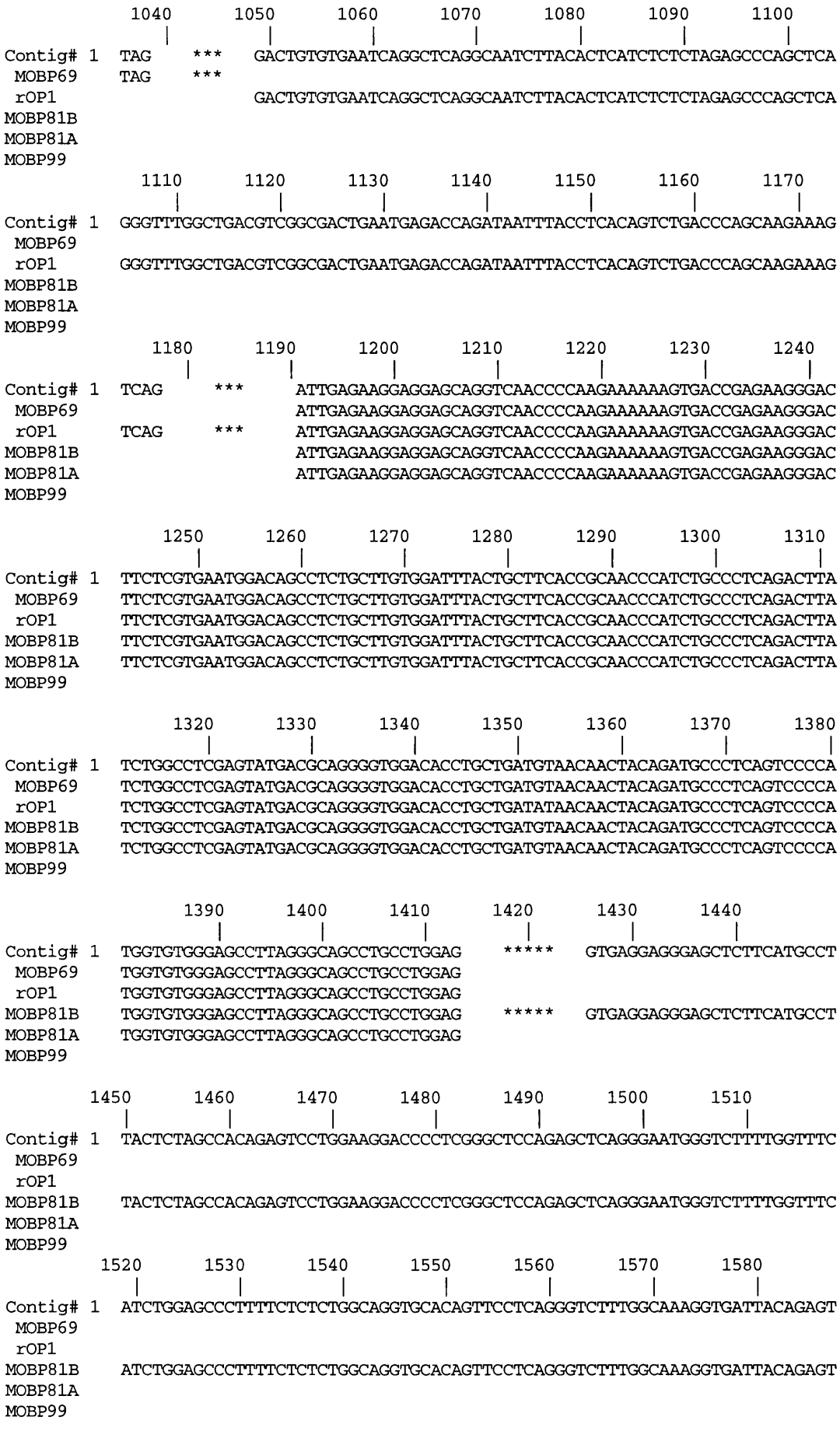
and exonic segment Z is transcribed in the generation of Mobp 71, Mobp 81A and Mobp 99. This result reinforces the suggestion that the *Mobp* gene undergoes a complex series of splicing events (Holz *et al.*, 1996) in order to generate the variety of protein isoforms and 3' UTRs predicted by transcript sequences (Yamamoto *et al.*, 1994 and Holz *et al.*, 1996).

2.3.1.2 Prediction of exon/exon junctions

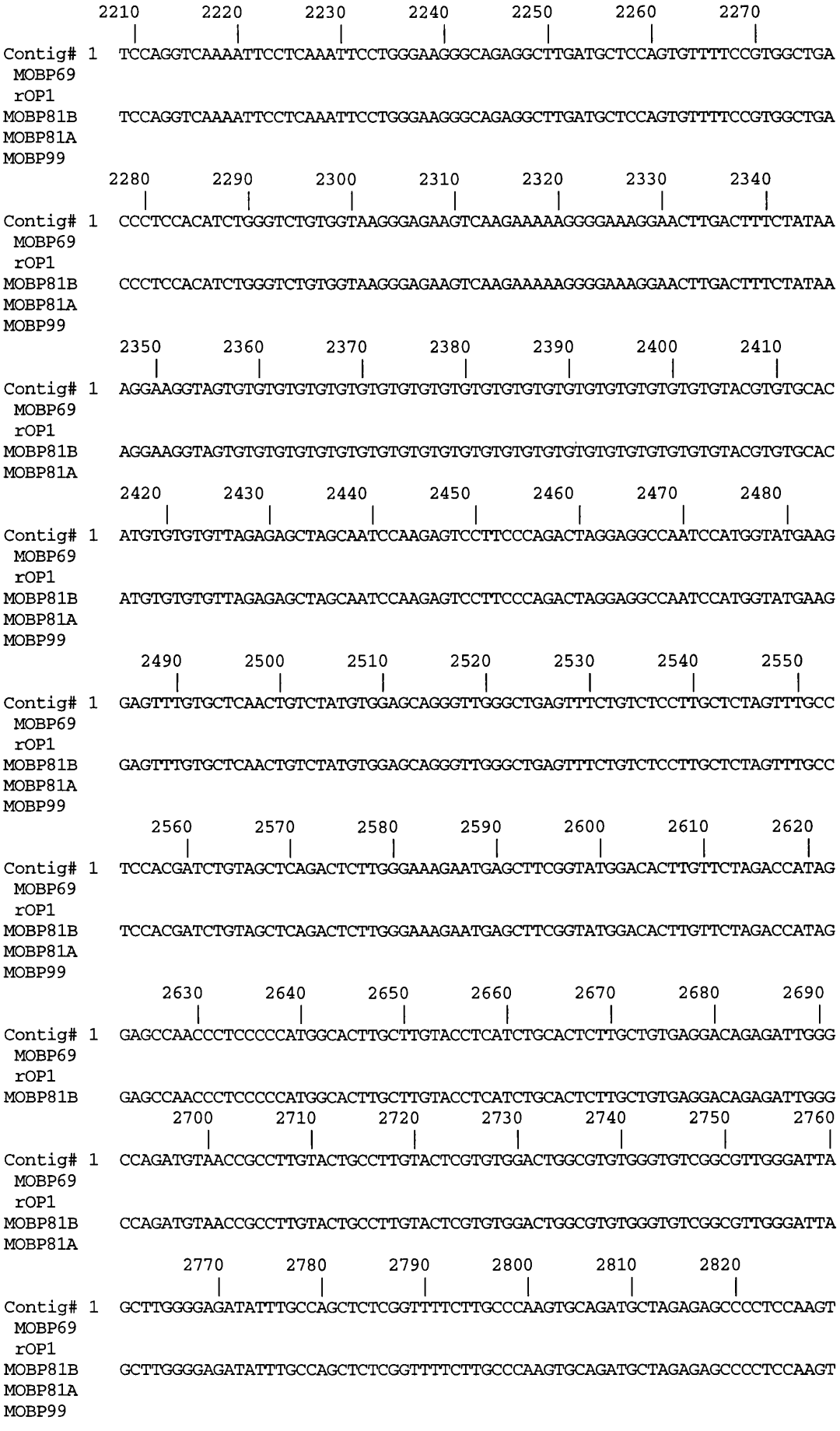
The 5' and 3' end sequences of the exonic segments (Q-Z) are defined in Table 2.3.4. It is noteworthy that four of the seven exonic segments (Q, R, U and Y), not terminating in polyadenylation signals, contain the exonic sequence portion (C/AAG|) of the extended 5' splice donor sequence (C/A)AG|GT(G/A)AGT (where the exon/intron boundary is denoted by the vertical line). The remaining three exonic segments (T, V and W) contain corrupted 5' donor sequences (T/GAG|) though it should be noted that this sequence is not highly conserved (Smith *et al.*, 1989). The DNA sequences across these putative exon/exon boundaries are defined in Table 2.3.4 and the encoded amino acid sequences are indicated in three-letter code. The amino acid sequence is numbered (1-68) from the first residue of the isoform MOBP 81 (Holz *et al.*, 1996) and from that point continues in accordance with the amino acid sequence of the individual isoform (Table 2.3.4). The predictions of exon/exon junctions, made on the basis of sequence comparison, indicate that R provides the first two nucleotides of the codon encoding amino acid residue 69, which may be Serine or Arginine dependent upon which downstream exonic segment, R is subsequently spliced to (Table 2.3.4).

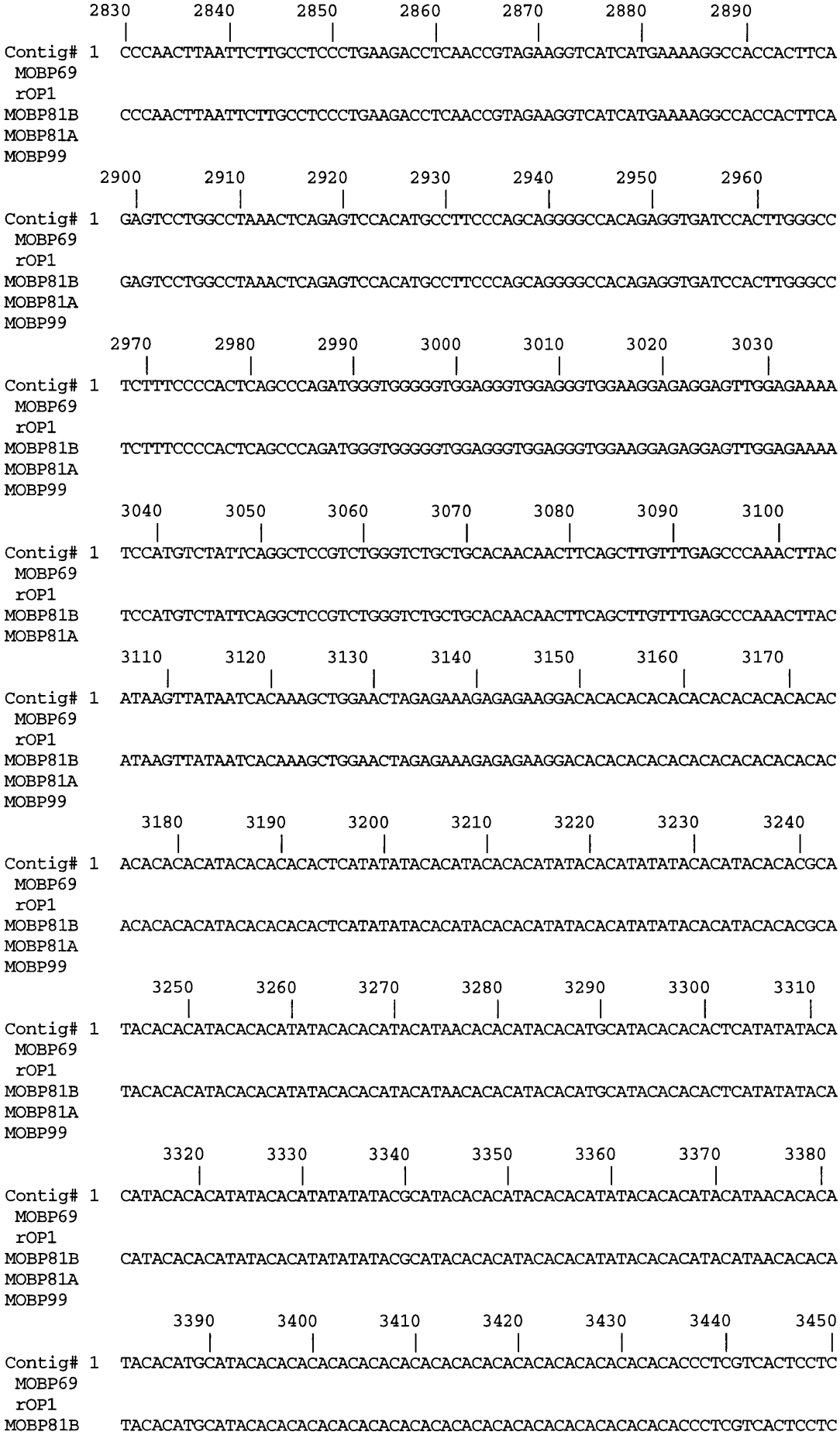

10 20 30 40 50 60
Contig# 1 TATCCACAGGGAACCTTTCACAGCAGCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGCGAGACA
rOPRP1 TGAGAGCGAGACA
MOBP69 TATCCACA GGAACCTTTCACAGCAGCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGCGAGACA
rOP1 TGAGAGCGAGACA
MOBP81B AGGGAACCTTTCACAGCAGCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGCGAGACA
MOBP81A CTTTCACAGCAGCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGCGAGACA
MOBP99 GCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGCGAGACA
70 80 90 100 110 120 130
Contig# 1 AGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAGTAAGATGAGTCAAAAAGTGGCC
rOPRP1 AGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAGTAAGATGAGTCAAAAAGTGGCC
MOBP69 AGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAGTAAGATGAGTCAAAAAGTGGCC
rOP1 AGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAGTAAGATGAGTCAAAAAGTGGCC
MOBP81B AGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAGTAAGATGAGTCAAAAAGTGGCC
MOBP81A AGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAGTAAGATGAGTCAAAAAGTGGCC
MOBP99 AGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAGTAAGATGAGTCAAAAAGTGGCC
140 150 160 170 180 190 200
Contig# 1 AAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGCACTTCAGCATCCACTGCTGCCCACCC
rOPRP1 AAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGCACTTCAGCATCCACTGCTGCCCACCC
MOBP69 AAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGCACTTCAGCATCCACTGCTGCCCACCC
rOP1 AAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGCACTTCAGCATCCACTGCTGCCCACCC
MOBP81B AAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGCACTTCAGCATCCACTGCTGCCCACCC
MOBP81A AAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGCACTTCAGCATCCACTGCTGCCCACCC
MOBP99 AAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGCACTTCAGCATCCACTGCTGCCCACCC
210 220 230 240 250 260 270
Contig# 1 TTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTTT
rOPRP1 TTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTTT
MOBP69 TTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTTT
rOP1 TTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTTT
MOBP81B TTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTTT
MOBP81A TTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTTT
MOBP99 TTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTTT
280 290 300 310 320 330 340
Contig# 1 TACCAGAAGAAGGAGGAGGACTGGATCTGCTGTGCCTGCCAGAAGACCAG *** CCGCCGTGC
rOPRP1 TACCAGAAGAAGGAGGAGGACTGGATCTGCTGTGCCTGCCAGAAGACCAG *** CCGCCGTGC
MOBP69 TACCAGAAGAAGGAGGAGGACTGGATCTGCTGTGCCTGCCAGAAGACCAG
rOP1 TACCAGAAGAAGGAGGAGGACTGGATCTGCTGTGCCTGCCAGAAGACCAG
MOBP81B TACCAGAAGAAGGAGGAGGACTGGATCTGCTGTGCCTGCCAGAAGACCAG
MOBP81A TACCAGAAGAAGGAGGAGGACTGGATCTGCTGTGCCTGCCAGAAGACCAG
MOBP99 TACCAGAAGAAGGAGGAGGACTGGATCTGCTGTGCCTGCCAGAAGACCAG
350 360 370 380 390 400 410
Contig# 1 CACATCCCCTCAGAAGCCCAAGCACCAGCCAGCTGCATCCCCTGTGGTGGTCAGAGCGCCGCCAGCCAA
rOPRP1 CACATCCCCTCAGAAGCCCAAGCACCAGCCAGCTGCATCCCCTGTGGTGGTCAGAGCGCCGCCAGCCAA
MOBP69
rOP1
MOBP81B
MOBP81A
MOBP99
420 430 440 450 460 470 480
Contig# 1 GCCAAAGTCCCCTCCGAGGCCAGCCAAGCCAAGGTCCCCTCCGATTCCAGCCAAGCCAAGGTCCCCTTC
rOPRP1 GCCAAAGTCCCCTCCGAGGCCAGCCAAGCCAAGGTCCCCTCCGATTCCAGCCAAGCCAAGGTCCCCTTC
MOBP69
rOP1
MOBP81B
MOBP81A
MOBP99

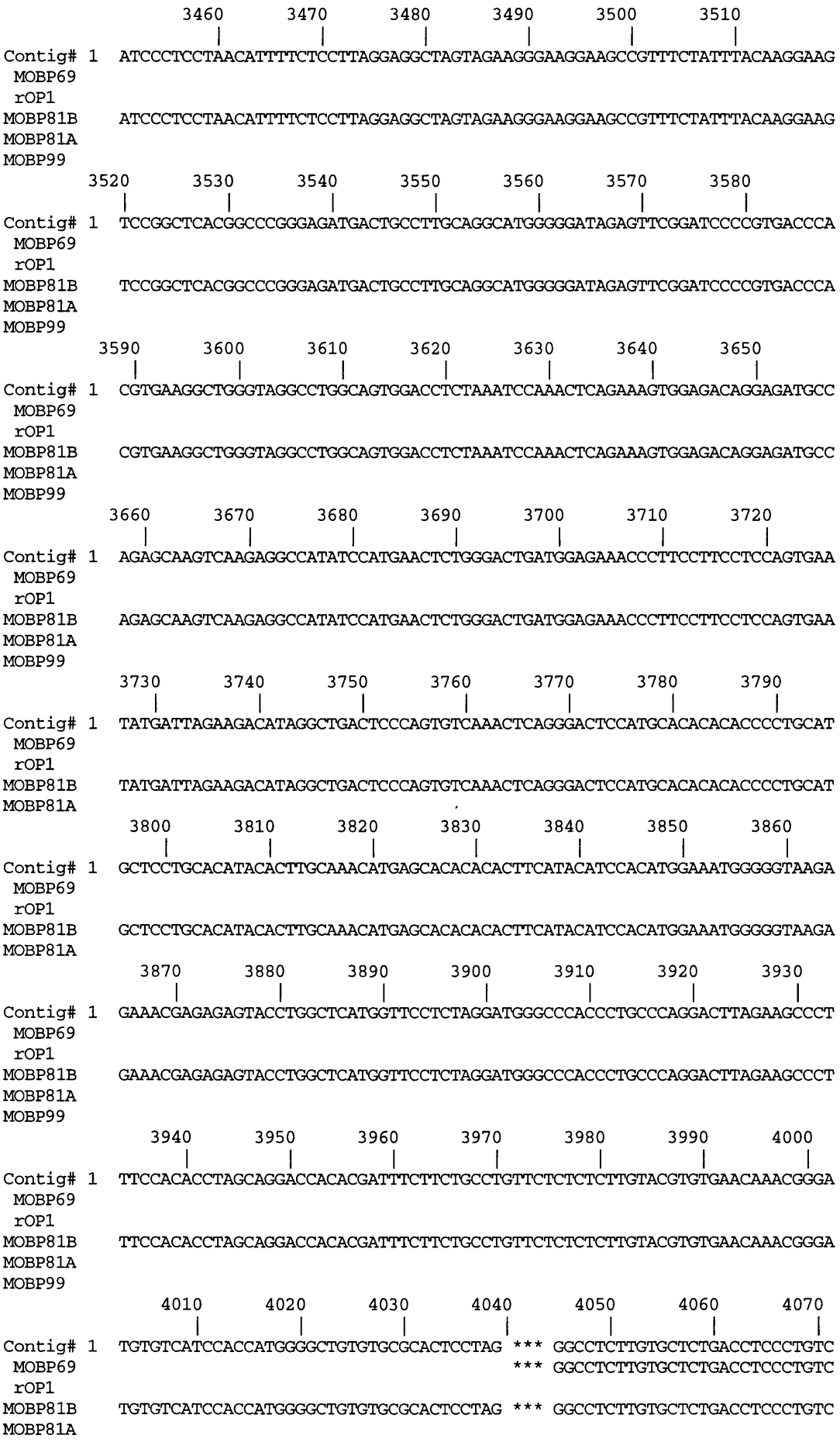

490 500 510 520 530 540 550
Contig# 1 AAGGACTGAGCGCCAGCCGCGTCCCCGCCCAGAGGTCCGACCACCACCAGCCAAGCAGAAGCCCCCTCA
rOPRP1 AAGGACTGAGCGCCAGCCGCGTCCCCGCCCAGAGGTCCGACCACCACCAGCCAAGCAGAAGCCCCCTCA
MOBP69
rOP1
MOBP81B
MOBP81A
MOBP99
560 570 580 590 600 610 620
Contig# 1 GAAGTCTAAACAGCCAGCACGAAGCAGCCCCCTCAGAGGGCCAGGCACCAGCCGCGGGGGGTCTCCCAC
rOPRP1 GAAGTCTAAACAGCCAGCACGAAGCAGCCCCCTCAGAGGGCCAGGCACCAGCCGCGGGGGGTCTCCCAC
MOBP69
rOP1
MOBP81B
MOBP81A
MOBP99
630 640 650 660 670 680 690
Contig# 1 CAGAGCTCCTAGGTTCTGGTAACACCATCTCTTGCCTTTTGTCCCCCCTAGCCTAAGGTCAGTAGCTGC
rOPRP1 CAGAGCTCCTAGGTTCTGGTAACACCATCTCTTGCCTTTTGTCCCCCCTAGCCTAAGGTCAGTAGCTGC
MOBP69
rOP1
MOBP81B
MOBP81A
MOBP99
700 710 720 730 740 750
Contig# 1 TCTCTGCAAGTACTAACATGGGGGATCTCTCCACAGTACCAAGCTGTGTAATCTACTCCTGCATTAAAC
rOPRP1 TCTCTGCAAGTACTAACATGGGGGATCTCTCCACAGTACCAAGCTGTGTAATCTACTCCTGCATTAAAC
MOBP69
rOP1
MOBP81B
MOBP81A
MOBP99
760 770 780 790 800 810 820
Contig# 1 CCCCTCTGTTTG ** ATAATAAGGTTCTATTCACTGTACTGGGGAGAGGAATATGATTTGTGA
rOPRP1 CCCCTCTGTTTG
MOBP69 ATAATAAGGTTCTATTCACTGTACTGGGGAGAGGAATATGATTTGTGA
rOP1
MOBP81B
MOBP81A
MOBP99
830 840 850 860 870 880 890
Contig# 1 TTCCCGAGGACCAAGAAGAGAGAAGCTGGCCCTGGTGTACACAGTAATCACTCCAATACTGCCGTGCCA
MOBP69 TTCCCGAGGACCAAGAAGAGAGAAGCTGGCCCTGGTGTACACAGTAATCACTCCAATACTGCCGTGCCA
rOP1
MOBP81B
MOBP81A
MOBP99
900 910 920 930 940 950 960
Contig# 1 GGATGTATGACCAAAGGAGCCCTGAATGGCGATTGTCCAGCTCTGCCATGGGAGCCTGAGGCTTCACAG
MOBP69 GGATGTATGACCAAAGGAGCCCTGAATGGCGATTGTCCAGCTCTGCCATGGGAGCCTGAGGCTTCACAG
rOP1
MOBP81B
MOBP81A
MOBP99
970 980 990 1000 1010 1020 1030
Contig# 1 AGAAAATTAGGTTGGAGCGGCGTACAGGAGAAGGATAGACAGAGAGGCACTTACCAATCTGCTCTGCCT
MOBP69 AGAAAATTAGGTTGGAGCGGCGTACAGGAGAAGGATAGACAGAGAGGCACTTACCAATCTGCTCTGCCT
rOP1
MOBP81B
MOBP81A
MOBP99

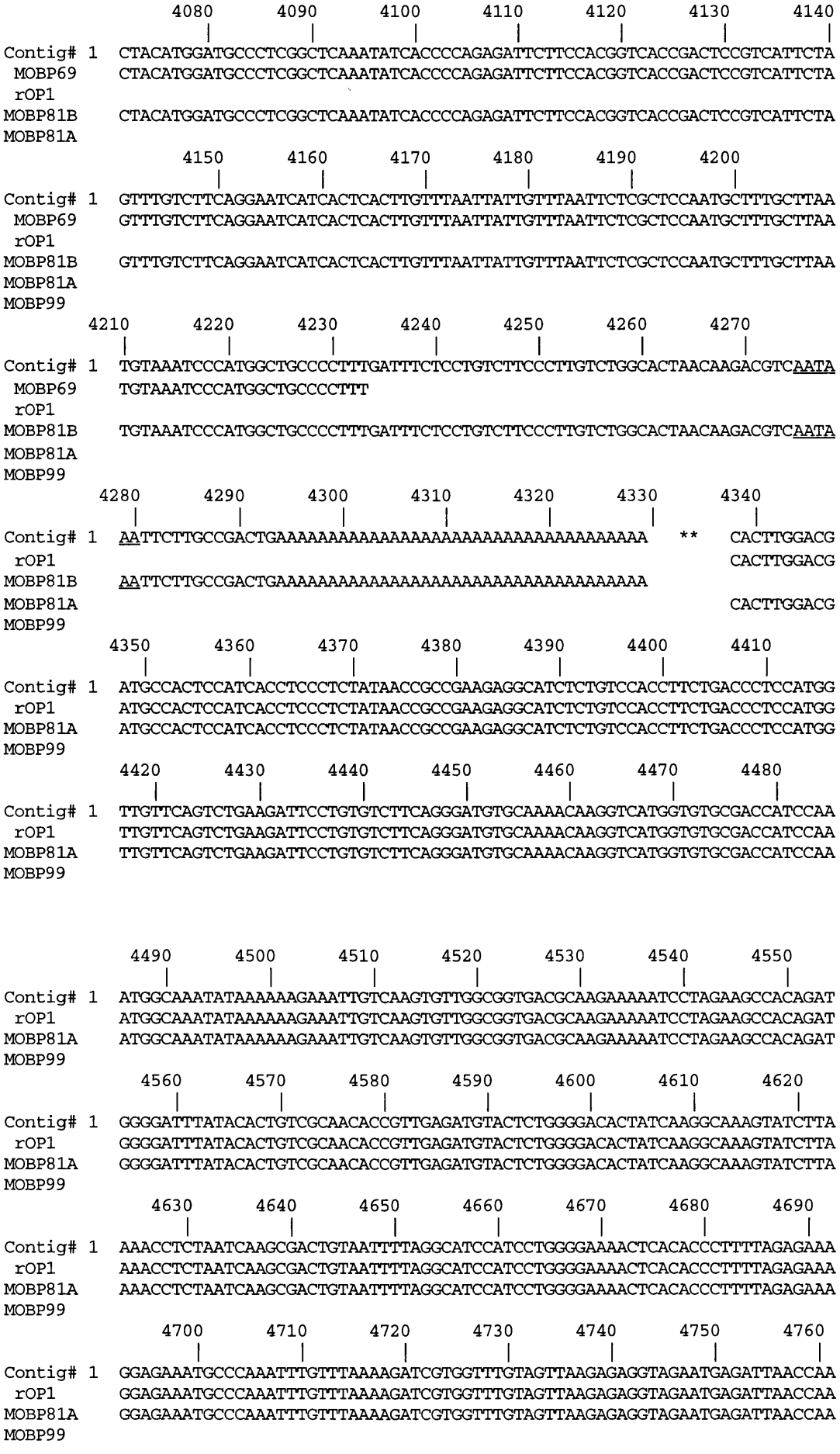

1040 1050 1060 1070 1080 1090 1100
Contig# 1 TAG *** GACTGTGTGAATCAGGCTCAGGCAATCTTACACTCATCTCTCTAGAGCCCAGCTCA
MOBP69 TAG ***
rOP1 GACTGTGTGAATCAGGCTCAGGCAATCTTACACTCATCTCTCTAGAGCCCAGCTCA
MOBP81B
MOBP81A
MOBP99
1110 1120 1130 1140 1150 1160 1170
Contig# 1 GGGTTTGGCTGACGTCGGCGACTGAATGAGACCAGATAATTTACCTCACAGTCTGACCCAGCAAGAAAG
MOBP69
rOP1 GGGTTTGGCTGACGTCGGCGACTGAATGAGACCAGATAATTTACCTCACAGTCTGACCCAGCAAGAAAG
MOBP81B
MOBP81A
MOBP99
1180 1190 1200 1210 1220 1230 1240
Contig# 1 TCAG *** ATTGAGAAGGAGGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGAC
MOBP69 ATTGAGAAGGAGGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGAC
rOP1 TCAG *** ATTGAGAAGGAGGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGAC
MOBP81B ATTGAGAAGGAGGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGAC
MOBP81A ATTGAGAAGGAGGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGAC
MOBP99
1250 1260 1270 1280 1290 1300 1310
Contig# 1 TTCTCGTGAATGGACAGCCTCTGCTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTA
MOBP69 TTCTCGTGAATGGACAGCCTCTGCTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTA
rOP1 TTCTCGTGAATGGACAGCCTCTGCTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTA
MOBP81B TTCTCGTGAATGGACAGCCTCTGCTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTA
MOBP81A TTCTCGTGAATGGACAGCCTCTGCTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTA
MOBP99
1320 1330 1340 1350 1360 1370 1380
Contig# 1 TCTGGCCTCGAGTATGACGCAGGGGTGGACACCTGCTGATGTAACAACTACAGATGCCCTCAGTCCCCA
MOBP69 TCTGGCCTCGAGTATGACGCAGGGGTGGACACCTGCTGATGTAACAACTACAGATGCCCTCAGTCCCCA
rOP1 TCTGGCCTCGAGTATGACGCAGGGGTGGACACCTGCTGATATAACAACTACAGATGCCCTCAGTCCCCA
MOBP81B TCTGGCCTCGAGTATGACGCAGGGGTGGACACCTGCTGATGTAACAACTACAGATGCCCTCAGTCCCCA
MOBP81A TCTGGCCTCGAGTATGACGCAGGGGTGGACACCTGCTGATGTAACAACTACAGATGCCCTCAGTCCCCA
MOBP99
1390 1400 1410 1420 1430 1440
Contig# 1 TGGTGTGGGAGCCTTAGGGCAGCCTGCCTGGAG ***** GTGAGGAGGGAGCTCTTCATGCCT
MOBP69 TGGTGTGGGAGCCTTAGGGCAGCCTGCCTGGAG
rOP1 TGGTGTGGGAGCCTTAGGGCAGCCTGCCTGGAG
MOBP81B TGGTGTGGGAGCCTTAGGGCAGCCTGCCTGGAG ***** GTGAGGAGGGAGCTCTTCATGCCT
MOBP81A TGGTGTGGGAGCCTTAGGGCAGCCTGCCTGGAG
MOBP99
1450 1460 1470 1480 1490 1500 1510
Contig# 1 TACTCTAGCCACAGAGTCCTGGAAGGACCCCTCGGGCTCCAGAGCTCAGGGAATGGGTCTTTTGGTTTC
MOBP69
rOP1
MOBP81B TACTCTAGCCACAGAGTCCTGGAAGGACCCCTCGGGCTCCAGAGCTCAGGGAATGGGTCTTTTGGTTTC
MOBP81A
MOBP99
1520 1530 1540 1550 1560 1570 1580
Contig# 1 ATCTGGAGCCCTTTTCTCTCTGGCAGGTGCACAGTTCCTCAGGGTCTTTGGCAAAGGTGATTACAGAGT
MOBP69
rOP1
MOBP81B ATCTGGAGCCCTTTTCTCTCTGGCAGGTGCACAGTTCCTCAGGGTCTTTGGCAAAGGTGATTACAGAGT
MOBP81A
MOBP99

```
          1590       1600       1610       1620       1630       1640       1650
             |          |          |          |          |          |          |
Contig# 1  CAATCCACACACATTCCTCCTCCTTCCAATTCCCCTCTCTCCCTGGGTCAGGTTGACAACCTGGCTCTT
 MOBP69
 rOP1
MOBP81B    CAATCCACACACATTCCTCCTCCTTCCAATTCCCCTCTCTCCCTGGGTCAGGTTGACAACCTGGCTCTT
MOBP81A
MOBP99

          1660       1670       1680       1690       1700       1710       1720
             |          |          |          |          |          |          |
Contig# 1  TTTCCTTTCCTTCCTAATTCTCTGCCTGGCCTCTCTTGATCACGTGACATCTGTCTCCCATTGACCTTC
 MOBP69
 rOP1
MOBP81B    TTTCCTTTCCTTCCTAATTCTCTGCCTGGCCTCTCTTGATCACGTGACATCTGTCTCCCATTGACCTTC
MOBP81A

          1730       1740       1750       1760       1770       1780       1790
             |          |          |          |          |          |          |
Contig# 1  TTAATCAGGAGTGGTCAGAGCCAACGTCCACACTGGTGAAGGCTCAGGTTATTTTTGGAGTAAGATACA
 MOBP69
 rOP1
MOBP81B    TTAATCAGGAGTGGTCAGAGCCAACGTCCACACTGGTGAAGGCTCAGGTTATTTTTGGAGTAAGATACA
MOBP81A
MOBP99

          1800       1810       1820       1830       1840       1850       1860
             |          |          |          |          |          |          |
Contig# 1  GAGTTGCTGTATGATCCTTAAATGCAGCAATGTTCTCATAAGAATGGGATCCCGTTAAATCTTCACGTC
 MOBP69
 rOP1
MOBP81B    GAGTTGCTGTATGATCCTTAAATGCAGCAATGTTCTCATAAGAATGGGATCCCGTTAAATCTTCACGTC
MOBP81A
MOBP99

          1870       1880       1890       1900       1910       1920       1930
             |          |          |          |          |          |          |
Contig# 1  AACTGTAGGTAACTTGTGTCAGAAACAGAGAAAAGGGGGCATCTTCCTCCCTCCCCCTCTGCAGGCCCA
 MOBP69
 rOP1
MOBP81B    AACTGTAGGTAACTTGTGTCAGAAACAGAGAAAAGGGGGCATCTTCCTCCCTCCCCCTCTGCAGGCCCA
MOBP81A
MOBP99

          1940       1950       1960       1970       1980       1990       2000
             |          |          |          |          |          |          |
Contig# 1  GTCTCCAGCGTTACACAGGGCTTTGTGAGCACCAGCTCCTGCCCCAGTCTGGGGATAACAACCCTCTAA
 MOBP69
 rOP1
MOBP81B    GTCTCCAGCGTTACACAGGGCTTTGTGAGCACCAGCTCCTGCCCCAGTCTGGGGATAACAACCCTCTAA
MOBP81A
MOBP99

          2010       2020       2030       2040       2050       2060       2070
             |          |          |          |          |          |          |
Contig# 1  ACACTGACTTTGACTTGTCCCCAGTGCAGTTGGCACTCAGCCTCACAGACTCTCAGTGCAGCTCTGACT
 MOBP69
 rOP1
MOBP81B    ACACTGACTTTGACTTGTCCCCAGTGCAGTTGGCACTCAGCCTCACAGACTCTCAGTGCAGCTCTGACT
          2080       2090       2100       2110       2120       2130
             |          |          |          |          |          |
Contig# 1  GCTCTCAATTATATAATTTAGGAAGGTCATGTGGCTTCAGCCCTGTCCCCCATTCCCCATCACCTCCCT
 MOBP69
 rOP1
MOBP81B    GCTCTCAATTATATAATTTAGGAAGGTCATGTGGCTTCAGCCCTGTCCCCCATTCCCCATCACCTCCCT
MOBP81A
MOBP99

        2140       2150       2160       2170       2180       2190       2200
          |          |          |          |          |          |          |
Contig# 1  TGAGGCATCCTGGACAAGTCCCATGTGCACTTGTATAGGTCTAGAAAGAATGCTTGGAGAGTTTCTGTC
 MOBP69
 rOP1
MOBP81B    TGAGGCATCCTGGACAAGTCCCATGTGCACTTGTATAGGTCTAGAAAGAATGCTTGGAGAGTTTCTGTC
MOBP81A
MOBP99
```

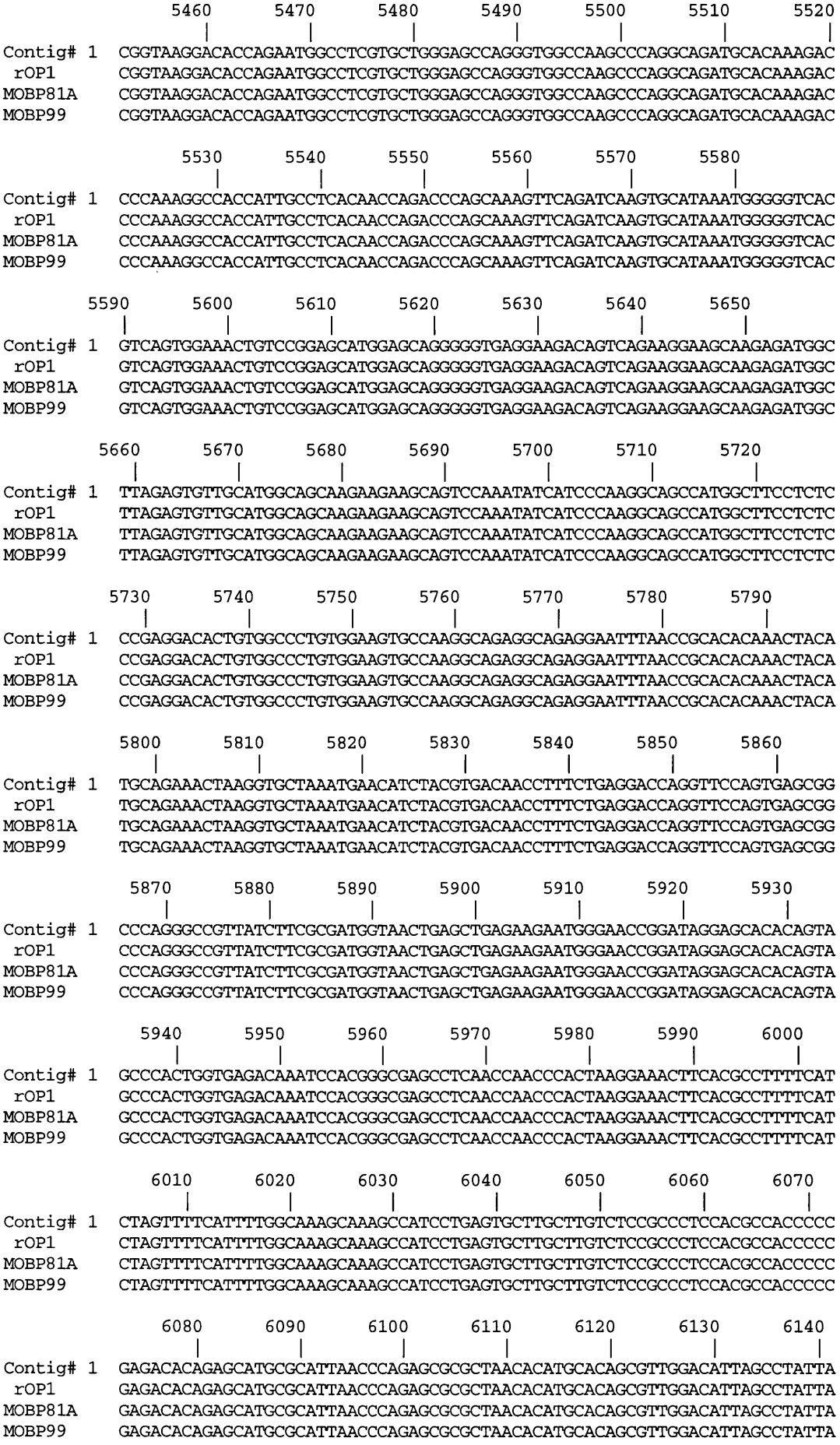

2210 2220 2230 2240 2250 2260 2270
Contig# 1 TCCAGGTCAAAATTCCTCAAATTCCTGGGAAGGGCAGAGGCTTGATGCTCCAGTGTTTTCCGTGGCTGA
MOBP69
rOP1
MOBP81B TCCAGGTCAAAATTCCTCAAATTCCTGGGAAGGGCAGAGGCTTGATGCTCCAGTGTTTTCCGTGGCTGA
MOBP81A
MOBP99
2280 2290 2300 2310 2320 2330 2340
Contig# 1 CCCTCCACATCTGGGTCTGTGGTAAGGGAGAAGTCAAGAAAAAGGGGAAAGGAACTTGACTTTCTATAA
MOBP69
rOP1
MOBP81B CCCTCCACATCTGGGTCTGTGGTAAGGGAGAAGTCAAGAAAAAGGGGAAAGGAACTTGACTTTCTATAA
MOBP81A
MOBP99
2350 2360 2370 2380 2390 2400 2410
Contig# 1 AGGAAGGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTACGTGTGCAC
MOBP69
rOP1
MOBP81B AGGAAGGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTACGTGTGCAC
MOBP81A
2420 2430 2440 2450 2460 2470 2480
Contig# 1 ATGTGTGTGTTAGAGAGCTAGCAATCCAAGAGTCCTTCCCAGACTAGGAGGCCAATCCATGGTATGAAG
MOBP69
rOP1
MOBP81B ATGTGTGTGTTAGAGAGCTAGCAATCCAAGAGTCCTTCCCAGACTAGGAGGCCAATCCATGGTATGAAG
MOBP81A
MOBP99
2490 2500 2510 2520 2530 2540 2550
Contig# 1 GAGTTTGTGCTCAACTGTCTATGTGGAGCAGGGTTGGGCTGAGTTTCTGTCTCCTTGCTCTAGTTTGCC
MOBP69
rOP1
MOBP81B GAGTTTGTGCTCAACTGTCTATGTGGAGCAGGGTTGGGCTGAGTTTCTGTCTCCTTGCTCTAGTTTGCC
MOBP81A
MOBP99
2560 2570 2580 2590 2600 2610 2620
Contig# 1 TCCACGATCTGTAGCTCAGACTCTTGGGAAAGAATGAGCTTCGGTATGGACACTTGTTCTAGACCATAG
MOBP69
rOP1
MOBP81B TCCACGATCTGTAGCTCAGACTCTTGGGAAAGAATGAGCTTCGGTATGGACACTTGTTCTAGACCATAG
MOBP81A
MOBP99
2630 2640 2650 2660 2670 2680 2690
Contig# 1 GAGCCAACCCTCCCCCATGGCACTTGCTTGTACCTCATCTGCACTCTTGCTGTGAGGACAGAGATTGGG
MOBP69
rOP1
MOBP81B GAGCCAACCCTCCCCCATGGCACTTGCTTGTACCTCATCTGCACTCTTGCTGTGAGGACAGAGATTGGG
2700 2710 2720 2730 2740 2750 2760
Contig# 1 CCAGATGTAACCGCCTTGTACTGCCTTGTACTCGTGTGGACTGGCGTGTGGGTGTCGGCGTTGGGATTA
MOBP69
rOP1
MOBP81B CCAGATGTAACCGCCTTGTACTGCCTTGTACTCGTGTGGACTGGCGTGTGGGTGTCGGCGTTGGGATTA
MOBP81A
2770 2780 2790 2800 2810 2820
Contig# 1 GCTTGGGGAGATATTTGCCAGCTCTCGGTTTTCTTGCCCAAGTGCAGATGCTAGAGAGCCCCTCCAAGT
MOBP69
rOP1
MOBP81B GCTTGGGGAGATATTTGCCAGCTCTCGGTTTTCTTGCCCAAGTGCAGATGCTAGAGAGCCCCTCCAAGT
MOBP81A
MOBP99


2830 2840 2850 2860 2870 2880 2890
Contig# 1 CCCAACTTAATTCTTGCCTCCCTGAAGACCTCAACCGTAGAAGGTCATCATGAAAAGGCCACCACTTCA
MOBP69
rOP1
MOBP81B CCCAACTTAATTCTTGCCTCCCTGAAGACCTCAACCGTAGAAGGTCATCATGAAAAGGCCACCACTTCA
MOBP81A
MOBP99
2900 2910 2920 2930 2940 2950 2960
Contig# 1 GAGTCCTGGCCTAAACTCAGAGTCCACATGCCTTCCCAGCAGGGGCCACAGAGGTGATCCACTTGGGCC
MOBP69
rOP1
MOBP81B GAGTCCTGGCCTAAACTCAGAGTCCACATGCCTTCCCAGCAGGGGCCACAGAGGTGATCCACTTGGGCC
MOBP81A
MOBP99
2970 2980 2990 3000 3010 3020 3030
Contig# 1 TCTTTCCCCACTCAGCCCAGATGGGTGGGGGTGGAGGGTGGAGGGTGGAAGGAGAGGAGTTGGAGAAAA
MOBP69
rOP1
MOBP81B TCTTTCCCCACTCAGCCCAGATGGGTGGGGGTGGAGGGTGGAGGGTGGAAGGAGAGGAGTTGGAGAAAA
MOBP81A
MOBP99
3040 3050 3060 3070 3080 3090 3100
Contig# 1 TCCATGTCTATTCAGGCTCCGTCTGGGTCTGCTGCACAACAACTTCAGCTTGTTTGAGCCCAAACTTAC
MOBP69
rOP1
MOBP81B TCCATGTCTATTCAGGCTCCGTCTGGGTCTGCTGCACAACAACTTCAGCTTGTTTGAGCCCAAACTTAC
MOBP81A
3110 3120 3130 3140 3150 3160 3170
Contig# 1 ATAAGTTATAATCACAAAGCTGGAACTAGAGAAAGAGAGAAGGACACACACACACACACACACACACAC
MOBP69
rOP1
MOBP81B ATAAGTTATAATCACAAAGCTGGAACTAGAGAAAGAGAGAAGGACACACACACACACACACACACACAC
MOBP81A
MOBP99
3180 3190 3200 3210 3220 3230 3240
Contig# 1 ACACACACATACACACACACTCATATATACACATACACACATATACACATATATACACATACACACGCA
MOBP69
rOP1
MOBP81B ACACACACATACACACACACTCATATATACACATACACACATATACACATATATACACATACACACGCA
MOBP81A
MOBP99
3250 3260 3270 3280 3290 3300 3310
Contig# 1 TACACACATACACACATATACACACATACATAACACACATACACATGCATACACACACTCATATATACA
MOBP69
rOP1
MOBP81B TACACACATACACACATATACACACATACATAACACACATACACATGCATACACACACTCATATATACA
MOBP81A
MOBP99
3320 3330 3340 3350 3360 3370 3380
Contig# 1 CATACACACATATACACATATATATACGCATACACACATACACACATATACACACATACATAACACACA
MOBP69
rOP1
MOBP81B CATACACACATATACACATATATATACGCATACACACATACACACATATACACACATACATAACACACA
MOBP81A
MOBP99
3390 3400 3410 3420 3430 3440 3450
Contig# 1 TACACATGCATACACACACACACACACACACACACACACACACACACACACACACACCCTCGTCACTCCTC
MOBP69
rOP1
MOBP81B TACACATGCATACACACACACACACACACACACACACACACACACACACACACACACCCTCGTCACTCCTC


3460 3470 3480 3490 3500 3510
Contig# 1 ATCCCTCCTAACATTTTCTCCTTAGGAGGCTAGTAGAAGGGAAGGAAGCCGTTTCTATTTACAAGGAAG
MOBP69
rOP1
MOBP81B ATCCCTCCTAACATTTTCTCCTTAGGAGGCTAGTAGAAGGGAAGGAAGCCGTTTCTATTTACAAGGAAG
MOBP81A
MOBP99
3520 3530 3540 3550 3560 3570 3580
Contig# 1 TCCGGCTCACGGCCCGGGAGATGACTGCCTTGCAGGCATGGGGGATAGAGTTCGGATCCCCGTGACCCA
MOBP69
rOP1
MOBP81B TCCGGCTCACGGCCCGGGAGATGACTGCCTTGCAGGCATGGGGGATAGAGTTCGGATCCCCGTGACCCA
MOBP81A
MOBP99
3590 3600 3610 3620 3630 3640 3650
Contig# 1 CGTGAAGGCTGGGTAGGCCTGGCAGTGGACCTCTAAATCCAAACTCAGAAAGTGGAGACAGGAGATGCC
MOBP69
rOP1
MOBP81B CGTGAAGGCTGGGTAGGCCTGGCAGTGGACCTCTAAATCCAAACTCAGAAAGTGGAGACAGGAGATGCC
MOBP81A
MOBP99
3660 3670 3680 3690 3700 3710 3720
Contig# 1 AGAGCAAGTCAAGAGGCCATATCCATGAACTCTGGGACTGATGGAGAAACCCTTCCTTCCTCCAGTGAA
MOBP69
rOP1
MOBP81B AGAGCAAGTCAAGAGGCCATATCCATGAACTCTGGGACTGATGGAGAAACCCTTCCTTCCTCCAGTGAA
MOBP81A
MOBP99
3730 3740 3750 3760 3770 3780 3790
Contig# 1 TATGATTAGAAGACATAGGCTGACTCCCAGTGTCAAACTCAGGGACTCCATGCACACACACCCCTGCAT
MOBP69
rOP1
MOBP81B TATGATTAGAAGACATAGGCTGACTCCCAGTGTCAAACTCAGGGACTCCATGCACACACACCCCTGCAT
MOBP81A
3800 3810 3820 3830 3840 3850 3860
Contig# 1 GCTCCTGCACATACACTTGCAAACATGAGCACACACACTTCATACATCCACATGGAAATGGGGGTAAGA
MOBP69
rOP1
MOBP81B GCTCCTGCACATACACTTGCAAACATGAGCACACACACTTCATACATCCACATGGAAATGGGGGTAAGA
MOBP81A
3870 3880 3890 3900 3910 3920 3930
Contig# 1 GAAACGAGAGAGTACCTGGCTCATGGTTCCTCTAGGATGGGCCCACCCTGCCCAGGACTTAGAAGCCCT
MOBP69
rOP1
MOBP81B GAAACGAGAGAGTACCTGGCTCATGGTTCCTCTAGGATGGGCCCACCCTGCCCAGGACTTAGAAGCCCT
MOBP81A
MOBP99
3940 3950 3960 3970 3980 3990 4000
Contig# 1 TTCCACACCTAGCAGGACCACACGATTTCTTCTGCCTGTTCTCTCTCTTGTACGTGTGAACAAACGGGA
MOBP69
rOP1
MOBP81B TTCCACACCTAGCAGGACCACACGATTTCTTCTGCCTGTTCTCTCTCTTGTACGTGTGAACAAACGGGA
MOBP81A
MOBP99
4010 4020 4030 4040 4050 4060 4070
Contig# 1 TGTGTCATCCACCATGGGGCTGTGTGCGCACTCCTAG *** GGCCTCTTGTGCTCTGACCTCCCTGTC
MOBP69 *** GGCCTCTTGTGCTCTGACCTCCCTGTC
rOP1
MOBP81B TGTGTCATCCACCATGGGGCTGTGTGCGCACTCCTAG *** GGCCTCTTGTGCTCTGACCTCCCTGTC
MOBP81A


4080 4090 4100 4110 4120 4130 4140
Contig# 1 CTACATGGATGCCCTCGGCTCAAATATCACCCCAGAGATTCTTCCACGGTCACCGACTCCGTCATTCTA
MOBP69 CTACATGGATGCCCTCGGCTCAAATATCACCCCAGAGATTCTTCCACGGTCACCGACTCCGTCATTCTA
rOP1
MOBP81B CTACATGGATGCCCTCGGCTCAAATATCACCCCAGAGATTCTTCCACGGTCACCGACTCCGTCATTCTA
MOBP81A
4150 4160 4170 4180 4190 4200
Contig# 1 GTTTGTCTTCAGGAATCATCACTCACTTGTTTAATTATTGTTTAATTCTCGCTCCAATGCTTTGCTTAA
MOBP69 GTTTGTCTTCAGGAATCATCACTCACTTGTTTAATTATTGTTTAATTCTCGCTCCAATGCTTTGCTTAA
rOP1
MOBP81B GTTTGTCTTCAGGAATCATCACTCACTTGTTTAATTATTGTTTAATTCTCGCTCCAATGCTTTGCTTAA
MOBP81A
MOBP99
4210 4220 4230 4240 4250 4260 4270
Contig# 1 TGTAAATCCCATGGCTGCCCCTTTGATTTCTCCTGTCTTCCCTTGTCTGGCACTAACAAGACGTCAATA
MOBP69 TGTAAATCCCATGGCTGCCCCTTT
rOP1
MOBP81B TGTAAATCCCATGGCTGCCCCTTTGATTTCTCCTGTCTTCCCTTGTCTGGCACTAACAAGACGTCAATA
MOBP81A
MOBP99
4280 4290 4300 4310 4320 4330 4340
Contig# 1 AATTCTTGCCGACTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA ** CACTTGGACG
rOP1 CACTTGGACG
MOBP81B AATTCTTGCCGACTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
MOBP81A CACTTGGACG
MOBP99
4350 4360 4370 4380 4390 4400 4410
Contig# 1 ATGCCACTCCATCACCTCCCTCTATAACCGCCGAAGAGGCATCTCTGTCCACCTTCTGACCCTCCATGG
rOP1 ATGCCACTCCATCACCTCCCTCTATAACCGCCGAAGAGGCATCTCTGTCCACCTTCTGACCCTCCATGG
MOBP81A ATGCCACTCCATCACCTCCCTCTATAACCGCCGAAGAGGCATCTCTGTCCACCTTCTGACCCTCCATGG
MOBP99
4420 4430 4440 4450 4460 4470 4480
Contig# 1 TTGTTCAGTCTGAAGATTCCTGTGTCTTCAGGGATGTGCAAAACAAGGTCATGGTGTGCGACCATCCAA
rOP1 TTGTTCAGTCTGAAGATTCCTGTGTCTTCAGGGATGTGCAAAACAAGGTCATGGTGTGCGACCATCCAA
MOBP81A TTGTTCAGTCTGAAGATTCCTGTGTCTTCAGGGATGTGCAAAACAAGGTCATGGTGTGCGACCATCCAA
MOBP99
4490 4500 4510 4520 4530 4540 4550
Contig# 1 ATGGCAAATATAAAAAAGAAATTGTCAAGTGTTGGCGGTGACGCAAGAAAAATCCTAGAAGCCACAGAT
rOP1 ATGGCAAATATAAAAAAGAAATTGTCAAGTGTTGGCGGTGACGCAAGAAAAATCCTAGAAGCCACAGAT
MOBP81A ATGGCAAATATAAAAAAGAAATTGTCAAGTGTTGGCGGTGACGCAAGAAAAATCCTAGAAGCCACAGAT
MOBP99
4560 4570 4580 4590 4600 4610 4620
Contig# 1 GGGGATTTATACACTGTCGCAACACCGTTGAGATGTACTCTGGGGACACTATCAAGGCAAAGTATCTTA
rOP1 GGGGATTTATACACTGTCGCAACACCGTTGAGATGTACTCTGGGGACACTATCAAGGCAAAGTATCTTA
MOBP81A GGGGATTTATACACTGTCGCAACACCGTTGAGATGTACTCTGGGGACACTATCAAGGCAAAGTATCTTA
MOBP99
4630 4640 4650 4660 4670 4680 4690
Contig# 1 AAACCTCTAATCAAGCGACTGTAATTTTAGGCATCCATCCTGGGGAAAACTCACACCCTTTTAGAGAAA
rOP1 AAACCTCTAATCAAGCGACTGTAATTTTAGGCATCCATCCTGGGGAAAACTCACACCCTTTTAGAGAAA
MOBP81A AAACCTCTAATCAAGCGACTGTAATTTTAGGCATCCATCCTGGGGAAAACTCACACCCTTTTAGAGAAA
MOBP99
4700 4710 4720 4730 4740 4750 4760
Contig# 1 GGAGAAATGCCCAAATTTGTTTAAAAGATCGTGGTTTGTAGTTAAGAGAGGTAGAATGAGATTAACCAA
rOP1 GGAGAAATGCCCAAATTTGTTTAAAAGATCGTGGTTTGTAGTTAAGAGAGGTAGAATGAGATTAACCAA
MOBP81A GGAGAAATGCCCAAATTTGTTTAAAAGATCGTGGTTTGTAGTTAAGAGAGGTAGAATGAGATTAACCAA
MOBP99

```
                4770       4780       4790       4800       4810       4820       4830
                  |          |          |          |          |          |          |
Contig# 1  AATGTCGGTCCACGGGAGAGCAAGAGAGTCAATCTTGGTATGTCCACAAAGTCCTAAGAGCAGTTAAAA
 rOP1      AATGTCGGTCCACGGGAGAGCAAGAGAGTCAATCTTGGTATGTCCACAAAGTCCTAAGAGCAGTTAAAA
MOBP81A    AATGTCGGTCCACGGGAGAGCAAGAGAGTCAATCTTGGTATGTCCACAAAGTCCTAAGAGCAGTTAAAA
MOBP99

                4840       4850       4860       4870       4880       4890
                  |          |          |          |          |          |
Contig# 1  GCCAATCTGTCAGACAGACCCATGCCGAGCAGGTAAATCTAGAAAGCAAACCAAGGACTATAAAAGTCA
 rOP1      GCCAATCTGTCAGACAGACCCATGCCGAGCAGGTAAATCTAGAAAGCAAACCAAGGACTATAAAAGTCA
MOBP81A    GCCAATCTGTCAGACAGACCCATGCCGAGCAGGTAAATCTAGAAAGCAAACCAAGGACTATAAAAGTCA
MOBP99

          4900       4910       4920       4930       4940       4950       4960
            |          |          |          |          |          |          |
Contig# 1  AGCACCAGGCAGTAAATCCAG|AATAAGAGCCTATGCATACATATTGACAGCTCACAATCCTGTACTGGT
 rOP1      AGCACCAGGCAGTAAATCCAG|AATAAGAGCCTATGCATACATATTGACAGCT ACAATCCTGTACTGGT
MOBP81A    AGCACCAGGCAGTAAATCCAG|AATAAGAGCCTATGCATACATATTGACAGCTCACAATCCTGTACTGGT
MOBP99                         |AATAAGAGCCTATGCATACATATTGACAGCTCACAATCCTGTACTGGT

          4970       4980       4990       5000       5010       5020       5030
            |          |          |          |          |          |          |
Contig# 1  TATTGCTTACATAAGGAGAAGCCACAGATCTCATTCTTTCTTCTAGAATGCAATAGAATTTAAAAATGA
 rOP1      TATTGCTTACATAAGGAGAAGCCACAGATCT ATTCTTTCTTCTAGAATGCAATAGAATTTAAAAATGA
MOBP81A    TATTGCTTACATAAGGAGAAGCCACAGATCTCATTCTTTCTTCTAGAATGCAATAGAATTTAAAAATGA
MOBP99     TATTGCTTACATAAGGAGAAGCCACAGATCTCATTCTTTCTTCTAGAATGCAATAGAATTTAAAAATGA

           5040       5050       5060       5070       5080       5090       5100
            |          |          |          |          |          |          |
Contig# 1  CAGCAGAACAAAAGGACACTAGAAATTATCCTCTCAGAAATAAAAGTATAAGAAATAATTTCTGGCTTG
 rOP1      CAGCAGAACAAAAGGACACTAGAAATTATCCTCTCAGAAATAAAAGTATAAGAAATAATTTCTGGCTTG
MOBP81A    CAGCAGAACAAAAGGACACTAGAAATTATCCTCTCAGAAATAAAAGTATAAGAAATAATTTCTGGCTTG
MOBP99     CAGCAGAACAAAAGGACACTAGAAATTATCCTCTCAGAAATAAAAGTATAAGAAATAATTTCTGGCTTG

           5110       5120       5130       5140       5150       5160       5170
            |          |          |          |          |          |          |
Contig# 1  AGGAAATAACCCAGGCTAAAAGCTACACAGGCCTGGCTTAACCATACCAGCAGGGCATACCTTTGTTAG
 rOP1      AGGAAATAACCCAGGCTAAAAGCTACACAGGCCTGGCTTAACCATACCAGCAGGGCATACCTTTGTTAG
MOBP81A    AGGAAATAACCCAGGCTAAAAGCTACACAGGCCTGGCTTAACCATACCAGCAGGGCATACCTTTGTTAG
MOBP99     AGGAAATAACCCAGGCTAAAAGCTACACAGGCCTGGCTTAACCATACCAGCAGGGCATACCTTTGTTAG

           5180       5190       5200       5210       5220       5230       5240
            |          |          |          |          |          |          |
Contig# 1  GAGAGGGTCAATTCGAAGTTTGTGAACAGCCTGGGCTGTTTCCGAGGCTGCATAAATCTTAAACACCAA
 rOP1      GAGAGGGTCAATTCGAAGTTTGTGAACAGCCTGGGCTGTTTCCGAGGCTGCATAAATCTTAAACACCAA
MOBP81A    GAGAGGGTCAATTCGAAGTTTGTGAACAGCCTGGGCTGTTTCCGAGGCTGCATAAATCTTAAACACCAA
MOBP99     GAGAGGGTCAATTCGAAGTTTGTGAACAGCCTGGGCTGTTTCCGAGGCTGCATAAATCTTAAACACCAA

           5250       5260       5270       5280       5290       5300       5310
            |          |          |          |          |          |          |
Contig# 1  AAATGTGTGTTTTGAATTAACTTTGGGTCTGGTCTACAGAGTCCGGCTTGTGATGGCTCCTGACACGGG
 rOP1      AAATGTGTGTTTTGAATTAACTTTGGGTCTGGTCTACAGAGTCCGGCTTGTGATGGCTCCTGACACGGG
MOBP81A    AAATGTGTGTTTTGAATTAACTTTGGGTCTGGTCTACAGAGTCCGGCTTGTGATGGCTCCTGACACGGG
MOBP99     AAATGTGTGTTTTGAATTAACTTTGGGTCTGGTCTACAGAGTCCGGCTTGTGATGGCTCCTGACACGGG

           5320       5330       5340       5350       5360       5370       5380
            |          |          |          |          |          |          |
Contig# 1  AGAAAAGGGGATAAGTTAATATAAGGCTCTGGGATTTTCCTTTGAGGAACAGGCAGGTAGGAAAGGACA
 rOP1      AGAAAAGGGGATAAGTTAATATAAGGCTCTGGGATTTTCCTTTGAGGAACAGGCAGGTAGGAAAGGACA
MOBP81A    AGAAAAGGGGATAAGTTAATATAAGGCTCTGGGATTTTCCTTTGAGGAACAGGCAGGTAGGAAAGGACA
MOBP99     AGAAAAGGGGATAAGTTAATATAAGGCTCTGGGATTTTCCTTTGAGGAACAGGCAGGTAGGAAAGGACA

           5390       5400       5410       5420       5430       5440       5450
            |          |          |          |          |          |          |
Contig# 1  CTAGAATGTTGGGGAATGTGGAGGGATGTAAAACAGACAGTATTATGATGGGGGGAACTCAGGAAGGAA
 rOP1      CTAGAATGTTGGGGAATGTGGAGGGATGTAAAACAGACAGTATTATGATGGGGGGAACTCAGGAAGGAA
MOBP81A    CTAGAATGTTGGGGAATGTGGAGGGATGTAAAACAGACAGTATTATGATGGGGGGAACTCAGGAAGGAA
MOBP99     CTAGAATGTTGGGGAATGTGGAGGGATGTAAAACAGACAGTATTATGATGGGGGGAACTCAGGAAGGAA
```


5460 5470 5480 5490 5500 5510 5520
Contig# 1 CGGTAAGGACACCAGAATGGCCTCGTGCTGGGAGCCAGGGTGGCCAAGCCCAGGCAGATGCACAAAGAC
rOP1 CGGTAAGGACACCAGAATGGCCTCGTGCTGGGAGCCAGGGTGGCCAAGCCCAGGCAGATGCACAAAGAC
MOBP81A CGGTAAGGACACCAGAATGGCCTCGTGCTGGGAGCCAGGGTGGCCAAGCCCAGGCAGATGCACAAAGAC
MOBP99 CGGTAAGGACACCAGAATGGCCTCGTGCTGGGAGCCAGGGTGGCCAAGCCCAGGCAGATGCACAAAGAC
5530 5540 5550 5560 5570 5580
Contig# 1 CCCAAAGGCCACCATTGCCTCACAACCAGACCCAGCAAAGTTCAGATCAAGTGCATAAATGGGGGTCAC
rOP1 CCCAAAGGCCACCATTGCCTCACAACCAGACCCAGCAAAGTTCAGATCAAGTGCATAAATGGGGGTCAC
MOBP81A CCCAAAGGCCACCATTGCCTCACAACCAGACCCAGCAAAGTTCAGATCAAGTGCATAAATGGGGGTCAC
MOBP99 CCCAAAGGCCACCATTGCCTCACAACCAGACCCAGCAAAGTTCAGATCAAGTGCATAAATGGGGGTCAC
5590 5600 5610 5620 5630 5640 5650
Contig# 1 GTCAGTGGAAACTGTCCGGAGCATGGAGCAGGGGGTGAGGAAGACAGTCAGAAGGAAGCAAGAGATGGC
rOP1 GTCAGTGGAAACTGTCCGGAGCATGGAGCAGGGGGTGAGGAAGACAGTCAGAAGGAAGCAAGAGATGGC
MOBP81A GTCAGTGGAAACTGTCCGGAGCATGGAGCAGGGGGTGAGGAAGACAGTCAGAAGGAAGCAAGAGATGGC
MOBP99 GTCAGTGGAAACTGTCCGGAGCATGGAGCAGGGGGTGAGGAAGACAGTCAGAAGGAAGCAAGAGATGGC
5660 5670 5680 5690 5700 5710 5720
Contig# 1 TTAGAGTGTTGCATGGCAGCAAGAAGAAGCAGTCCAAATATCATCCCAAGGCAGCCATGGCTTCCTCTC
rOP1 TTAGAGTGTTGCATGGCAGCAAGAAGAAGCAGTCCAAATATCATCCCAAGGCAGCCATGGCTTCCTCTC
MOBP81A TTAGAGTGTTGCATGGCAGCAAGAAGAAGCAGTCCAAATATCATCCCAAGGCAGCCATGGCTTCCTCTC
MOBP99 TTAGAGTGTTGCATGGCAGCAAGAAGAAGCAGTCCAAATATCATCCCAAGGCAGCCATGGCTTCCTCTC
5730 5740 5750 5760 5770 5780 5790
Contig# 1 CCGAGGACACTGTGGCCCTGTGGAAGTGCCAAGGCAGAGGCAGAGGAATTTAACCGCACACAAACTACA
rOP1 CCGAGGACACTGTGGCCCTGTGGAAGTGCCAAGGCAGAGGCAGAGGAATTTAACCGCACACAAACTACA
MOBP81A CCGAGGACACTGTGGCCCTGTGGAAGTGCCAAGGCAGAGGCAGAGGAATTTAACCGCACACAAACTACA
MOBP99 CCGAGGACACTGTGGCCCTGTGGAAGTGCCAAGGCAGAGGCAGAGGAATTTAACCGCACACAAACTACA
5800 5810 5820 5830 5840 5850 5860
Contig# 1 TGCAGAAACTAAGGTGCTAAATGAACATCTACGTGACAACCTTTCTGAGGACCAGGTTCCAGTGAGCGG
rOP1 TGCAGAAACTAAGGTGCTAAATGAACATCTACGTGACAACCTTTCTGAGGACCAGGTTCCAGTGAGCGG
MOBP81A TGCAGAAACTAAGGTGCTAAATGAACATCTACGTGACAACCTTTCTGAGGACCAGGTTCCAGTGAGCGG
MOBP99 TGCAGAAACTAAGGTGCTAAATGAACATCTACGTGACAACCTTTCTGAGGACCAGGTTCCAGTGAGCGG
5870 5880 5890 5900 5910 5920 5930
Contig# 1 CCCAGGGCCGTTATCTTCGCGATGGTAACTGAGCTGAGAAGAATGGGAACCGGATAGGAGCACACAGTA
rOP1 CCCAGGGCCGTTATCTTCGCGATGGTAACTGAGCTGAGAAGAATGGGAACCGGATAGGAGCACACAGTA
MOBP81A CCCAGGGCCGTTATCTTCGCGATGGTAACTGAGCTGAGAAGAATGGGAACCGGATAGGAGCACACAGTA
MOBP99 CCCAGGGCCGTTATCTTCGCGATGGTAACTGAGCTGAGAAGAATGGGAACCGGATAGGAGCACACAGTA
5940 5950 5960 5970 5980 5990 6000
Contig# 1 GCCCACTGGTGAGACAAATCCACGGGCGAGCCTCAACCAACCCACTAAGGAAACTTCACGCCTTTTCAT
rOP1 GCCCACTGGTGAGACAAATCCACGGGCGAGCCTCAACCAACCCACTAAGGAAACTTCACGCCTTTTCAT
MOBP81A GCCCACTGGTGAGACAAATCCACGGGCGAGCCTCAACCAACCCACTAAGGAAACTTCACGCCTTTTCAT
MOBP99 GCCCACTGGTGAGACAAATCCACGGGCGAGCCTCAACCAACCCACTAAGGAAACTTCACGCCTTTTCAT
6010 6020 6030 6040 6050 6060 6070
Contig# 1 CTAGTTTTCATTTTGGCAAAGCAAAGCCATCCTGAGTGCTTGCTTGTCTCCGCCCTCCACGCCACCCCC
rOP1 CTAGTTTTCATTTTGGCAAAGCAAAGCCATCCTGAGTGCTTGCTTGTCTCCGCCCTCCACGCCACCCCC
MOBP81A CTAGTTTTCATTTTGGCAAAGCAAAGCCATCCTGAGTGCTTGCTTGTCTCCGCCCTCCACGCCACCCCC
MOBP99 CTAGTTTTCATTTTGGCAAAGCAAAGCCATCCTGAGTGCTTGCTTGTCTCCGCCCTCCACGCCACCCCC
6080 6090 6100 6110 6120 6130 6140
Contig# 1 GAGACACAGAGCATGCGCATTAACCCAGAGCGCGCTAACACATGCACAGCGTTGGACATTAGCCTATTA
rOP1 GAGACACAGAGCATGCGCATTAACCCAGAGCGCGCTAACACATGCACAGCGTTGGACATTAGCCTATTA
MOBP81A GAGACACAGAGCATGCGCATTAACCCAGAGCGCGCTAACACATGCACAGCGTTGGACATTAGCCTATTA
MOBP99 GAGACACAGAGCATGCGCATTAACCCAGAGCGCGCTAACACATGCACAGCGTTGGACATTAGCCTATTA

```
                  6150       6160       6170       6180       6190       6200
                    |          |          |          |          |          |
Contig# 1  ATGGGAATAACCAGAAACTACTAAATTTTGTTTGTTGTAGAAATAAAACCTGGGGATGGATCAAAA
 rOP1      ATGGGAATAACCAGAAACTACTAAATTTTGTTTGTTGTAGAAATAAAACCTGGGGATGGATC
MOBP81A    ATGGGAATAACCAGAAACTACTAAATTTTGTTTGTTGTAGAAATAAAACCTGGGGATGGATCAAAA
MOBP99     ATGGGAATAACCAGAAACTACTAAATT
```

Figure 2.3.1 Sequence multi-alignment of published (Yamamoto *et al.*, 1994; Holz *et al.*, 1996) rat splice variants. Putative exon/exon junctions are indicated by | and ***. Polyadenylation signals are identified in the sequence as ATTAAA and AATAAA. Exonic sequences containing polyadenylation signals are separated from subsequent exonic sequence units by **. N.B. Nucleotide numbers, listed above sequence, are purely arbitrary and bear no relevance to the genomic organisation of the gene. Contig #1 is an arbitrary contiguous sequence, generated by joining the exonic segments Q-Z.

Transcript identification	EMBL accession number	Publication
rOP1 (Mobp 71)	D28110	Yamamoto *et al.*, 1994
rOPRP1 (Mobp 170)	D28111	Yamamoto *et al.*, 1994
Mobp 69	X90402	Holz *et al.*, 1996
Mobp 81A	X87900	Holz *et al.*, 1996
Mobp 81B	X89637	Holz *et al.*, 1996
Mobp 99	X89638	Holz *et al.*, 1996

Table 2.3.1 *Mobp* transcript sequences derived from rat tissue defined, according to their respective EMBL accession number.

Exonic sequence	Q	R	S	T	U	V	W	X	Y	Z
Nucleotide numbers	1-117	118-326	337-753	781-1038	1049-1177	1191-1413	1426-4329	4039-4277	4338-4920	4920-6182

Table 2.3.2 Exonic sequences defined according to arbitrary nucleotide numbers listed above sequence multialignment (Figure 2.3.1).

Transcript	Q	R	S	T	U	V	W	X	Y	Z
Mobp 71	+	+	-	-	+	+	-	-	+	+
Mobp 170	+	+	+	-	-	-	-	-	-	-
Mobp 69	+	+	-	+	-	+	-	+	-	-
Mobp 81A	+	+	-	-	-	+	-	-	+	+
Mobp 81B	+	+	-	-	-	+	+	+	-	-
Mobp 99	+	+	-	-	-	-	-	-	-	+

Table 2.3.3 Identification of sequence units shared between the published transcript sequences, based on sequence multialignment (Figure 2.3.1). Q-Z, exonic sequences; +/-, presence or absence of exonic sequence.

Exonic sequence	5' Donor		Polyadenylation signal	Intron	3' Acceptor	Exonic sequence
Q	CAC	AG	/	Size unknown	T GAG	R
R	ACC Thr^{68}	AG SeS ArT ArU ArZ	/	Size unknown	C CGC r^{69} Arg^{70}	S
S	CCTGC		...ATTAAA...	Size unknown	A^{69} TAA g OCH	T
T	CTTAG		/	Size unknown	G^{69} ACT^{70} g Thr	U
U	GAAAG		/	Size unknown	A^{69} TTG^{70} g Leu	V
V	TGGAG		/	Size unknown	GTGAG	W
W	CCTAG		/	Size unknown	GGCCT	X
X	ACGTC		...AATAAA...	Size unknown	CACTT	Y
Y	TCCAG				A^{69} ATA^{70} g Ile	Z
Z	GTAGA		...AATAAA...			

Table 2.3.4 Sequences across putative exon/exon boundaries of messenger RNA species from the rat *Mobp* gene (Yamamoto *et al.*, 1994 and Holz *et al.*, 1996). The DNA sequences across putative exon/exon boundaries are shown and the encoded amino acid sequences are indicated in three-letter code. The last amino acid residue encoded by Ex 2 may be Ser or Arg, dependent upon which downstream exonic segment they are subsequently spliced to. The downstream exonic segment is indicated, in each case, by the subscript letters S, T, U or Z. The amino acid sequence is numbered from the first residue of the isoform MOBP 81 (Holz *et al.*, 1996). Polyadenylation signals are indicated by the sequence ATTAAA (exon 3) and AATAAA (exons 8 and 10). Exonic segments containing the exonic sequence portion (C/AAG) of the extended 5' splice donor sequence [(C/A)AG|GT(G/A)AGT] are underlined.

2.3.2 Discussion

Sequence from the six transcripts of the novel rat gene *Mobp*, identified in the studies of Yamamoto *et al.* (1994) and Holz *et al.* (1996), have been compared. This study confirms the conclusion that the transcripts share in common their 5' regions (Yamamoto *et al.*, 1994 and Holz *et al.*, 1996); encoding amino acids 1-68 of their predicted proteins. The report of Holz *et al.* (1996) further states that each of the predicted proteins has a unique C-terminal segment. However, the authors almost completely fail to recognise that the exonic segments, encoding the carboxyl termini and providing the respective 3'UTRs of the *Mobp* transcripts, do share sequences to varying degrees. Concession is made in the report of Holz *et al.* (1996) that sequences are shared by rOP1 and MOBP 81A. In fact 6 of the 10 exonic segments identified in this study are shared by two or more of the transcripts reported by Yamamoto *et al* (1994) and Holz (1996). These findings lend strength to the suggestion that MOBP isoforms are generated by a complex series of alternative splicing events (Holz *et al.*, 1996 and Montague *et al.*, 1997). Exclusive use of certain exonic segments does occur. Exonic segments S, T, U and W are used exclusively in the transcripts rOPRP1, MOBP 69, rOP1 and MOBP 81B.

The predictions of exon/exon boundaries are based on sequence comparison and identification of putative 5' donor splice site remnants. These predictions are tentative and serve only as a model for examination in an endeavour to elucidate the genomic organisation of the *Mobp* gene. Thus, the precise relationships between the alternatively spliced mRNA species and the single copy gene remain unknown, at this stage. However, examination of the genomic region encompassing *Mobp* will resolve this (Montague *et al.*, 1997; McCallion *et al.*, manuscript in preparation; refer also to Section 2, 2.4).

2.4 Isolation and characterization of the murine gene (*Mobp*) encoding myelin-associated oligodendrocytic basic protein

2.4.1 Isolation of the *Mobp* gene

Two mouse 129/Sv genomic libraries, prepared in λ 2001 (Warren *et al.*, 1994) and λ PS (Nehls *et al.*, 1994), were used to isolate the *Mobp* gene. These libraries were kindly donated by Dr Andrew Smith (University of Edinburgh). Screening of the libraries was performed, at high stringency, as described in Chapter 2 using as probe a 1.6 kb cDNA fragment (defined in Chapter 2). Using this approach forty four independent clones were obtained; thirty seven from the λ 2001 library and seven from the λ PS library.

2.4.2 Genomic structure of the murine gene (*Mobp*) encoding myelin-associated oligodendrocytic basic protein

2.4.2.1 DNA fingerprinting of bacteriophage clones containing *Mobp* sequences

The procedure used for fingerprinting the clones is outlined in Figure 2.1 (Chapter 2) and was performed (as described in Chapter 2, 2.2.4) utilising all 37 clones isolated from the λ 2001 library. Cloned DNA was incubated in the presence of two restriction enzymes with 6 bp specificities (*Eco*RI and *Hin*dIII) that leave staggered ends. These ends were simultaneously radioactively labeled by end-filling with the inclusion of a radioactive dNTP. The enzymes were then heat inactivated and the DNA fragments cleaved again, this time with an enzyme with a 4 bp specificity (*Sau*3AI). The resulting fragments were separated in denaturing polyacrylamide (sequencing type) gels. The signals obtained on autoradiographs were arbitrarily identified as I-XXXV. All data were compared with those derived from AS18 (arbitrary selection). The number of signals shared between AS18 and any other clone was taken as an indication of the degree of overlap of the cloned fragments. This overlap was expressed in arbitrary units. The data arising from this assay are shown in Table 2.4.1. The extent of overlaps, inferred by this data, amongst selected clones is illustrated in Figure 2.4.2. It is noteworthy that *Sau* 3AI was the enzyme utilised in the construction of both libraries.

Clone identification	Number of interpretable signals	Signal identification	Number of signals shared with ASM 18
ASM 1	8	I, II, IV, VII, X, XIV, XV, XXI	3
ASM 2	9	I, II, IV, XIII, XVI, XXI, XXV, XXXIII	5
ASM 8	12	I, V, VII, VIII, IX, XXI, XXII, XXVI, XXVIII, XXIX, XXXI, XXXII	10
ASM 11	5	I, VIII, XXII, XXVIII, XXXII	2
ASM 12	16	I, V, VII, VIII, IX, X, XI, XV, XXI, XXII, XXVI, XXVII, XXVIII, XXIX, XXXI, XXXII	13
ASM 14	4	I, IV XII, XXVI	3
ASM 15	15	I, II, III, IV, V VII, VIII, IX, XXI, XXII, XXVI, XXVIII, XXIX, XXXI, XXXII	13
ASM 16	7	I, V VII, XXIX, XXXI, XXXIII, XXXV	6
ASM 17	9	I, XII, XIII, (XXII), XXIII, XXIV, XXV, XXVI, XXXV	4
ASM 18	22	I, II, III, IV, V VII, VIII, IX, X, XI, XVI, XVII, XVIII, XXI, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXIII, XXXV	22
ASM 19	20	I, II, III, IV, V, VI, VII, VIII, X, XI, XVI, XVIII, XXI, XXVI, XXVIII, XXIX, XXX, XXXI, XXXII, XXXIII	19
ASM 20	17	I, IV, V, VII, VIII, IX, X, XI, XVI, XXI, XXII, XXVIII, XXIX, XXX, XXXI, XXXII, XXXV	15
ASM 21	21	I, II, III, IV, V, VI, VII, VIII, X, XI, XVI, XVIII, XXI, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXIII, XXXIV	18

Table 2.4.1

Clone identification	Number of interpretable signals	Signal identification	Number of signals shared with ASM 18
ASM 22	10	I, XII, XIII, (XXII), XXIII, XXIV, XXV, XXVI, XXXII, XXXV	5
ASM 23	8	I, VI, IX, XXII, XXVIII, XXXII, XXXV, XXXVI	3
ASM 24	7	I, II, XII, XVII, XXII, XXVI, XXXII	3
ASM 25	7	I, II, XII, XVII, XXII, XXVI, XXXII	3
ASM 26	12	I, V, VII, VIII, IX, XXI, XXII, XXVI, XXVIII, XXIX, XXXI, XXXII	9
ASM 27	12	I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XVI, XVIII, XXI, XXII, XXIII, XXIV, XXV, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXII, XXXIII, XXXIV, XXXV	2
ASM 33	14	I, V, VII, VIII, XXI, XXII, XXIII, XXIV, XXVI, XXVIII, XXIX, XXXI, XXXII, XXXV	9
ASM 35	13	I, V, VII, VIII, XXI, XXII, XXV, XXVI, XXVIII, XXIX, XXXI, XXXII, XXXV	10
ASM 39	21	I, IV, V, VII, VIII, IX, X, XI, XVI, XVII, XXI, XXII, XXV, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXII, XXXIII	18
ASM 40	13	I, IV, V, VII, VIII, IX, XXI, XXII, XXV, XXVI, XXVIII, XXXI, XXXII	8
ASM 41	18	I, IV, V, VII, VIII, IX, X, XI, XVI, XVII, XXI, XXII, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXII	16

Table 2.4.1 continued

Clone identification	Number of interpretable signals	Signal identification	Number of signals shared with ASM 18
ASM 42	18	I, IV, V, VII, VIII, IX, X, XI, XVI, XVII, XXI, XXII, XXVI, XXVII, XXVIII, XXIX, XXX, XXXI, XXXII	16
ASM 45	11	I, V, VII, VIII, IX, XXI, XXII, XXVIII, XXIX, XXXI, XXXII	9
ASM 46	11	I, IX, XXII, XXV, XXVIII, XXXII	3

Table 2.4.1 Data generated by the DNA fingerprinting assay. The data displayed above indicate, in arbitrary units, the degree of overlap between examined clones and the selected reference clone ASM 18. Clones ASM 3, 28, 29, 32, 34, 48 and 49 failed to provide interpretable results. Signal (XXII) was considered close to but not equal to the signal identified as XXII.

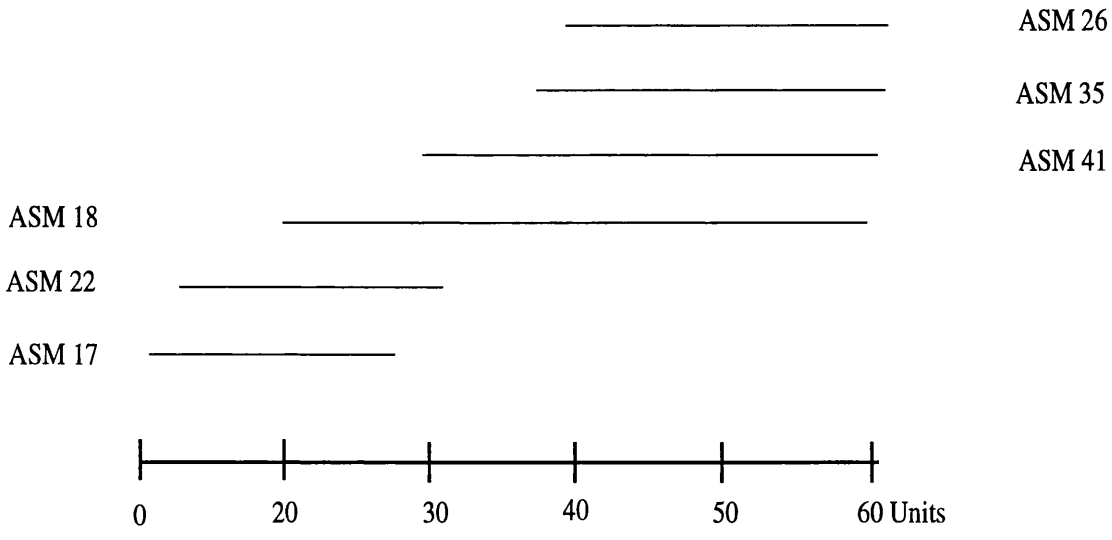


Figure 2.4.2 Schematic representation of overlaps between selected cloned DNAs based on data generated by the fingerprinting assay, section 2.4.2.1 (Table 2.4.1)

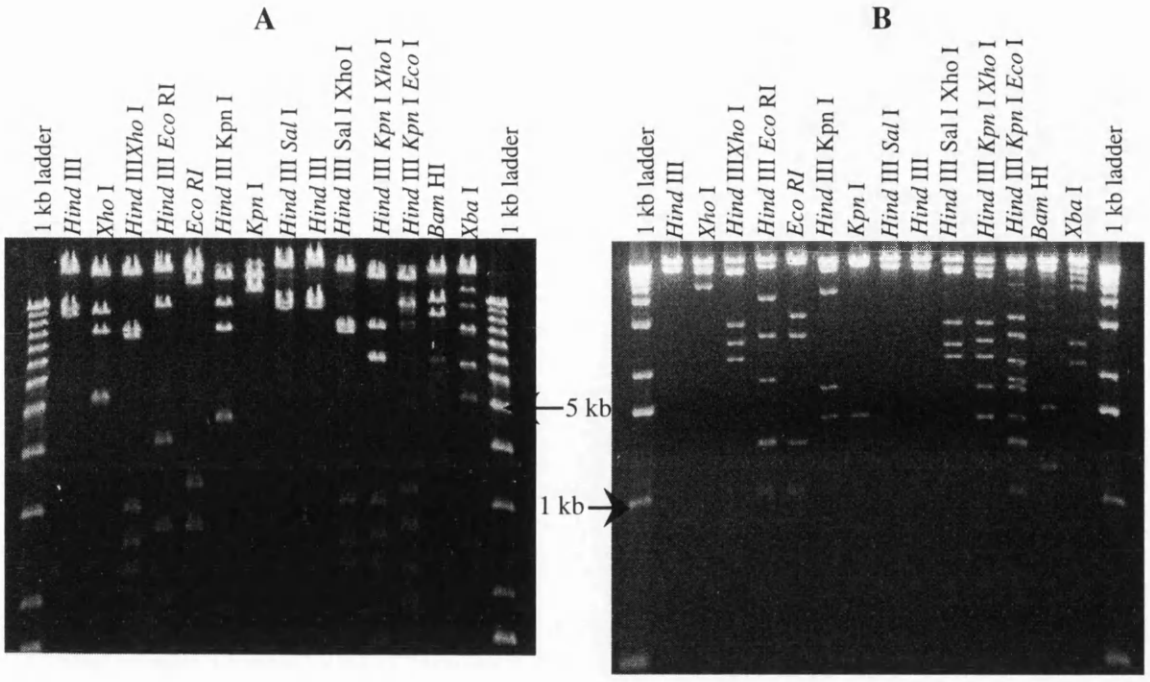
2.4.2.2 Restriction mapping of bacteriophage clones containing *Mobp* sequences

Restriction mapping of selected clones ([λ 2001] ASM 17, 18, 22, 26, 29, 35 and 41; [λ PS] ASM 50, 55 and 59) was performed using a battery of restriction enzymes including: *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn* I, *Nde* I, *Sal* I, *Sst* I, *Pst* I and *Xho* I. The resulting cleaved fragments were separated by electrophoresis in agarose gels (0.5 - 1.5%) as described in Chapter 2. Figure 2.4.3 (A and B) illustrates an example of electrophoresis of DNA fragments resulting from the restriction enzyme cleavage of cloned DNAs, and list the corresponding fragment sizes. Figure 2.4.3 (C) illustrates the corresponding restriction maps. Figure 2.4.4 illustrates the restriction maps resulting from this series of assays and demonstrates the degree of overlap amongst selected, cloned DNAs. Overlaps are elucidated on the basis of the identification of restriction fragments that are shared by two or more cloned DNAs. The restriction map is also in close agreement with the map (Figure 2.4.2) generated from the DNA fingerprinting assay.

2.4.3 Identification of exonic sequences in cloned genomic fragments

Subsequent to electrophoresis, DNAs were transferred to nylon membrane as described in Chapter 2. To locate exonic sequences identified in the studies of Yamamoto *et al.*, 1994; Holz *et al.*, 1996 and Montague *et al.*, 1997, a series of oligonucleotide and PCR generated probes, designed to published *Mobp* sequences, were hybridised to these membrane bound DNAs as described in Chapter 2. The probes utilised in this study are defined in Table 2.4.2. Figures 2.4.5 and 2.4.6 A/B show examples of hybridisation experiments identifying sequence (defined by PCR oligonucleotides ASM 201 F/R; Chapter 2, Table 2.3) corresponding to an exonic segment encoding a portion of the 68 amino acids shared by all MOBP isoforms (ASM501) and to 3'UTR (defined by PCR oligonucleotides ASM 215 F/R; Table 2.4, Chapter 2) in splice variants Mobp 81A and rOP1 (ASM502). Figure 2.4.4b illustrates the position of the probes utilised, on the restriction map (Figure 2.4.4a) generated in section 2.4.2.

Sequences defined by oligonucleotides ASM 242 and ASM 244 (Chapter 2, Table 2.4), were not detected amongst the overlapping lambda cloned DNAs (Figure 2.4.4a). Oligonucleotides ASM 242 and ASM 244 were designed to sequence within exon 1 and exon 2 respectively. These exons provide the majority of the length (4 nt are provided by exon 3) of the 5'UTRs identified in rat and mouse *Mobp* transcripts (rat: Yamamoto *et al.*, 1994; Holz *et al.*, 1996 and mouse: McCallion, PhD thesis, 1997 [Section 2.5]).



Hin dIII	11.5 kb, 11 kb, > 15 kb
Xho I	22 kb, 9 kb, 10.8 kb, >5 kb
Hin dIII *Xho* I	8.5 kb, 2.9 kb, 2.6 kb, 2.2 kb
Hin dIII *Sal* I	2 × > 15 kb
Hin dIII *Eco* RI	20 kb, 9 kb, 3.9 kb, 2.7 kb, 1.9 kb, 1.3 kb, 1 kb
Eco RI	22 kb, > 11 kb, 3.3 kb, 2.7 kb, 1.3 kb, 1 kb
Hin dIII *Kpn* I	> 22 kb,12 kb, 9 kb, 4.2 kb, 1.8 kb
Kpn I	> 13 kb, 10.5 kb, 9 kb, 1.5 kb, 1 kb
Hin dIII *Sal* I *Xho* I	> 22 kb, 9 kb, 8.5 kb, 2.9 kb, 2.4 kb, 2.2 kb
Hin dIII *Kpn* I *Xho* I	> 15 kb, 9 kb, 6.6 kb, 2.9 kb, 2.6 kb, 2.2 kb, 1.8 kb
Hin dIII *Kpn* I *Eco* RI	> 15 kb, 9 kb, 3.1 kb, 2.7 kb, 2.2 kb, 1.9 kb, 1.8 kb, 1.4 kb, 1.4 kb, 1 kb

C

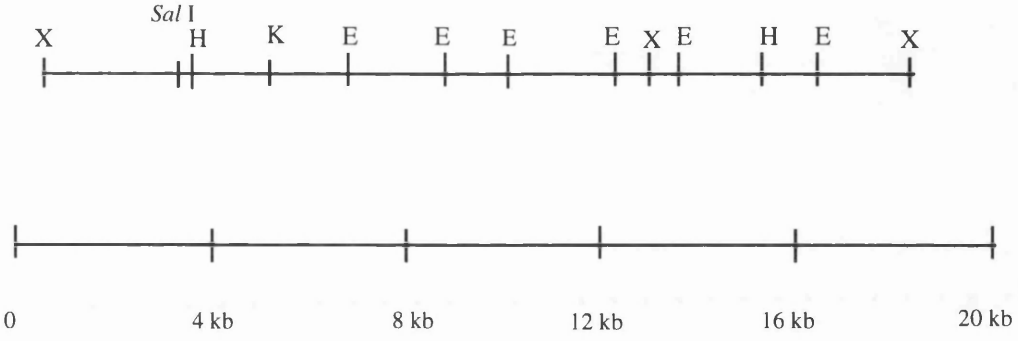


Figure 2.4.3 Electrophoresis of cleaved DNA fragments from clone ASM 41, subsequent to incubation with the restriction enzymes and enzyme combinations indicated above. Figure 2.4.3 (A and B [DNA fragments electrophoresed in 0.5 % and 1.5 % agarose gels respectively]) illustrates an example of electrophoresis of DNA fragments resulting from the restriction enzyme cleavage of cloned DNAs, and list the corresponding fragment sizes. Figure 2.4.3 (C) illustrates the corresponding restriction map. Restriction maps resulting from this and other restriction mapping experiments are illustrated in Figure 2.4.4. E, H, K and X represent restriction enzymes *Eco* RI, *Hin* dIII, *Kpn* I and *Xho* I respectively.

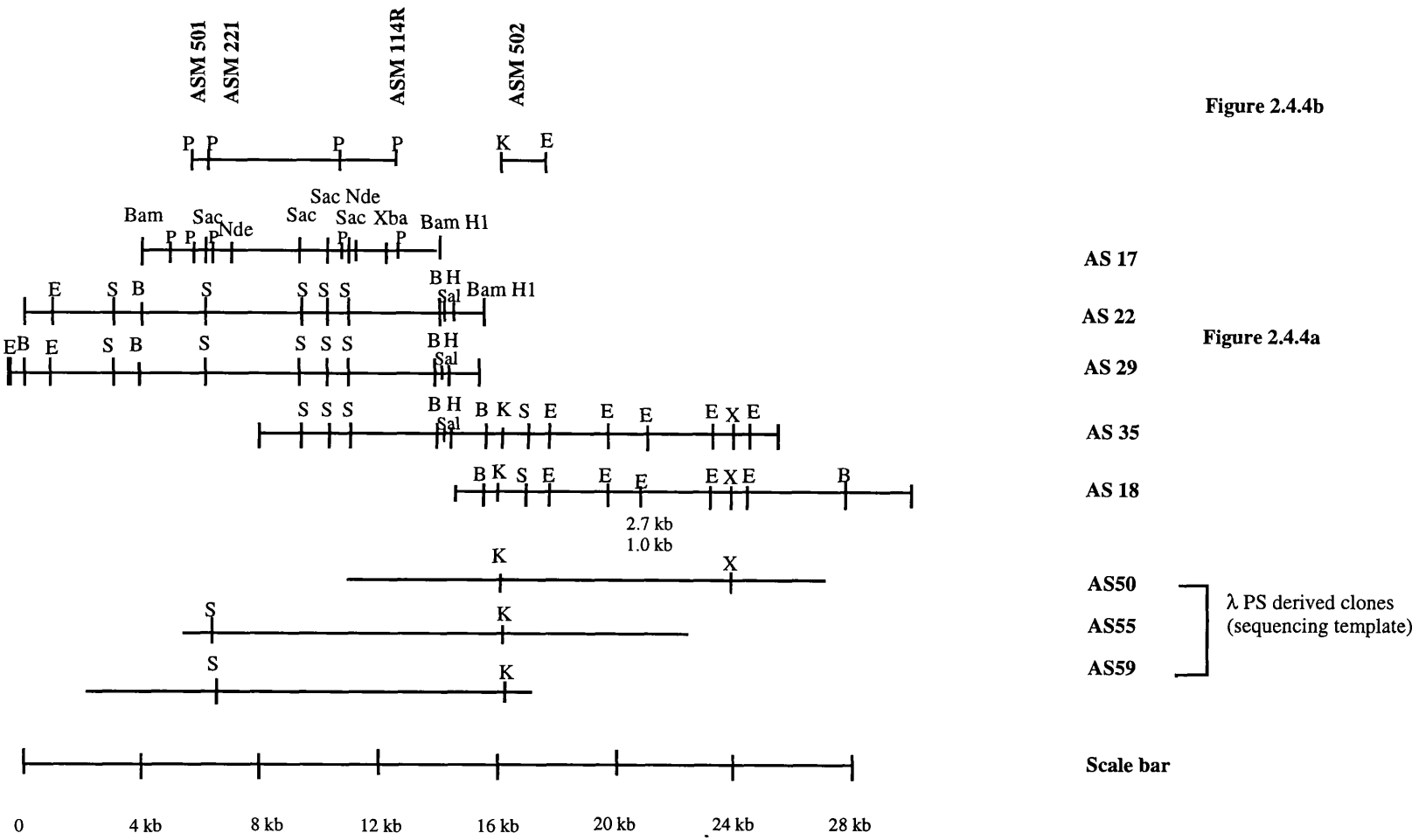


Figure 2.4.4 a and b (a) Restriction map constructed from data generated in 2.4.1.2. (b) Schematic representation of the position of probes (defined in table 2.4.2) on the restriction map (figure 2.4.4a). Signal from probes ASM 242 and ASM 244 were not detected amongst these cloned DNAs, in these experiments (2.4.3). E, B, H, K, S and X represent restriction enzymes *Eco* RI, *Bam* HI, Hin dIII, *Kpn* I, *Sst* I and *Xho* I respectively. Sal, *Sal* I. Probes ASM 501, ASM 221, ASM 114R and ASM 502 correspond to exons 3, 4, 7c and 8b (Table 2.4.2).

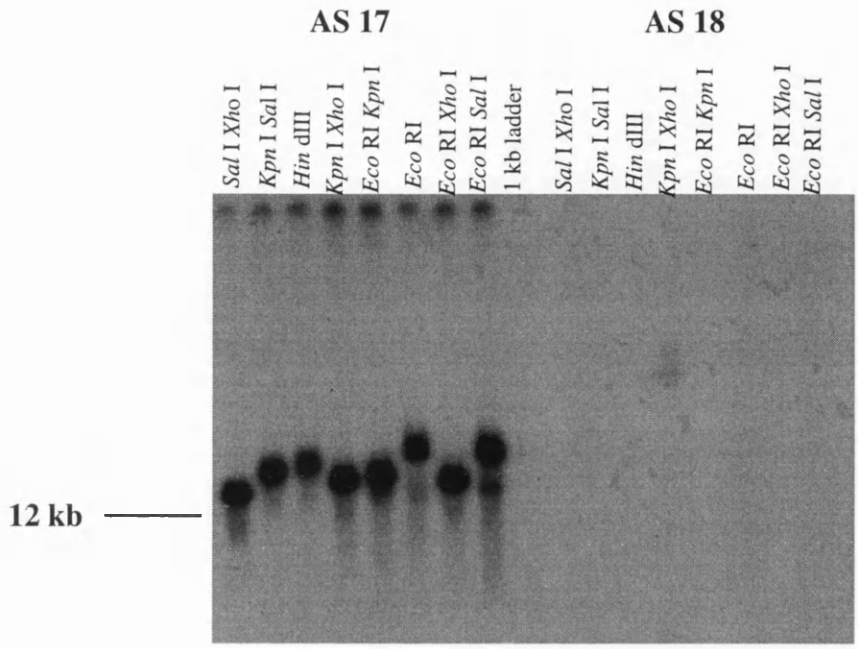


Figure 2.4.5 Autoradiograph resulting from hybridisation of the radioactively labelled probe ASM 501 (defined in Table 2.4.2) to membrane bound cleaved DNA fragments from clones ASM 17 and ASM 18 (refer also to Figure 2.4.4 a). This demonstrates the presence of *Mobp* exon 3 within ASM 17 (refer also to Figure 2.4.4 b).

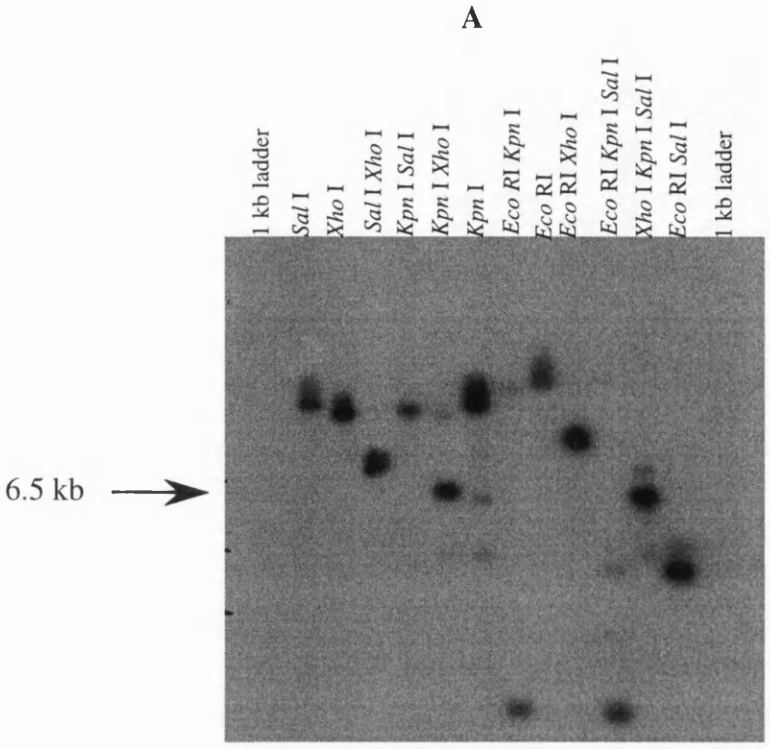


Figure 2.4.6 Autoradiograph resulting from hybridisation of the radioactively labelled probe ASM 502 (defined in Table 2.4.2) to membrane bound cleaved DNA fragments from clone ASM 35 (refer also to Figure 2.4.6 B). This demonstrates the presence of *Mobp* exon 8 within ASM 35 (refer also to Figure 2.4.4 A and B). Figure 2.4.6 A corresponds to DNA fragments that were electrophoresed in 0.5% ordinary agarose gels.

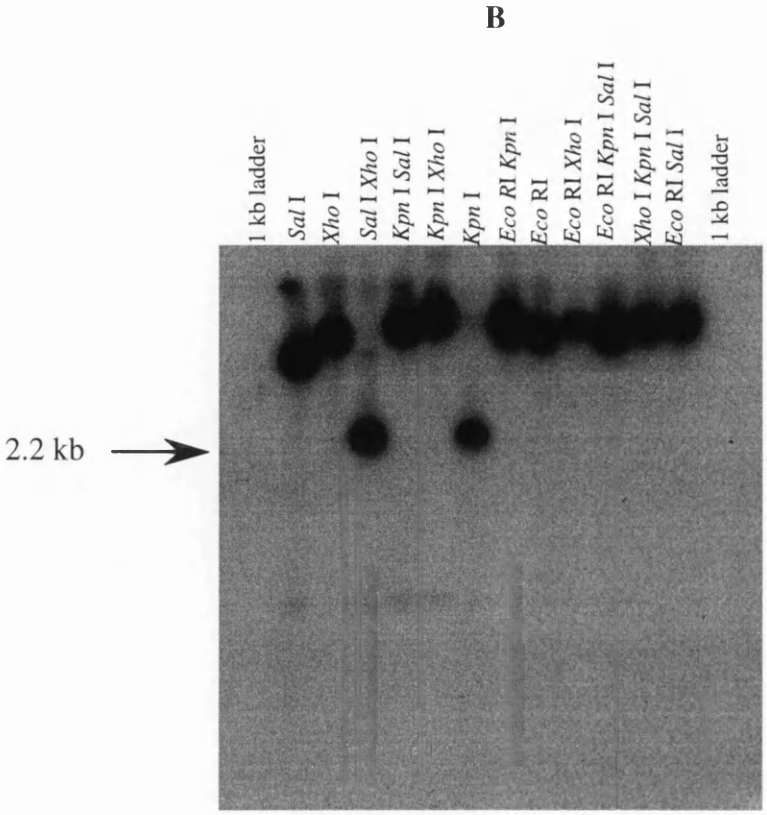


Figure 2.4.6 Autoradiograph resulting from hybridisation of the radioactively labelled probe ASM 502 (defined in Table 2.4.2) to membrane bound cleaved DNA fragments from clone ASM 35 (refer also to Figure 2.4.6 B). This demonstrates the presence of *Mobp* exon 8 within ASM 35 (refer also to Figure 2.4.4 A and B). Figure 2.4.6 B corresponds to DNA fragments that were electrophoresed in 1.5% ordinary agarose gels.

Probe identification	Origin of sequence	Publication	Oligonucleotide/s defining probe	Gene region
ASM 501	Rat	Yamamoto *et al.*, 1994	ASM 201 F/R	Exon 3
ASM 502	Mouse	Savioz and Davies, 1994	ASM 215 F/R	Exon 8b
ASM 242	Mouse	McCallion *et al.*, manuscript submitted	ASM 242	Exon 1
ASM 244	Mouse	McCallion *et al.*, manuscript submitted	ASM 244	Exon 2
ASM 221	Mouse	Montague *et al.*, 1997	ASM 221	Exon 4
ASM 114 R	Rat	Holz *et al.*, 1996	ASM 114 R	Exon 7c

Table 2.4.2 Probes utilised Southern blot-hybridisation experiments 2.4.3. Oligonucleotides utilised as probes or in the PCR generation of probes are defined in Table 2.4, Chapter 2. The *Mobp* gene regions to which these probes are designed are identified above. The identification of exons is based on data obtained in the completed studies reported in Section 2 (2.4) and Section 2 (2.5).

2.4.5 Sequence analysis of *Mobp*

2.4.5.1 Sequence analysis of cloned DNA fragments containing *Mobp* sequences

Dye-terminator *Taq* cycle sequencing was performed, as described in Chapter 2, utilising double-stranded plasmid DNA as template. Sequence was obtained from cloned DNAs; as λ PS derived plasmids (AS 50, AS 55, AS 59), generated by automatic subcloning (Chapter 2) or from DNA fragments derived from the *Sst* I restriction enzyme cleavage of clones AS17 and AS26, subcloned in a pBS KS^+ plasmid backbone. The first oligonucleotides to be synthesised (ASM 101, 102, 103, ASM 201 F/R, ASM 114 R; defined in Table 2.3, Chapter 2) for sequencing were designed to sequences within reported rat transcripts (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). Subsequently oligonucleotides were designed to sequence generated by this study. All oligonucleotides utilised in this study are defined in Table 2.3 (Chapter 2). Figure 2.4.7 illustrates how the sequence obtained corresponds to the restriction map of the region encompassing *Mobp* and to the genomic organisation of the gene. The combined length of contiguous sequence units (contigs) 1, 2 and 3 comprises >10 kb. Overlapping sequences were generated using, in most instances, one strand of DNA from any single region. However, all sequencing reactions have been performed at least twice with each oligonucleotide. Only unambiguously identified bases were used in the generation of contiguous sequence units. Contigs 1, 2 and 3 are listed in Appendices VII, VIII and IX respectively. The sequence defined as contig 1 (> 3.5 kb) encompasses exons 3-5 (defined in Table 2.4.3). The sequence defined as contig 2 (> 5 kb) encompasses exons 6 and 7 (defined in Table 2.4.3). The sequence defined as contig 3 (> 2.8 kb) encompasses exon 8 (defined in Table 2.4.3). Sequence across intron/exon junctions are reported in section 2.4.5 and in Table 2.4.3.

N.B. All sequencing reaction for study of the *Mobp* gene were performed with the technical support of Dr G.J. Stewart.

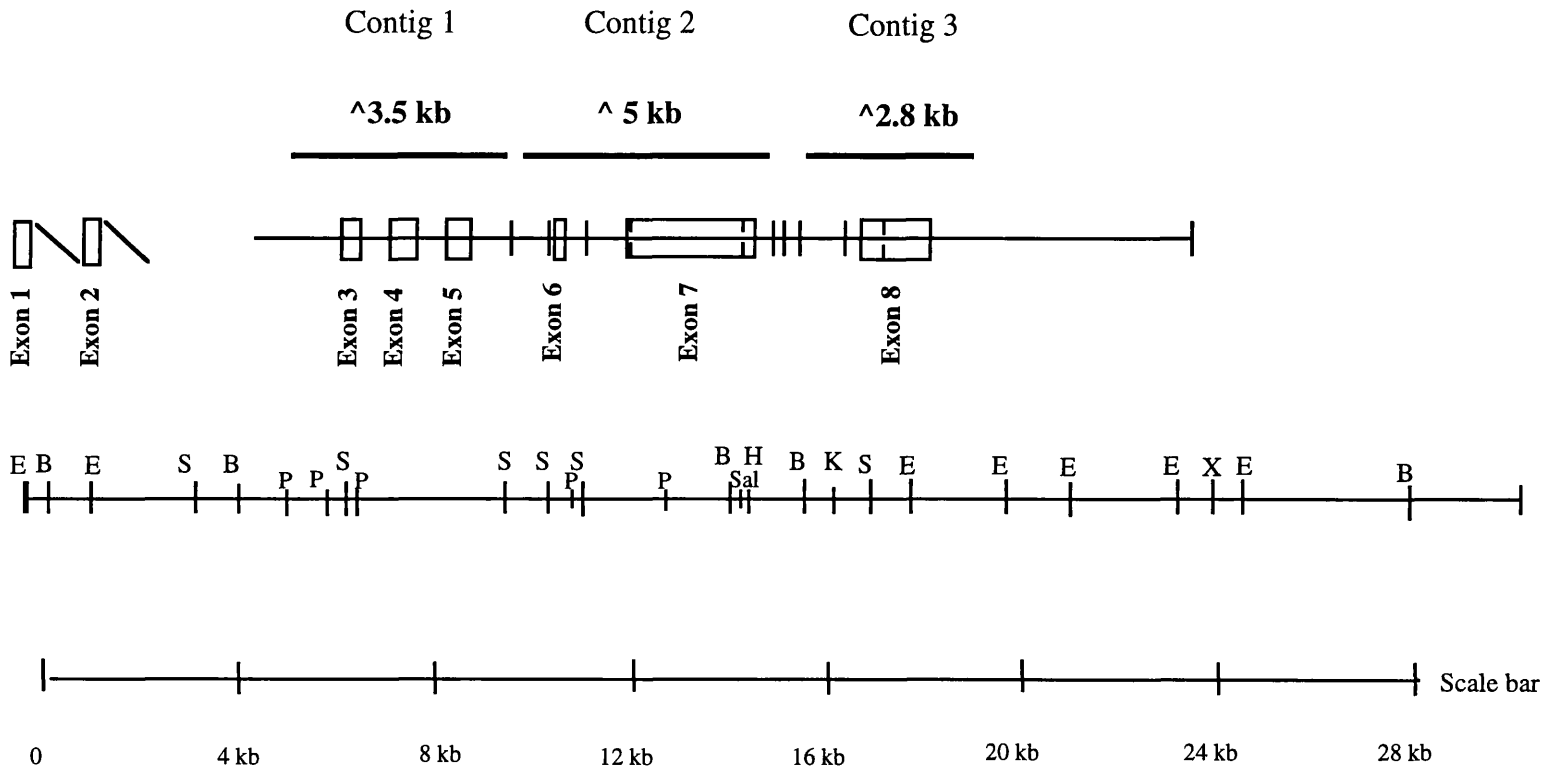


Figure 2.4.7 a

Figure 2.4.7 b

Figure 2.4.7 c

Figure 2.4.7 Correspondence between the restriction map of the region encompassing *Mobp*, the genomic orginisation of *Mobp* and contiguous sequence units (contigs) 1, 2 and 3. (a) Composite of contiguous sequence units (1, 2 and 3) generated across the region encompassing *Mobp*. (b) Orginisation of the gene encoding MOBP. (c) Composite restriction map of the cloned genomic fragments ASM 17, 18, 22, 26, 28, 35 and 41. E, B, H, K, S and X represent restriction enzymes *Eco* RI, *Bam* HI, Hin dIII, *Kpn* I, *Sst* I and *Xho* I respectively. N.B. Exons 1 and 2 are separated from each other and from exon three by an, as yet, undetermined distance.

2.4.5.2 Generation of a sequence multi-alignment of coding exon and 3'UTR sequences from the mouse and rat genes encoding myelin-associated oligodendrocytic basic protein

Sequences were imported to the Genejockey software (Cambridge, Biosoft), from the database according to their EMBL accession numbers (defined in Table 2.3.1 and Section 2, 2.5.1) and aligned by eye within the Genejockey sequence multialignment feature. The results of this analysis are illustrated in Appendix X.

Sequence from exons 1 and 2 of the murine gene *Mobp* demonstrate ≤ 84% identity with the 5' UTR sequences within the published rat splice variants (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). Exon 3 demonstrates 99% identity, at the nucleotide level, with its corresponding sequence in rat. These sequences demonstrate 100% identity between the amino acids that they encode. Exon 4 demonstrates 98% identity, at the nucleotide level, with its corresponding sequence in rat. This corresponds to a 96% similarity at the amino acid level. The codon encoding amino acid 69 (Arg) of MOBP 69, encoded by exon 5, is conserved between rat and mouse. Exon 6, which in rat encodes amino acids 69-71 (Arg Thr Val) of MOBP 71, in mouse encodes 5 amino acids (Arg Thr Val Arg Lys) resulting in a predicted protein of 73 amino acids. Exon 7a encodes amino acids 69-81 of MOBP 81 in mouse, and demonstrates 100% identity, at the amino acid level, with amino acids 69-81 of the isoform MOBP 81. The 3'UTRs provided by exons 7b/c and 8b demonstrate >85% sequence identity with their corresponding sequences in rat.

2.4.5.3 Identification of repeated sequence structures within the *Mobp* gene

The region encompassing the *Mobp* gene contains a number of repeat structures. These are listed in Table 2.4.4.

2.4.6 Linear organisation of *Mobp* exons

Analysis of isolated clones by restriction mapping, Southern blot hybridisation and DNA sequencing indicated that this gene comprises 8 exons encompassing a region in excess of 15 kb. Exons 7 and 8 also utilise internal splice donor/acceptor sequences and are consequently further defined as Exon 7 (7a, 7b and 7c) and exon 8 (8a and 8b). The sequences across the intron/exon boundaries are shown in Table 2.4.3.
The evidence provided by the presence of an extended exonic 5' donor sequence (C/AAGl, where the exon/intron junction is denoted by a vertical line), and the PCR amplification of a 1.2 kb genomic fragment (refer to study performed in Section 2, 2.5) using primers ASM

242 and ASM 244 (defined; Table 2.3, Chapter 2) supports the hypothesis that the identified 5'UTRs (rat: Yamamoto *et al.*, 1994; Holz *et al.*, 1996 and mouse: McCallion, PhD thesis, 1997 [Section 2.5]) are interrupted by an intron. Subsequently, the exonic segments to which these oligonucleotides were designed have been arbitrarily identified as discrete exons. Sequence analysis of cloned cDNA fragments, generated in an endeavour to identify the transcriptional start sites associated with *Mobp* (Oligo-capping; Maruyama and Sugano, 1994; Section 2.5), indicated alternative use of the second of these sequence units (exon 2) but constitutive use of the first (exon 1). No such alternative splicing in the 5'UTRs of rat RNAs has been reported (Yamamoto *et al.*, 1994; Holz *et al.*, 1996). Exons downstream of the arbitrarily named exons 1 and 2 have been named exons 3-8 in their corresponding linear order in the mouse genome.

Exon 3 was found to encode the amino acid residues 1 - 68 present in all previously reported MOBP isoforms (Yamamoto *et al.*, 1994; Holz *et al.*, 1996 and Montague *et al.*, 1997). The complex relationship between exons 4 - 8, the six splice variants of the *Mobp* transcript, reported in rat and mouse and the proteins that they encode (Yamamoto *et al.*, 1994; Holz *et al.*, 1996 and Montague *et al.*, 1997), is illustrated in Figure 2.4.8 and in Figure 2.4.9 (refer also to Section 2, 2.3). It is noteworthy that the transcripts Mobp 81A and Mobp 81B, though identical in their coding regions, are distinguished by their use of exons 7b/c and 8a/b respectively, in the generation of their different 3'UTRs. It has been reported that these transcripts demonstrate differing subcellular localisation and are expressed at different developmental timepoints (Montague *et al.*, 1997; Montague *et al.*, unpublished data).

Exon 4 encodes the 102 amino acids unique to the isoform MOBP 170 ([rOPRP1] Yamamoto *et al.*, 1994; Montague *et al.*, 1997). Exon 5 was found to encode amino acid 69 of the isoform MOBP 69. Exon 6 was found to encode amino acids 69-73 of predicted protein MOBP 73. The corresponding sequence unit in rat encodes amino acids 69-71 of isoform rOP1 [MOBP 71] (Yamamoto *et al.*, 1994). Exon 7a was found to encode amino acids 69-81 of isoform MOBP81 and was completely conserved between the rat and mouse *Mobp* genes. However, the same sequence also provides a portion of the 3'UTRs corresponding to transcripts rOP1 and Mobp 69. Exon 7b provides a 2.6 kb portion of 3'UTR unique to the transcript Mobp 81B. Exon 7c provides a portion of 3'UTR found within the transcripts Mobp 69 and Mobp 81B. Exon 8a/b provides a portion of 3'UTR found within transcript Mobp 81A and Mobp 73. However exon 8b, whose corresponding sequence in rat encodes amino acids 69-99 of isoform MOBP 99, does not contain an ORF reading in frame with that encoding amino acids 1-68. This suggests that no isoform,

corresponding to the polypeptide MOBP 99 in rat, is produced from the mouse *Mobp* gene. However, the fact that exon 8b demonstrates > 85% sequence identity with its corresponding sequence in rat, supports the suggestion that the 3'UTRs of *Mobp* splice variants may be important e.g. in the translocation of RNAs to various parts of the oligodendrocyte (Holz *et al.*, 1996; Montague *et al.*, 1997).

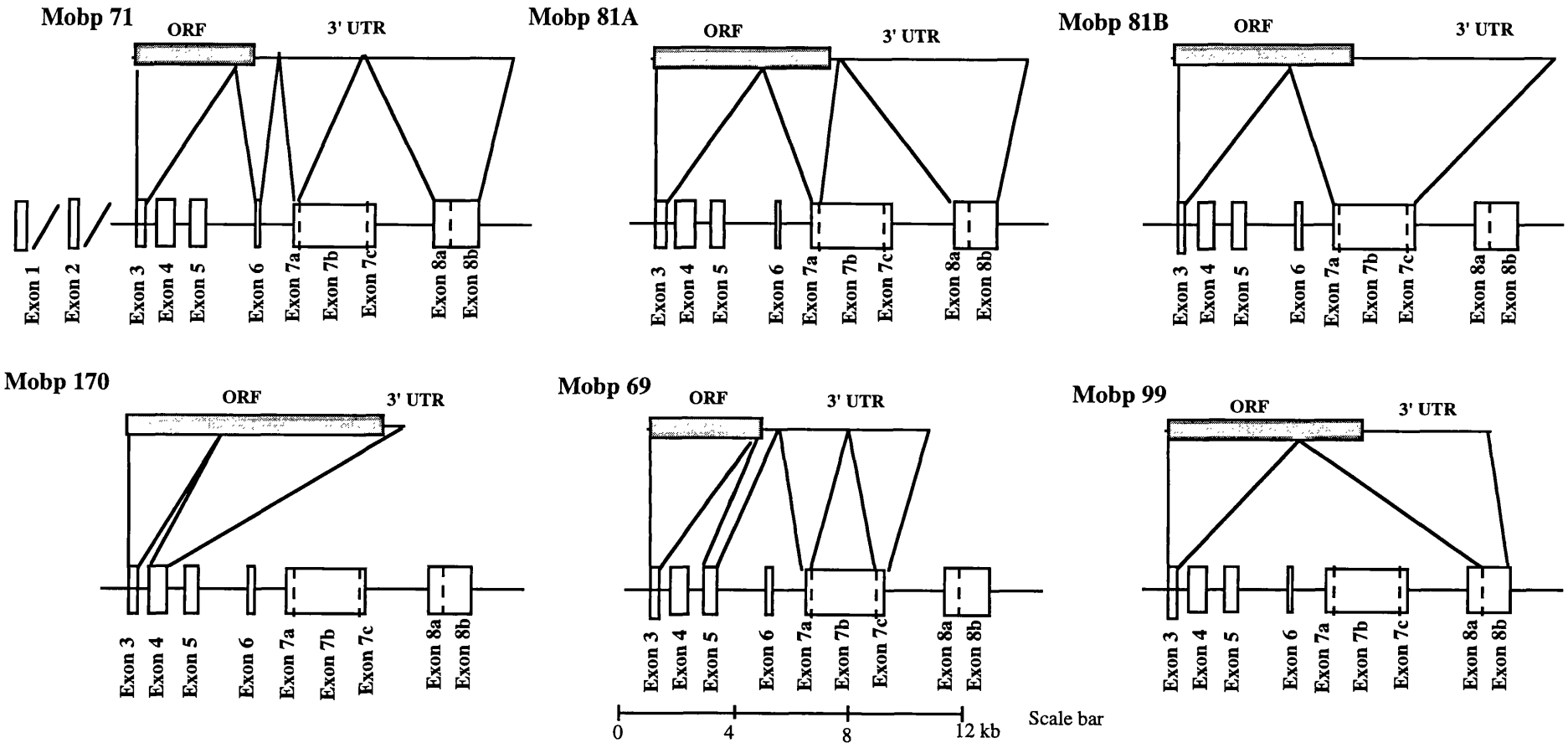


Figure 2.4.8 Schematic representation of postulated splicing events, in the murine *Mobp* gene, generating transcripts equivalent to those reported in rat by Yamamoto *et al.*, 1994; Holz *et al.*, 1996. Exons 1 and 2 are separated from each other and from exon three by an, as yet, undetermined distance. Protein coding regions (Open reading frames; ORFs) are indicated by shaded areas. N.B. Scale bar corresponds to genomic region only; transcripts are not drawn to scale. Mobp 71 and Mobp 170 correspond to transcripts rOP1 and rOPRP1, as defined by Yamamoto *et al.*, 1994

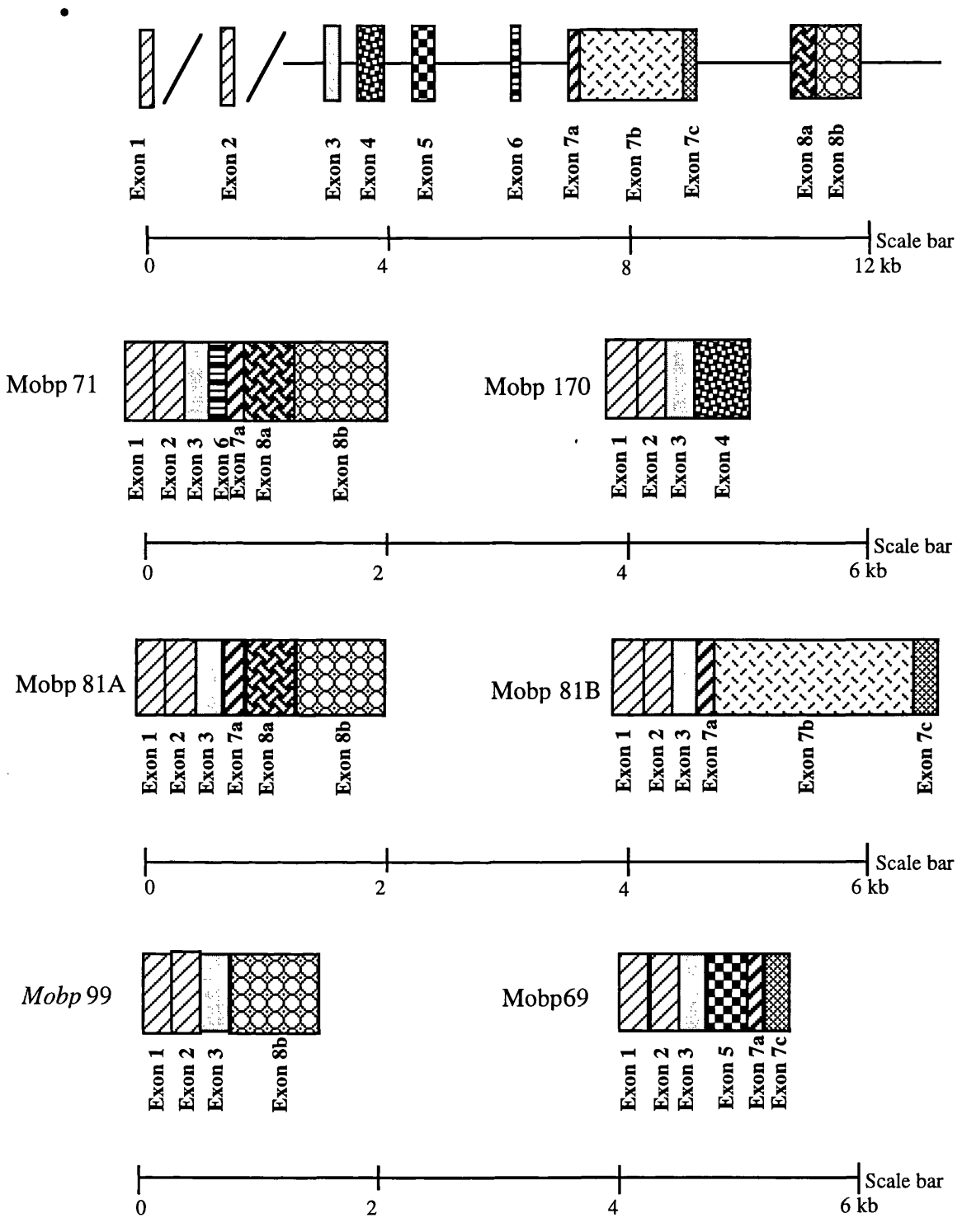


Figure 2.4.9 Schematic representation of exon use in the generation of splice variants, of the murine *Mobp* gene transcript (Figure 2.4.10). •, Schematic represenation of the genomic region encompassing *Mobp*. The indicated constitutive use of exons 1 and 2, by predicted splice variants, is purely arbitrary. The use of exons 1 and 2, within individual splice variants has not yet been elucidated. Mobp 71 and Mobp 170 correspond to transcripts rOP1 and rOPRP1, as defined by Yamamoto *et al.* (1994). Scale bars are situated immediately below their corresponding schematic diagram(s).

2.4.7 Identification of RNA signals in *Mobp* gene

The minimum consensus sequences (Mount, 1982; Smith *et al.*, 1989) for splice junctions, GT and AG (the intron 5' donor and 3' acceptor sites respectively), were found at the appropriate ends of all six introns where intronic sequence is available (Table 2.4.3). An exact match to the extended 5' donor site (C/A)AG|GT(G/A)AGT (where the exon/intron boundary is denoted by the vertical line) is not found at any of the 5' exon/intron junctions in the *Mobp* gene. The sequence at the 3' end of exon 6 is the most divergent from the consensus sequence; only six out of nine bases are in agreement. It is noteworthy that this sequence is not highly conserved (Smith *et al.*, 1989). However, the extended 3' acceptor site (C/T)AG|G is found in two of the 3' intron/exon junctions (exon 6 and exon 7c). The other 6 intron/exon boundaries, where intronic sequence is available, differ from the consensus sequence by only a single nucleotide. Alternative 3' acceptor sites are identified in exon 7 (7a and 7c) and in exon 8 (8a and 8b). Exon 7 also contains an internal 5' donor site at the exon/exon junction between exon 7a and 7b. Polyadenylation signals are indicated by the sequence ATTAAA (exon 4) and AATAAA (exons 7c and 8b).

2.4.7 Discussion

I report here the organisation of the murine gene *Mobp* encoding myelin-associated oligodendrocytic basic protein. Comparison of these data with those for reported transcripts generated from the rat and mouse Mobp genes (rat: Yamamoto *et al.*, 1994; Holz *et al.*, 1996 and mouse: Montague *et al.*, 1997) provides insight into the complex series of splicing events required in their generation (Holz *et al.*, 1996; Montague *et al.*, 1997). The evident use of alternative 3' splice acceptor sites (exons 7a, 7c, 8a and 8b) and 5' splice donor sites (exons 7a, 7c) adds to the, now obvious, complexity of the *Mobp* transcriptional unit. Whether intron/exon boundaries are conserved between these species will require further study of the rat gene (*Mobp*).

A knowledge of the organisation of this gene permits the conceptualisation of gene disruption strategies that will permit the elimination of expression of all MOBP isoforms or of individual isoforms. I have already initiated work to establish a mouse model in which the *Mobp* gene is disrupted, removing expression of all MOBP isoforms. This work is reported in Section 2.6.

In summary; the murine *Mobp* gene encoding myelin-associated oligodendrocytic basic protein is a complex transcriptional unit, comprising 8 exons and encompassing a genomic region in excess of 15 kb. The relationship between its genomic orginisation and its

transcriptional complexity may now be deduced, at least in part. Elucidation of the genomic orginisation of *Mobp* provides a platform for functional study through gene disruption strategies (refer also to Section 2, 2.6).

Exon			5' Donor	Intron size	3' Acceptor			Exon
Exon 1	CGC	AG	/	/	/	G	GCA	Exon 2
Exon 2	CCC	AG	/	/	.atggtttccag.	T	GAG	Exon 3
Exon 3	ACT	AG	.gtgagtgcggg.	0.2 kb	.ctctgcaccag.	C	CGC	Exon 4
	Thr^{68}	Se_4				r^{69}	Arg	
		Ar_5						
		Ar_6						
		Se_{8b}						
Exon 4	CCT	GC	...ATTAAA...	0.3 kb	.ctgctttatag.	A^{69}	TAA	Exon 5
						g	OCH	
Exon 5	CTC	GG	.gttagaccagt.	2.5 kb	.ttcctcc**cag**.	**G**^{69}	ACT	Exon 6
						g	Thr	
Exon 6	GGA	AG	.gtggacggctt.	3.0 kb	.ctataaatcag.	A^{69}	TTG	Exon 7a
						g	Leu	
Exon 7a	TGG	AG	.gtgagaagggg.	1.2 kb	...		...	Exon 7b
Exon 7b	...		...		.gcactcc**tag**.	**G**	GCCT	Exon 7c
Exon 7c	ACG	TC	...AATAAA...	3.0 kb	.tctcccctcag.		CATGT	Exon 8a
Exon 8a	...		...		.agtaaatgcag.	T^{69}	ATG	Exon 8b
						r	Met	
Exon 8b	GTA	GA	...AATAAA...					

Table 2.4.3 Sequences across the intron / exon boundaries of the murine *Mobp* gene. The DNA sequences across the intron / exon boundaries are shown and the encoded amino acid sequences are indicated in three-letter code. The last amino acid residue encoded by exon three may be Ser or Arg dependent on which downstream exon (4, 5, 6 or 8b) they are subsequently spliced to. The exon in question is indicated, in each case, by the subscript numbers 4, 5, 6 or 8b. The amino acid sequence is numbered (superscript) from the first residue of the isoform MOBP 81 (Holz *et al.*, 1996). Nucleotides in upper-case letters represent exonic sequences, while nucleotides in lower-case letters represent intronic sequences. Alternative 3' acceptor sites are identified in exon 7 (7a and 7c) and in exon 8 (8a and 8b). Exon 7 contains an internal 5' donor site at the exon/exon junction between exon 7a and 7b. The most 3' end of this exon (exon 7c) terminates with the polyadenylation signal AATAAA. Polyadenylation signals are indicated by the sequence ATTAAA (exon 4) and AATAAA (exons 7c and 8b). Extended 5' splice donor sites and 3' acceptor sites are underlined. The extended 3' acceptor site (C/T)AG|G, found in the 3' intron/exon junctions of exon 6 and exon 7c respectively, are in bold type.

Repeat unit	Number of repeat units	Description	Location within *Mobp*	Contig Number	Location within Contig
CA	18	Perfect	Intron 2	1	
GA	26	Perfect	Intron 2	1	
TTTCAAA	4	Imperfect	Intron 5	2	
GT/C	12	Imperfect	Exon 7b	2	
GT/C	19	Imperfect	Exon 7b	2	
CA	5	Perfect	Exon 7b	2	
CG/A	27	Imperfect	Exon 7b	2	
CA	20	Perfect	Exon 7b	2	
CCACA	5	Imperfect	Exon 7b	2	

Table 2.4.4 Repeat structures within the *Mobp* gene

Locations within contigs correspond to nucleotide (nt) positions within contigs 1, 2 and 3 (Appendices VII, VIII and IX respectively).

2.5 Promoter region of the murine *Mobp* gene

2.5.1 Identification of transcriptional start sites

The oligo-capping procedure, first reported by Maruyama and Sugano (1994) was utilised in an endeavour to identify the transcriptional start sites of the murine gene *Mobp*. This procedure is made possible by the fact that the 5' end of eukaryotic mRNA has a cap structure (Furuichi and Miura, 1975) comprising a 7-methylated GTP attached to the first nucleotide of the message via two pyrophosphate linkages (Furuichi and Miura, 1975). The protocol is described in full in Chapter 2. In brief, poly A^+ RNA was dephosphorylated in the presence of bacterial alkaline phosphatase (BAP) and the methylated cap structure removed by incubation with tobacco acid pyrophosphatase (TAP). An oligoribonucleotide (r-oligo) was then ligated to the 5' end of previously capped mRNAs. It should be noted that: (a) BAP removes phosphates from 5' uncapped RNAs, leaving a hydroxyl at the 5' end; (b) TAP removes the cap structure from capped intact RNAs leaving a 5' phosphate whilst leaving internal phosphodiester bonds unchanged; (c) RNA ligase requires a 3'-hydroxyl (Acceptor) and a 5'-phosphate (Donor) for ligation of substrates. Consequently, where mRNAs are treated first with BAP and then with TAP, the RNA ligase can only ligate an r-oligo to previously capped RNAs. The 5' end of *Mobp* RNAs were then specifically reverse transcribed and PCR amplified (RT-PCR), and cloned into the cloning vector pCR™ 2.1 (defined in Table 2.2, Chapter 2) and *Taq* cycle sequencing, of the resulting clones, performed (as described in Chapter 2).

2.5.2 RT-PCR amplification of the 5' ends of *Mobp* RNAs

RT was performed using the oligonucleotide ASM 202 (defined in Table 2.3, Chapter 2). Subsequent PCR amplification was performed using the oligonucleotides TAG 135 (defined in Table 2.3, Chapter 2) and ASM 202. A nested PCR amplification was then performed, utilising as template product from the primary amplification (1/10 reaction volume), using the oligonucleotides TAG 135 and ASM 201 (defined in Table 2.3, Chapter 2). The sequence to which ASM 201 was designed is upstream to the sequence to which ASM 202 was designed. The amplification products were separated by electrophoresis (as described in Chapter 2) in 2% ordinary agarose gels (Figure 2.5.1). Four signals were observed, corresponding to DNA fragments in a size range 50-300 bp. These fragments were band isolated (as described in Chapter 2) and each was amplified using the oligonucleotides TAG 135 and ASM 201. Subsequently, 1 µl of each reaction was ligated with the cloning vector pCR™ 2.1 and transformed into the bacterial strain

INVα F' (defined in Table 2.2, Chapter 2). *Taq* cycle sequencing was performed using as template, cDNA clones corresponding to these amplified products. The sequences derived from cDNA fragments generated using this RT-PCR combination are illustrated in Figure 2.5.2 and are all prefixed by OC. This RT-PCR combination was repeated using the oligonucleotide ASM 203 (defined in Table 2.3, Chapter 2) in the RT step. The amplification stages utilised the oligonucleotide TAG 135 in conjunction with the oligonucleotides ASM 203 and ASM 204 (defined in Table 2.3, Chapter 2), in the primary and secondary PCR steps respectively. The sequence to which oligonucleotide ASM 203 was designed was downstream of the sequence to which ASM 204 was designed. This RT-PCR combination also identified four major signals when products were separated by electrophoresis (as described in Chapter 2) in 2% ordinary agarose gels. These four signals, corresponded to DNA fragments in a size range 50-350 bp. These fragments were band isolated (as described in Chapter 2) and each was amplified using the oligonucleotides TAG 135 and ASM 204. PCR products were isolated, cloned and sequenced as indicated above (described in Chapter 2). The sequences derived from cDNA fragments generated using this RT-PCR combination are illustrated in Figure 2.5.2 and are all prefixed by OCII. It should be noted that oligonucleotides ASM 201, ASM 202, ASM 203 and ASM 204 were all designed to sequences within the primary coding exon (exon 3) of *Mobp*. It should be noted that all *Mobp* exons have been named subsequent to the conclusion of all studies contributing to this thesis.

2.5.3 Sequence identification of transcriptional start sites

In total, 70 clones provided interpretable sequence. The corresponding sequences and distribution of start sites are illustrated in Figure 2.5.2. Five major transcriptional start sites (tsp), or clusters of start sites, are readily observable (Figure 2.5.2) amongst sequences derived from cDNA fragments in experiments OC and OCII. The lengths of these sequences (28-280 nt) correlate well with the observed fragment sizes (Figure 2.5.1), when the lengths of the corresponding oligonucleotide sequences are included. It should be noted that amplified cDNA fragments with a product size range of 50-60 bp would be indistinguishable from PCR artefacts resulting from primer-dimer formation, in the above mentioned agarose gel electrophoresis. A major tsp (>17% of cDNAs sequenced) is present at 140 nt upstream of the translation start site ($\underline{ATG}_{1049\text{-}1051}$, where T_{1050} is the point relative to which all tsp positions are defined and the subscript number 1050 corresponds to the nucleotide position in Contig 1) (Figure 2.5.3) used in the production of all previously identified MOBP isoforms (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997). A cluster of tsp are present at 75-68 nt (>7% of cDNAs sequenced) upstream of T_{1050} (Figure 2.5.3). It is noteworthy that a third tsp, comprising

> 5% (OC 4.41 inv, OC 3.1, OC 3.19 and OC 13) of sequences in this study, contains sequence from exons 1 and 3 in the absence of exon 2 (Figure 2.5.2). These tsp are found at a point 60 nt (>5% of cDNAs sequenced) upstream of T_{1050} (Figure 2.5.2).

Two further tsp are found clustered at points 25 nt (>35% of cDNAs sequenced) and 101 nt (>10% of cDNAs sequenced) downstream of T_{1050} (Figure 2.5.2). As only previously capped RNAs are available for ligation of an r-oligo (Maruyama and Sugano, 1994; Kurihara *et al.*, 1997), these data would indicate the presence of a transcriptional unit whose transcription is initiated within the *Mobp* gene but utilises a translation start point that is downstream of the ATG used in the generation of reported isoforms (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997).

2.5.3.1 Identification of a putative novel transcriptional unit within *Mobp*

To assess the possibility of the presence of a downstream transcriptional unit existing within *Mobp*, sequence analysis of intron 3 (separating exons 3 and 4, Figure 2.4.7, Contig 1) was performed. This analysis identified a putative translation start site (ATG) within the intronic sequence. This methionine codon was followed by an open reading frame (ORF) encoding a putative protein of some 155 amino acids. Furthermore, this ORF extends into exon 4 and reads in frame with the ORF encoding the 102 C-terminal amino acids of the isoform MOBP 170 (rOPRP 1, Yamamoto *et al.*, 1994). The sequence across intron 3 and exon 4 is shown in Figure 2.5.3. The amino acid sequence of the protein encoded by the novel transcript would predict that like MOBP 170, it too is a small (26 kD) basic hydrophilic positively charged (pI = 12.25 and net charge at pH 7.5 = +26) proline rich (35/155 aa residues) protein (Figure 2.5.3).

Oligonucleotides (ASM 245 and ASM 246; defined in Table 2.3, Chapter 2) were designed to sequence within intron 3 (within the novel ORF) and exon 4 respectively. These oligonucleotides were used in an RT-PCR assay for transcript corresponding to sequence within this novel ORF. A preliminary study indicated that a corresponding sequence was detectable in cDNA from mouse brain (Postnatal day 43 [P43]). Time limitation precluded pursuit of this line of study by this author. However, the study of this putative novel transcript has continued in collaboration with the laboratory of Dr I. Griffiths (University of Glasgow). Results to date indicate that this transcript is detectable at levels equivalent to or slightly lower than Mobp 81A (Montague *et al.*, unpublished data). This novel transcript is first detected, by PCR, in concert with or

just after that of Mobp 73 and Mobp 81A; first detected in the brain at P5 (Montague *et al.*, 1997). Detection of this putative novel transcript lags behind that of Mobp 170, first detected at embryonic day 12 (E12). Mobp 170 is the only other transcript reported to utilise sequence corresponding to exon 4 (Yamamoto *et al.*, 1994; Montague *et al.*, 1997). Expression of this putative novel transcript would also appear to be oligodendrocyte specific (Montague *et al.*, unpublished data). It is noteworthy that this transcript lacks the sequence encoding amino acids 1-68 of all previously reported MOBP isoforms (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997), including MOBP 170.

2.5.4 Identification of alternative splicing in the 5' untranslated sequence of *Mobp* pre-mRNAs

As described above (2.5.3) mRNAs corresponding to tsp at 60 nt upstream of T_{1050} (Figure 2.5.2) contain sequence from exon 1 and exon 3 but lack sequence from exon 2. It should be noted that the exon defined as exon 1, possesses a 5' donor splice sequence (CAG|; Section 2.4, Table 2.4.3; Figure 2.5.2) at its 3' end. These data indicate that the 5'UTR sequence of all mRNAs generated by splicing events within the *Mobp* transcript contain sequence from exon 1 and a 4 nt portion of exon 3. Sequence from exon 2 is present in < 95% of mRNAs examined in this study. Whether the presence or absence of sequence from exon 2 is regulated developmentally or restricted to a subpopulation of *Mobp* splice variants is not known. Exon 2 may be separated from exon 1 by an intronic sequence or may be contiguous with exon 1 and demonstrate alternative use of an available 5' donor splice site. If the latter is correct, prediction may be made of the size of a PCR product amplified using oligonucleotides designed to sequence within exons 1 and 2. In order to address this question oligonucleotides for PCR were designed to sequence within exon 1 and exon 2. In the event that these exons are contiguous, the size of the amplified genomic DNA fragment should be no more than 135 bp in length. Figure 2.5.4 illustrates the results of a sequence alignment of 5'UTR sequences of the published splice variants of the rat *Mobp* gene and a cDNA fragment from the mouse *Mobp* transcript, which contains the predicted exons 1 and 2.

Figure 2.5.1 PCR

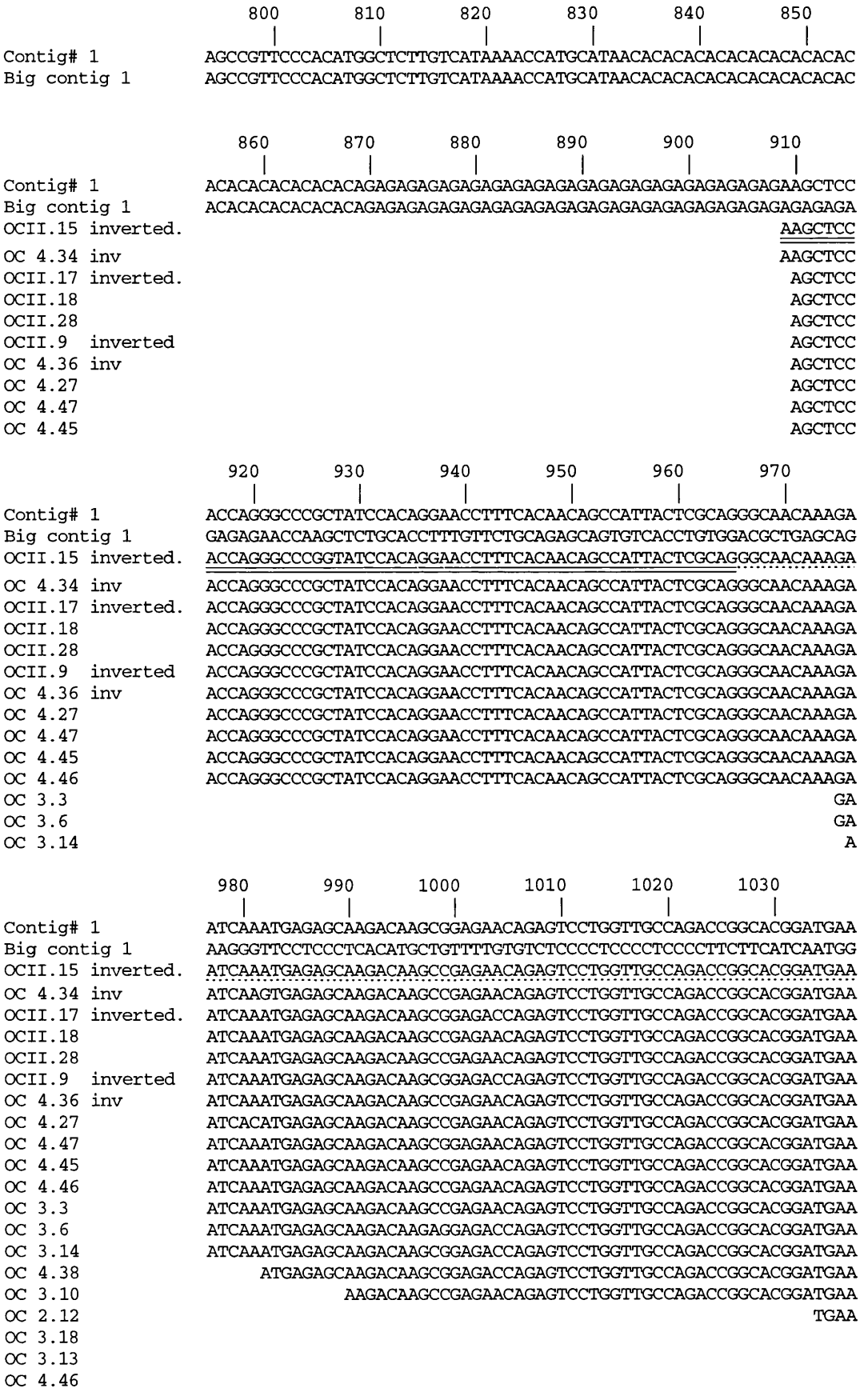

800 810 820 830 840 850
Contig# 1 AGCCGTTCCCACATGGCTCTTGTCATAAAACCATGCATAACACACACACACACACACAC
Big contig 1 AGCCGTTCCCACATGGCTCTTGTCATAAAACCATGCATAACACACACACACACACACAC
860 870 880 890 900 910
Contig# 1 ACACACACACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAGCTCC
Big contig 1 ACACACACACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
OCII.15 inverted. AAGCTCC
OC 4.34 inv AAGCTCC
OCII.17 inverted. AGCTCC
OCII.18 AGCTCC
OCII.28 AGCTCC
OCII.9 inverted AGCTCC
OC 4.36 inv AGCTCC
OC 4.27 AGCTCC
OC 4.47 AGCTCC
OC 4.45 AGCTCC
920 930 940 950 960 970
Contig# 1 ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
Big contig 1 GAGAGAACCAAGCTCTGCACCTTTGTTCTGCAGAGCAGTGTCACCTGTGGACGCTGAGCAG
OCII.15 inverted. ACCAGGGCCCGGTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OC 4.34 inv ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OCII.17 inverted. ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OCII.18 ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OCII.28 ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OCII.9 inverted ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OC 4.36 inv ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OC 4.27 ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OC 4.47 ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OC 4.45 ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OC 4.46 ACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCAACAAAGA
OC 3.3 GA
OC 3.6 GA
OC 3.14 A
980 990 1000 1010 1020 1030
Contig# 1 ATCAAATGAGAGCAAGACAAGCGGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
Big contig 1 AAGGGTTCCTCCCTCACATGCTGTTTTGTGTCTCCCCTCCCCTCCCCTTCTTCATCAATGG
OCII.15 inverted. ATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 4.34 inv ATCAAGTGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OCII.17 inverted. ATCAAATGAGAGCAAGACAAGCGGAGACCAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OCII.18 ATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OCII.28 ATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OCII.9 inverted ATCAAATGAGAGCAAGACAAGCGGAGACCAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 4.36 inv ATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 4.27 ATCACATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 4.47 ATCAAATGAGAGCAAGACAAGCGGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 4.45 ATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 4.46 ATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 3.3 ATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 3.6 ATCAAATGAGAGCAAGACAAGAGGAGACCAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 3.14 ATCAAATGAGAGCAAGACAAGCGGAGACCAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 4.38 ATGAGAGCAAGACAAGCGGAGACCAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 3.10 AAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCACGGATGAA
OC 2.12 TGAA
OC 3.18
OC 3.13
OC 4.46

Figure 2.5.2

```
                                980         990         1000        1010        1020        1030
                                 |           |           |           |           |           |

OC 4.41 inv                              AGCTCCACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTA
OC 3.1                                   AGCTCCACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTA
OC 3.19                                  AGCTCCACCAGGGCCCGCTATCCACAGGAACCTTTCACAACAGCCATTA
OC 13                                    AGCTCCACCAGGGCCCGCTAGCCACAGGAACCTTTCACAACAGCCATTA

                           1040        1050        1060        1070        1080        1090
                            |           |           |           |           |           |
Contig# 1                  AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
Big contig 1               TTTCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OCII.15 inv                AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTTCAAGAACCAGAA
OC 4.34 inv                AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OCII.17 inv                AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OCII.18                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OCII.28                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OCII.9  inv                AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 4.36 inv                AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 4.27                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 4.47                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 4.45                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 4.46                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 3.3                     AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 3.6                     AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 3.14                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 4.38                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGTCCCAGGCTCTCCAAGAACCAGAA
OC 3.10                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 2.12                    AACCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCGAGAACCAGAA
OC 3.18                     ACCCACTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACAAGAA
OC 3.13                      CCCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 4.46                       CCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA

OC 4.41 inv                CTCGCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 3.1                     CTCGCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 3.19                    CTCGCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 13                      CTCGCAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA

OC 13                          CAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 2.12                        CAGTGAGATGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCGAGAACCAGAA
OC 4.43 inv                          TGAGTCAGAAAATGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 37   inv                                    AAGGAGGGCCCCAGGCTCTCCAAGAACCAGAA
OC 6                                                CCCCAGGCTCTCCAAGAACCAGAA
OC 15                                               CCCAGGCTCTCCAAGAACCAGAA
OC 35                                                CCAGGCTCTCCAAGAACCAGAA
OC 44                                                CCAGGCTCTCCAAGAACCAGAA
OC 4.48                                              CCAGGCTCTCCAAGAACCAGAA
OC 4.44                                              CCAGGCTCTCCAAGAACCAGAA
OC 4.42                                              CCAGGCTCTCCAAGAACCAGAA
OC 3.8                                               CCAGGCTCTCCAAGAACCAGAA
OC 3.12                                              CCAGGCTCTCCAAGAACCAGAA
OC 3.11                                              CCAGGCTCTCCAAGAACCAGAA
OC 19                                                CCAGGCTCTCCAAGAACCAGAA
OC 9                                                 CCAGGCTCTCCAAGAACCAGAA
OC 41                                                CCAGGCTCTCCAAGAACCAGAA
OC 39 inv                                            CCAGGCTCTCCAAGAACCAGAA
OC 32 inv                                            CCAGGCTCTCCAAGAACCAGAA
OC 22                                                CCAGGCTCTCCAAGAACCAGAA
OC 15                                                CCAGGCTCTCCAAGAACCAGAA
OC 14                                                CCAGGCTCTCCAAGAACCAGAA
OC 12                                                CCAGGCTCTCCAAGAACCAGAA
OC 3.5                                               CCAGGCTCTCCAAGAACCAGAA
OC 39 inv                                            CCAGGCTCTCCAAGAACCAGAA
OC 41                                                CCAGGCTCTCCAAGAACCAGAA
```

Figure 2.5.2

```
                  1040       1050       1060       1070       1080       1090
                    |          |          |          |          |          |
OC 9                                                            CCAGGCTCTCCAAGAACCAGAA
OC 22                                                           CCAGGCTCTCCAAGAACCAGAA
OC 32 inv                                                       CCAGGCTCTCCAAGAACCAGAA
OC 44                                                           CCAGGCTCTCCAAGAACCAGAA
OC 35                                                           CCAGGCTCTCCAAGAACCAGAA
OC 17                                                           CCAGGCACTCCAAGAACCAGAA
OC 12                                                           CCAGGCTCTCCAAGAACCAGAA
OC 14                                                           CCAGGCTCTCCAAGAACCAGAA
OC 10                                                           CCAGGCTCTCCAGGAACCAGAA
OC 19                                                             AGGCTCTCCAAGAACCAGAA
OC 16 inv                                                               CAAGAACCAGAA
OC 4.40                                                                  AAGAACCAGAA
OC 46                                                                       AACCAGAA
OC 47 inv                                                                        AA
OC 48 inv                                                                        AA

                 1100       1110       1120       1130       1140       1150
                    |          |          |          |          |          |
Contig# 1          GTTCTCCGAGCACTTCAGCATCCACTGAAGCCCACCCTTCACCTTCCTCAACTCCAAGCGT
Big contig 1       GTTCTCCGAGCACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGT
OCII.15 inv        GTTCTCCGAGCACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGT
OC 4.34 inv        GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OCII.17 inv        GTTCTCCGAGCACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGT
OCII.18            GTTCTCCGAGCACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGT
OCII.28            GTTCTCCGAGCACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGT
OCII.9  inv        GTTCTCCGAGCACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCGAAGCGT
OC 4.36 inv        GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.27            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.47            GTTCTCCGAGCACTTCAGCATCCACTGCTGCC
OC 4.45            GTTCTCCGAGCACTTCAGCATCCACTGCTGCC
OC 4.46            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.3             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.6             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.14            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.38            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.10            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 2.12            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.18            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.13            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.46            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.41 inv        GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.1             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.19            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 13              GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 13              GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 2.12            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.43 inv        GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 37   inv        GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 6               GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 15              GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 35              GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 44              GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.48            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.44            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.42            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.8             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.12            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.11            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 19              GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 9               GTTCTCCGAGCACTACAGCATCCACTGAAGCC
OC 41              GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 39 inv          GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 32 inv          GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 22              GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
```

Figure 2.5.2

```
                1100       1110       1120       1130       1140       1150
                  |          |          |          |          |          |
OC 15             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 14             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 12             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 3.5            GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 39 inv         GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 41             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 9              GTTCTCCGAGCACTACAGCATCCACTGAAGCC
OC 22             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 32 inv         GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 44             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 35             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 17             GTTCTCCGAGCACTTCATCATCCACTGAACCC
OC 12             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 14             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 10             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 19             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 16 inv         GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 4.40           GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 46             GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 47 inv         GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 48 inv         GTTCTCCGAGCACTTCAGCATCCACTGAAGCC
OC 38                    GAGCACTTCAGCATCCACTGAAGCC
OC7                      GAGCACTTCAGCATCCACTGAAGCC
OC 2.2                   GAGCACTTCAGCATCCACTGAAGCC
OC 2.1                   GAGCACTTCAGCATCCACTGAAGCC

               1160       1170       1180       1190       1200       1210       1220
                  |          |          |          |          |          |          |
Contig# 1         GAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTAAGCCCAGAAGAAGGAGG
Big contig 1      GAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGG
OCII.15 inverted. GAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTAAGCC
OC 4.34 inv       GAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTAAGCC
OCII.17 inverted. GAGATCGTGGATCGCAAGTACAGCATCTGCAAGAGCGGTTGCTAAGCC
OCII.18           GAGATCGTGGATCGCAAGTACAGCATCTGCAAGAGCGGTTGCTAAGCC
OCII.28           GAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTAAGCC
OCII.9 inverted.  GAGATCGTGGACCGCAAGTACAGCATCTGCAAGAGCGGTTGCTAAGCC

                       1230       1240       1250       1260       1270       1280
                         |          |          |          |          |          |
Contig# 1         AGGACTGGATCTGCTGTGCCTGCCAGAAGACTAGGTGAGTGCGGGCTCTGGGCCAGGTCTA
Big contig 1      AGGACTGGATCTGCTGTGCCTGCCAGAAGACTAGGTGAGTGCGGGCTCTGGGCCAGGTCTA
```

Figure 2.5.2 Multi-alignment of sequences from generated by oligo-capping assay to identify transcription start points (tsp) of the *Mobp* transcript. Exons 1 and 2 are underlined with solid (double) and dotted lines respectively. **ATG**$_{1049\text{-}1051}$, where T_{1050} is the point relative to which all tsp positions are defined and the subscript number 1050 corresponds to the nucleotide position in Contig 1. N.B. Nucleotide numbers, listed above sequence, are purely arbitrary and bear no relevance to the genomic organisation of the gene.

```
1030      1040          1050          1060          1070          1080
 |          |             |             |             |             |
ATG ACT GTC AGC TCC CTG TTA GGA AGG TTC TGG AGG GCG AGC AGA AGG TGG GAG GGA GGG
Met Thr Val Ser Ser Leu Leu Gly Arg Phe Trp Arg Ala Ser Arg Arg Trp Glu Gly Gly

1090      1100          1110          1120          1130          1140
 |          |             |             |             |             |
CCG GGC CTT CCT GCT GTT TCC AGT TCC TGC CAG CAT CCC TTA AGC GAC GGT GGA GAG CCC
Pro Gly Leu Pro Ala Val Ser Ser Ser Cys Gln His Pro Leu Ser Asp Gly Gly Glu Pro

1150      1160          1170          1180          1190          1200
 |          |             |             |             |             |
AGG CTG CAG CCT CCA GCT TCT TTT GGC CCT CTC TGC ACC AGC CGC CGT GCC ACG TCC CCT
Arg Leu Gln Pro Pro Ala Ser Phe Gly Pro Leu Cys Thr Ser Arg Arg Ala Thr Ser Pro

1210      1220          1230          1240          1250          1260      1270
 |          |             |             |             |             |
CAG AGG CCC AAG CAC CAG CCA GCT GCG TCC CCA GTG GTG GTC AGA GCG CCG CCA GCC AAG
Gln Arg Pro Lys His Gln Pro Ala Ala Ser Pro Val Val Val Arg Ala Pro Pro Ala Lys

1270      1280          1290          1300          1310          1320
 |          |             |             |             |             |
CCA AAG TCC CCT CTG ATG CCA GCC AAG CCA AGG TCC CCA CCG AGG CCA GCC AAG CCA AGG
Pro Lys Ser Pro Leu Met Pro Ala Lys Pro Arg Ser Pro Pro Arg Pro Ala Lys Pro Arg

1330      1340          1350          1360          1370          1380
 |          |             |             |             |             |
TCC CCT TCA AGG ACT GCT GAT GCC AGC CAA GCC AAG GTC CCC ACC GAG GCC AGC CAA GCC
Ser Pro Ser Arg Thr Glu Arg Gln Pro Arg Pro Arg Pro Glu Val Arg Pro Pro Pro Ala

1390      1400          1410          1420          1430          1440
 |          |             |             |             |             |
AAG GTC CCC TTC AAG GAC TGA GCG CCA GCC GCG TCC CCG CCC AGA GGT CCG ACC ACC ACC
Lys Gln Lys Pro Pro Gln Lys Ser Lys Gln Pro Ala Arg Ser Ser Pro Leu Arg Gly Pro

1450      1460          1470          1480          1490          1500
 |          |             |             |             |             |
AGC CAA GCA GAA GCC CCC TCA GAA ACC AGA GCT CCT AGG TTC TGG TAA CAC C
Gly Thr Ser Arg Gly Gly Ser Pro Thr Arg Ala Pro Arg Phe Trp
```

Figure 2.5.3 Alignment of DNA and amino acid sequences corresponding to the novel transcript identified in Section 2, 2.5. The translation start point (ATG) and stop codon (TAA), predicted to be utilised by the novel transcript are indicated in bold type. Nucleotide (nt) numbers above these sequences correspond to the nucleotide sequence within Contig 1 (Appendix VII). The 3' acceptor site corresponding to exon 4 of the *Mobp* gene is at nt 1191.

```
                                10        20        30        40        50        60
                                |         |         |         |         |         |
OC II 15inv 5'UTR  AAGCTCCACCAGGGCCCGGTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCA
MOBP69 5'                                TATCCACAGGAACCTTTCACAGCAGCCAATACCTGCAGGGCA
MOBP81 B 5'                                   AGGGAACCTTTCACAGCAGCCAATACCTGCAGGGCA
MOBP81A 5'                                       CTTTCACAGCAGCCAATACCTGCAGGGCA
MOBP99 5'                                                GCCAATACCTGCAGGGCA
rOP1 5'
rOPRP1 5'

                                70        80        90        100       110       120
                                |         |         |         |         |         |
OC II 15inv 5'UTR  ACAAAGAATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCAC
MOBP69 5'          ACAAAGAATCAAATGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
MOBP81 B 5'        ACAAAGAATCAAATGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
MOBP81A 5'         ACAAAGAATCAAATGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
MOBP99 5'          ACAAAGAATCAAATGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
rOP1 5'                       TGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
rOPRP1 5'                     TGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/

                          130
                           |
OC II 15inv 5'UTR  GGATGAAAACCCAG|***
MOBP69 5'          GCTTGAAAACACAG|***
MOBP81 B 5'        GCTTGAAAACACAG|***
MOBP81A 5'         GCTTGAAAACACAG|***
MOBP99 5'          GCTTGAAAACACAG|***
rOP1 5'            GCTTGAAAACACAG|***
rOPRP1 5'          GCTTGAAAACACAG|***
```

Figure 2.4.4 Sequence multi-alignment of 5' UTR sequences from published (Yamamoto *et al.*, 1994; Holz *et al.*, 1996) rat splice variants and sequence from a 5' cDNA fragment generated in an oligo-capping assay to identify transcriptional start sites (Section 2, 2.5). 5' donor splice sites are indicated by |***. N.B. Nucleotide numbers, listed above sequence, are purely arbitrary and bear no relevance to the genomic organisation of the gene. /, symbol arbitrarily introduced into 5'UTR sequence of rat splice variants to maintain the alignment of rat and mouse sequences. This does not represent an unidentified nucleotides. Exons 1 and 2, of the murine *Mobp,* gene are underlined with solid (double) and dotted lines respectively.

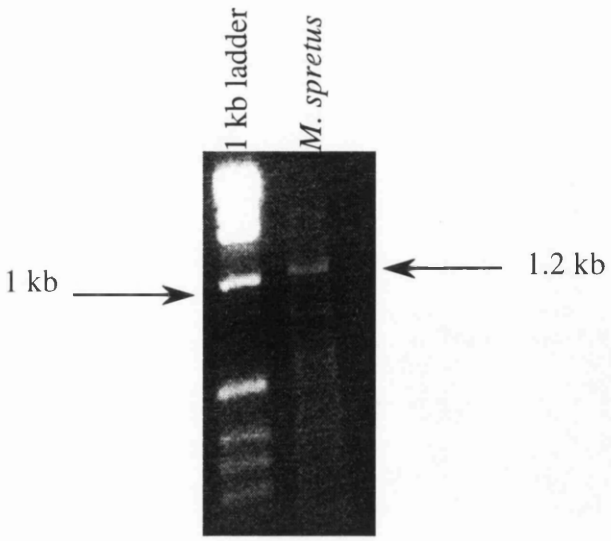


Figure 2.5.5 1.2 kb DNA fragment which was PCR amplified from mouse genomic (*M. spretus*) template DNA using oligonucleotides ASM 242 and ASM 244 (defined in Chapter 2, Table 2.4). 1 kb ladder, molecular weight ladder.

2.5.4.1 PCR amplification and sequencing of a mouse genomic fragment using oligonucleotides specific to the 5'UTR of *Mobp*

In an endeavour to assess whether the exons defined as exon 1 and exon 2 were interrupted by an intron, PCR amplification of the genomic region flanked by these exonic sequences was performed. Oligonucleotides ASM 242 and ASM 244 (Table 2.3, Chapter 2) were designed to sequence within exon 1 and exon 2 respectively. A 1.2 kb fragment was detected upon PCR amplification of mouse (C57BL/6) genomic DNA, using these oligonucleotides (Figure 2.5.5). It should be noted that none of the reported *Mobp* transcripts (rat: Yamamoto *et al.*, 1994; Holz *et al.*, 1996 and mouse: Montague *et al.*, 1997) demonstrate 5'UTRs in excess of 120 nt. Direct sequence analysis of this PCR fragment has failed to provide interpretable results. Time limitation precluded pursuit of this line of study. However, the study of this DNA fragment, and the 5' region of this gene, has continued in collaboration with the laboratory of Dr I. Griffiths (University of Glasgow). This work will involve cloning the PCR fragment into the vector pCR™2.1 (TA cloning vector; defined in Table 2.1, Chapter 2) for subsequent sequence analysis. Only upon confirmation of the identity of this DNA fragment, may it be used as a probe for hybridisation to cleaved genomic or cloned DNAs, and the genomic structure of the 5' region of the gene be fully elucidated.

As discussed in Section 2.4 the evidence, provided by the PCR amplification of a 1.2 kb genomic fragment using primers ASM 242 and ASM 244 and by the presence of an extended exonic 5' donor sequence (C/AAGl) in exon 1, supports the above hypothesis, that the identified 5'UTRs (rat: Yamamoto *et al.*, 1994; Holz *et al.*, 1996 and mouse: refer also to Section 2.5]) are produced by discrete exons that are interrupted by an intron. Subsequently, the exonic segments to which these oligonucleotides were designed have been arbitrarily identified as discrete exons. Data arising from this study demonstrate that only sequences within exons 1 and a 4 nt portion of exon 3 are present in the 5'UTR sequence of all mRNAs generated by splicing events within the *Mobp* transcript. This suggests that alternative splicing, including or excluding sequence from exon 2, may be involved in the generation of splice variants of the *Mobp* transcript. To date, the mode of regulation of these alternative splicing events remains unknown. As reported in Section 2.4, exons downstream of the arbitrarily named exons 1 and 2 have been named exons 3-8 in their corresponding linear order in the mouse genome. Appendix XI illustrates a sequence multi-alignment of 5'UTR sequences from exons 1 and 2 of the rat and mouse *Mobp* genes. These sequences demonstrate > 90% sequence identity.

2.5.5 Discussion

Three groups of tsp have been identified (140 nt [> 17%], 75-68 nt [> 7%] and 60 nt [> 5%]) upstream of T_{1050}, where $\underline{ATG}_{1049\text{-}1051}$ is the translation start site used in the production of all previously identified MOBP isoforms (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997). These data present the first report of the use of alternative splicing in the generation of 5'UTR for *Mobp* transcripts. It is noteworthy that 5'UTRs, lacking sequence from exon 2, are present in >5% of the cDNA fragments sequenced. The large number of cDNA fragments sequenced in this study is far in excess of previous reports (Yamamoto *et al.*, 1994; Holz *et al.*, 1996).

Two further major tsp were identified downstream of T_{1050}, where $\underline{ATG}_{1049\text{-}1051}$ is the translation start site used in the production of all previously identified MOBP isoforms (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997). These tsp, at 25 nt and 101 nt, downstream of T_{1050}, comprise >35% and >10% of start points respectively. The 5' end of previously uncapped RNA should, in theory, be unavailable for ligation of an r-oligo. The hypothesis that these sequences represent a portion of a previously unidentified transcript is strengthened by the subsequent amplification of a cDNA species using an oligonucleotide, specific to sequence within this putative novel transcript in conjunction with an oligonucleotide designed to sequence within the fourth exon of *Mobp*. Preliminary study suggests that this putative transcriptional unit is indeed transcribed (Montague *et al.*, unpublished data). Further study is required in order to assess whether all tsp identified, at points upstream of T_{1050}, are associated with previously identified *Mobp* splice variants. In order to address this question 3' oligonucleotides, designed to sequence that is unique to the novel transcript i.e. within intron 3 of *Mobp*, will be used in a further oligo-capping experiment. This experiment may exclude or include the use of tsp at points upstream of T_{1050} in the production of the novel transcript. Northern blot analysis, utilising as probe sequence that is unique to the putative novel transcript, is presently being performed in an endeavour to establish the veracity of the RT-PCR results. Confirmation of the transcription of this sequence unit and its translation to protein may yet increase the complexity of the question that must be asked in studies designed to elucidate the function of the *Mobp* gene and related transcriptional units.

As yet, the identification of exon sequences 1 and 2, as discrete exons, remains incompletely proven. Further analysis of the promoter region of *Mobp* will require

elucidation of the locations of exons 1 and 2. In an endeavour to achieve this, study of the 5' region of the *Mobp* gene continues in collaboration with the laboratory of Dr I Griffiths (University of Glasgow).

2.6 Targeted disruption of *Mobp* in embryonic stem (ES) cells

2.6.1 Gene disruption and assessment of gene function

Gene targeting provides the potential to generate mice of a desired phenotype (Capecchi, 1989). The discovery that somatic cells possess the ability to mediate homologous recombination between non-replicating DNA molecules (Folger *et al.*, 1982) has facilitated the development of gene targeting technologies. Gene targeting involves: (1) the use of recombinant DNA technology to introduce the desired mutation into a cloned DNA of a selected locus; (2) transfer of the mutation to the genome of a pluripotent, embryonic stem (ES) cell; (3) microinjection of the mutant cell into blastocysts with the aim of generating chimeric mice, containing the targeted allele within the genome of germline cells; (4) the mating of heterozygote siblings to generate mice, homozygous for the null allele in question (Capecchi, 1989). Homologous recombination between a cognate DNA sequence and an endogenous chromosomal locus may now be used in the correction, disruption, modification or replacement of mammalian genes (Capecchi, 1989; Rothstein, 1991; Bradley, 1993; Scherer and Adair, 1994). A number of recent studies (Aguzzi *et al.*, 1994; Higgins and Cordell, 1995; Aguzzi *et al.*, 1996; Lees *et al.*, 1996; Price *et al.*, 1996) have indicated the utility of gene targeting and transgenic technologies in the study of the mammalian nervous system. In recent years many studies have been directed towards gene disruption [Knockout, KO] (refer also to the review article by Scherer and Adair, 1994). Such null mutations define the basal function of a gene (Deng *et al.*, 1993).

Mutations in genes encoding myelin components have been demonstrated to underlie a number of inherited neuropathies in man (Griffiths *et al.*, 1995; Zielasek *et al.*, 1996) (refer also to Chapter 1, 1.7) and mouse (Mikoshiba *et al.*, 1991) (refer also to Chapter 1, 1.7). Loss of function mutations, reported in the genes encoding myelin protein zero [P0] (Giese *et al.*, 1992; Martini *et al.*, 1995a; Zielasek *et al.*, 1996) and peripheral myelin protein [PMP22] (Adlkofer *et al.*, 1995), have been suggested to resemble the human neuropathies Charcot-Marie-Tooth disease type 1B and Déjérne-Sottas and Charcot-Marie-Tooth type 1A respectively. It is also known that mutations in the gene encoding myelin proteolipid protein [PLP/DM20] are responsible for neuropathies in a number of species including man (Pelizaeus Merzbacher disease, X-linked Spastic Paraplegia). However, loss of function mutations, in this gene, demonstrate relatively minor clinical effects (Griffiths *et al.*, 1995). It has been suggested that MOBP plays a role in stabilising the myelin sheath (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997). It

can therefore reasonably be predicted that the generation of a null mutation in the gene encoding MOBP may interfere with the compact nature and appearance of the MDL of central myelin. Like *shiverer*, the CNS myelin of an *Mobp*$^{-/-}$ mouse may be diffuse and lack an MDL.

2.6.2 Design of a targeting construct

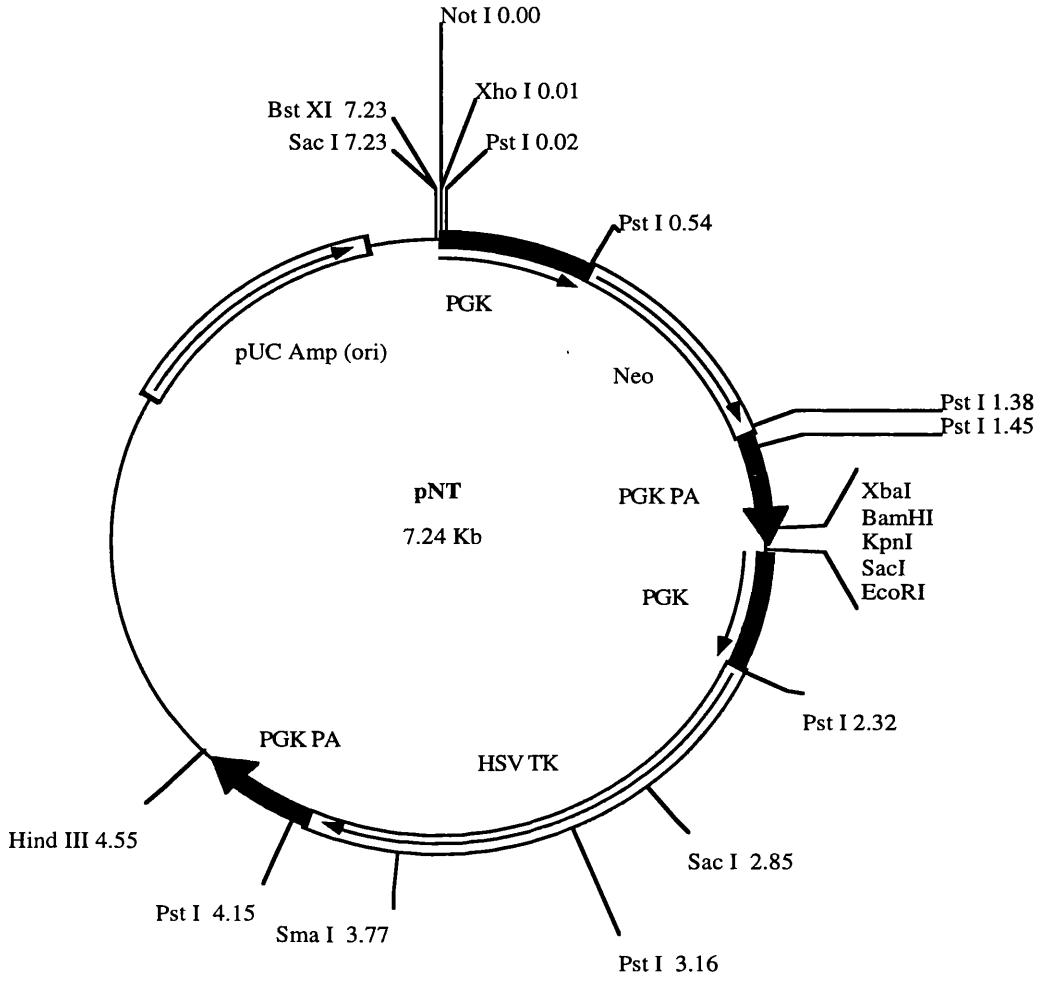
One recent report (Deng and Capecchi, 1992) suggested that a number of well defined questions must be addressed by any experimenter wishing to undertake a gene targeting project. These include: (1) whether to use an insertion or a replacement vector; (2) how much homology to use in the generation of the construct; (3) whether there is any advantage in the use of isogenic DNA, in preference to non-isogenic DNA; (4) whether it is more expedient to use enrichment procedures in cell culture or to screen larger numbers of cell lines in the search for targeted cell lines.

I will address these concerns in the order in which they are given above. Firstly, a number of studies (Thomas and Capecchi, 1987; Deng *et al.*, 1993) have assessed the utility of both vector types. It has been suggested that increased number of cell culture stages, required when using insertion vectors, may decrease the ability of a cell line to contribute to the germline (Deng *et al.*, 1993). Replacement vectors require fewer culture stages. In generating the desired mutation a replacement vector retains structure of the targeting construct, between the sites of recombination. This recombinant structure includes the gene cassette encoding neomycin resistance (neo^r). This is not a disadvantage in the generation of a gene KO. In this instance the investigator desires loss of expression of the gene in question and may select the insertion site for the neo^r gene cassette within the transcriptional unit. Secondly, it has been demonstrated (Thomas and Capecchi, 1987; Hasty *et al.*, 1991) that there is an exponential relationship between the length of homologous sequence, in a targeting vector, and the targeting frequency. This exponential relationship exists across the range of 2-10 kb. It is further noted that the exponential nature of this relationship saturates (plateaus) at approximately 14 kb (Thomas and Capecchi, 1987). Thirdly, it has been demonstrated (Letsou and Liskay, 1987) that intrachromosomal gene conversion is sensitive to single basepair mismatch. A number of recent studies (Waldman and Liskay, 1987; Deng and Capecchi, 1992; Miller, 1992; te Reile *et al.*, 1992; van Deursen and Wieringa, 1992) have compared the efficiency of using isogenic and non-isogenic DNA in targeting vectors. These suggest that efficient recombination requires uninterrupted homology (Waldman and Liskay, 1987) and further conclude that even limited heterogeneity can dramatically reduce targeting efficiency

(Deng and Capecchi, 1992; Miller *et al.*, 1992; te Reile *et al.*, 1992; van Deursen and Weringen, 1992). Finally, the use of positive-negative selection (PNS) is known to enrich for cells containing targeted mutations (Mansour *et al.*, 1988; Capecchi, 1989). Positive selection utilises the presence of a novel selectable marker in assaying for the integration of a targeting vector into the host genome. Negative selection, selects against random insertion targeting events (Mansour *et al.*, 1988; Capecchi, 1989). The gene targeting vector, for the murine *Mobp* gene (p*Mobp*-N/TK), was designed and generated as described below.

2.6.3 Generation of a targeting construct

Based on the information given in the studies cited above, I undertook to generate a replacement construct, using isogenic DNA. The purpose of this construct was to generate a loss of function (KO) mutation in the *Mobp* gene by deleting a genomic region encompassing exons 3, 4 and 5. As discussed in Section 2.4, exon 3 encodes amino acid residues 1 - 68; present in all MOBP isoforms. The murine gene encoding myelin-associated oligodendrocytic basic protein was isolated and characterised as described in Section 2 (2.4 and 2.5). The genomic region encompassing exons 3-8 of the *Mobp* gene is illustrated in Figure 2.6.2 A. The genomic libraries used in the isolation of the gene, were prepared using DNA from the mouse strain 129/Sv. It should be noted that the ES cells used are also derived from this mouse strain. Figure 2.6.1 illustrates the vector backbone (pNT) used in the construction of the targeting construct (p*Mobp*-N/TK). Figure 2.6.2 illustrates the steps involved in its construction. Firstly, a 3.9 kb *Bam* HI fragment, isolated from clone AS17 which lies approximately 2 kb upstream of the first coding exon (exon 3 [Figure 2.6.2 B, Step 1]), was subcloned into the unique *Bam* HI site of the pNT vector. Secondly, a 6.9 kb *Xho* I-*Sal* I fragment, isolated from AS 35 (encompassing exons 6-8 [Figure 2.6.2 B, Step 2]), was subcloned into the unique *Xho* I site of the pNT vector. It is noteworthy that the *Xho* I site present in this fragment derives from the multiple cloning site of the λ 2001 bacteriophage vector. The size and orientation of fragments (illustrated in Figure 2.6.2) were confirmed after subcloning, in each instance, by restriction enzyme cleavage with *Sst* I. Figure 2.6.3 (A, B and C) illustrate the results of *Sst* I restriction enzyme cleavage of the completed construct, indicating the expected sizes of fragments. The completed construct is shown linearised, at the unique *Not* I site in the pNT vector (Figure 2.6.2 C), and circularised (Figure 2.6.4). The targeting vector p*Mobp*-N/TK comprises 10.8 kb of homology to the *Mobp* gene locus.



Plasmid name: pNT
Plasmid size: 7.24 kb

Figure 2.6.1 Schematic representation of the vector backbone (pNT) used in the construction of the targeting construct (p*Mobp*-N/TK).

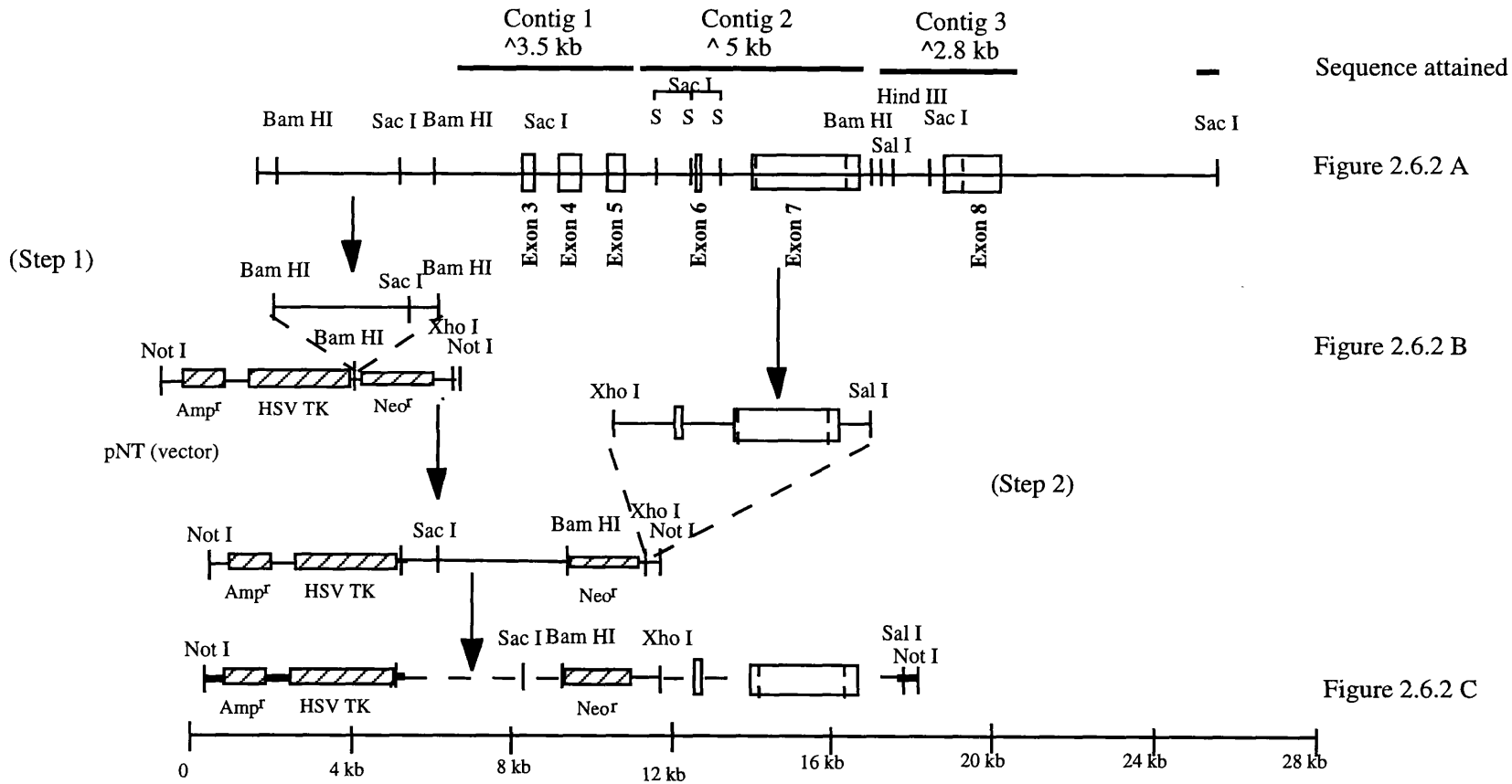


Figure 2.6.2 Schematic representation of the steps involved in the generation of the p*Mobp*-N/TK targeting construct. Figure 2.6.2 A illustrates the organisation of the *Mobp* gene region encompassing exons 3-8. Figure 2.6.2 B (Step 1) represents the isolation and cloning of a 3.9 kb *Bam* H1 fragment into the unique *Bam* H1 site of vector (pNT). Figure 2.6.2 B (Step 1) represents the isolation and cloning of a 6.9 kb *Xho* I - *Sal* I fragment cloned into the unique *Xho* I site of vector (pNT). N.B. *Xho* I site of *Xho* I - *Sal* I fragment is present in the multiple cloning site (MCS) of λ 2001 and exists in the *Xho* I - *Sal* I fragment as an endpoint of clone AS 35. Figure 2.6.2 C represents the completed construct p*Mobp*-N/TK, linearised at the unique *Not* I site therein. Sequence attained indicates the correspondence between contigs 1-3 and the *Mobp* gene region encompassing exons 3-8. Dotted lines and solid black lines (Figure 2.6.2C) correspond to genomic and vector DNA respectively

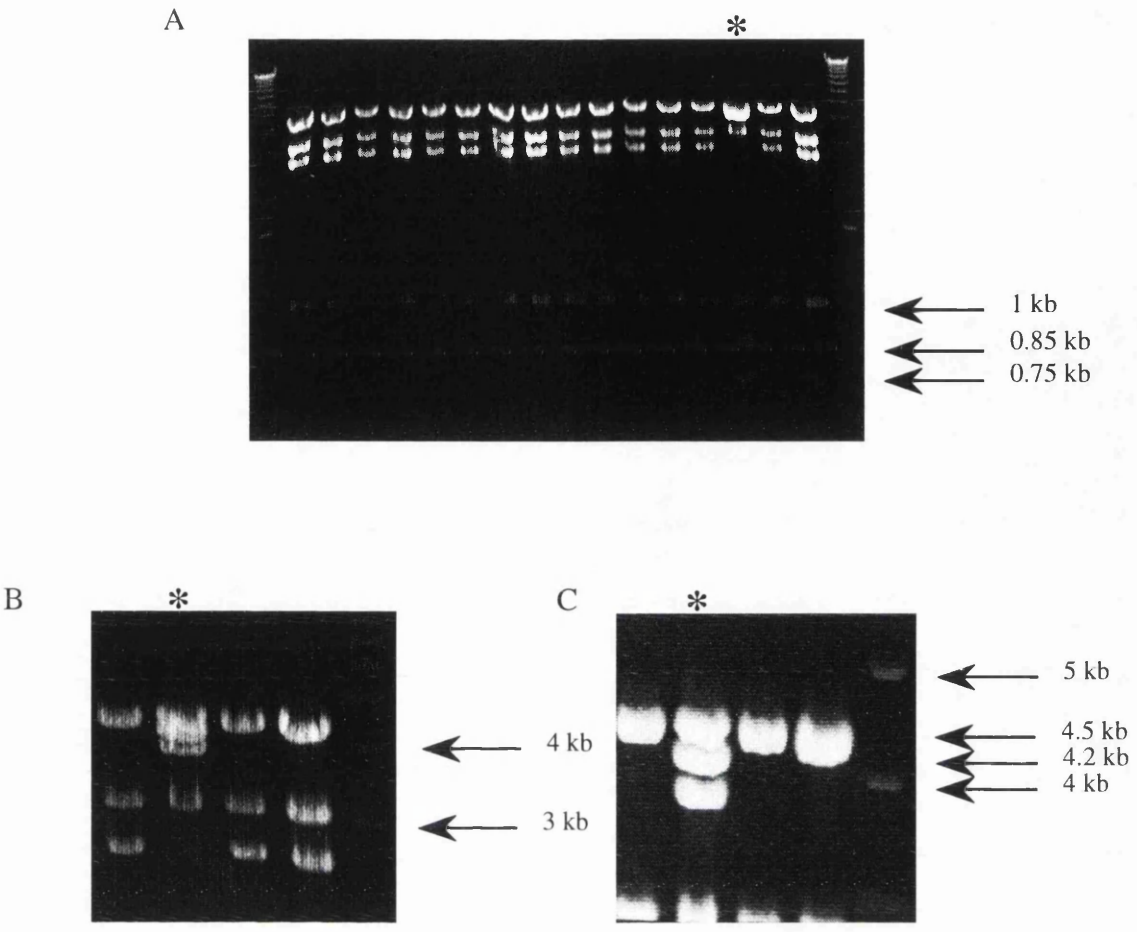
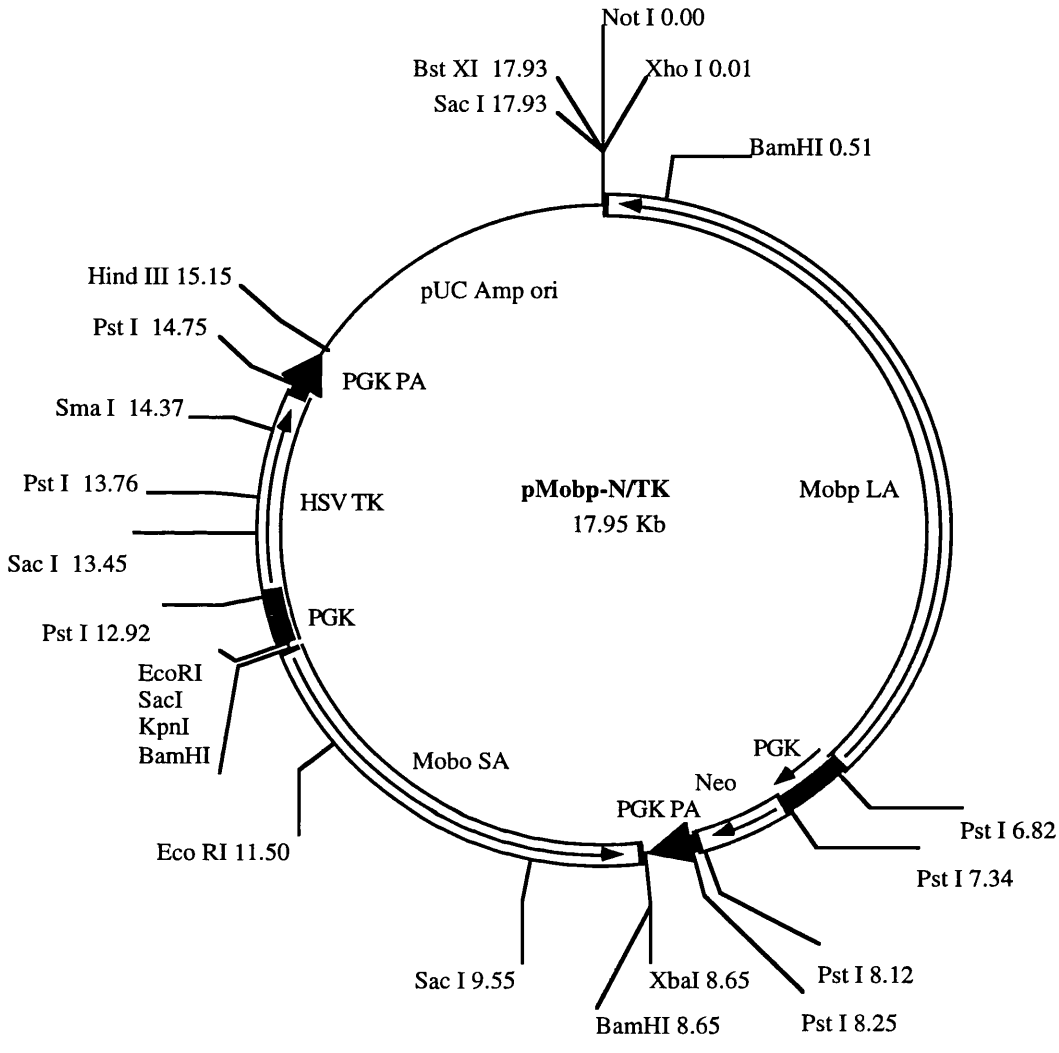


Figure 2.6.3 Identification of cleaved DNA from completed construct p*Mobp*-N/TK by electrophoresis of DNAs cleaved with the restriction enzyme *Sst* I. Restriction enzyme cleavage of the completed construct with *Sst* I yields 7 fragments of 0.75 - 4.5 kb in length (Lane *). The three images (A,B and C) illustrate the progress of DNA, with time, through a 1% agarose gel. The fragments are, as expected, 0.75 kb, 0.85 kb, 1 kb, 3 kb, 4 kb, 4.2 kb and 4.5 kb.



Plasmid name: pMobp-N/TK
Plasmid size: 17.95 kb

Figure 2.6.4 Schematic representation of the circularised completed construct p*Mobp*-N/TK, corresponding to Figure 2.6.2 C in which the completed construct is linearised at the unique *Not* I site therein.

2.6.4 Embryonic stem (ES) cell culture

The vector p*Mobp*-N/TK was linearised by cleavage with the restriction enzyme *Not* I (Figure 2.6.2 C) and introduced into ES cells via electroporation (as described in Chapter 2). Aliquots of the resulting cell suspension were distributed evenly between 15 tissue culture plates and nutrient medium added (as described in Chapter 2). ES cell culture was performed as described in Chapter 2. N.B. one plate was identified as an experimental positive control and was incubated with a medium lacking Gancyclovir. When the nutrient medium was replaced by medium, containing G418 (0.2 mg/ml) and Gancyclovir (0.55 μg/ml) two different preparations of media were used. These media, arbitrarily named A and B, were prepared individually but utilised the same stock materials (defined in Chapter 2). Two days after initiating selection, the cells grown in medium A perished. The cells were observed to lose their characteristic shape becoming rounded and subsequently underwent cell death. The cells incubated in the presence of medium B suffered no such reaction. Distinct colonies, resistant to the drugs used in this study, could be observed with the naked eye within eight days of electroporation. The average number of colonies, demonstrating resistance to G418 (neomycin; positive selection), was found to be 40 per experimental plate (total number of colonies = 280). The control plate contained 115 colonies demonstrating resistance to G418 (neomycin). This indicated that the enrichment factor, for correctly targeted events, was approximately 3 fold. Enrichment, for correctly targeted events, was achieved by selection against the presence of the Herpes Simplex virus (HSV I) gene encoding Thymidine kinase, using the drug Gancyclovir. Gancyclovir is cytotoxic to ES cells expressing HSV-TK (Mansour *et al.*, 1988). In total 240 colonies, resistant to G418, were identified, picked and grown to confluence in the absence of any selection as described in Chapter 2. These cell lines were trypsinised (as described in Chapter 2) upon reaching confluence (7-12 days post picking, Chapter 2). 4/10 volume, of each trypsinised cell line, were taken for DNA isolation (as described in Chapter 2) and 6/10 for storage under liquid nitrogen.

2.6.5 Identification of targeted cell lines

DNA was isolated from each cell line (as described in Chapter 2). 1/6 volume of each DNA preparation was incubated with the restriction enzyme *Eco* RI and the resulting fragments separated by electrophoresis in 0.5% agarose gels. These DNAs were then transferred to nylon membrane (as described in Chapter 2) for hybridisation with selected probes. It was believed that any one of a number of DNA fragments would be informative in the identification of targeted cell lines. The restriction sites (*Sal*I - *Eco*RI, corresponding to the 14-18 kb marks on the scale bar and spanning exon 8) permitting the

generation of these fragments may be observed in Figure 2.4.8. However, hybridisation experiments utilising the fragments *Hin*dIII -*Bam*HI(< 1.5 kb), *Bam*HI - *Sst* I (< 2 kb) and *Sst*I - *Eco*RI, as probes did not provide interpretable results. These probes hybridised in a non-specific manner and did not provide discrete signals. It was known from sequencing data (Contigs 1-3) that the genomic region encompassing this gene was rich in repeated sequence units (refer also to Section 2.4). Consequently a 1.3 kb genomic DNA fragment corresponding to a portion of exon 8b (defined in Table 2.3, Chapter 2) was utilised as a probe. The sequence of this DNA fragment is known not to contain repeated sequence units (Contig 3 nt 1216-2555, Appendix IX). Time limitation has precluded hybridisation experiments, utilising this DNA fragment as a probe, being performed. Figure 2.6.5 illustrates the homologous recombination event resulting in a correctly targeted allele of *Mobp*. The predicted size of DNA fragment detected by this probe is 16 kb. The diagnostic size shift in DNA from a targeted cell line, cleaved by restriction enzyme *Eco*RI, is predicted to be from 16 kb - 13 kb (wild type - targeted respectively). The appearance of these bands is illustrated schematically in Figure 2.6.6. It should be noted that both band sizes are expected in a targeted cell line. This is because normally only one allele will be targeted for an autosomal locus.

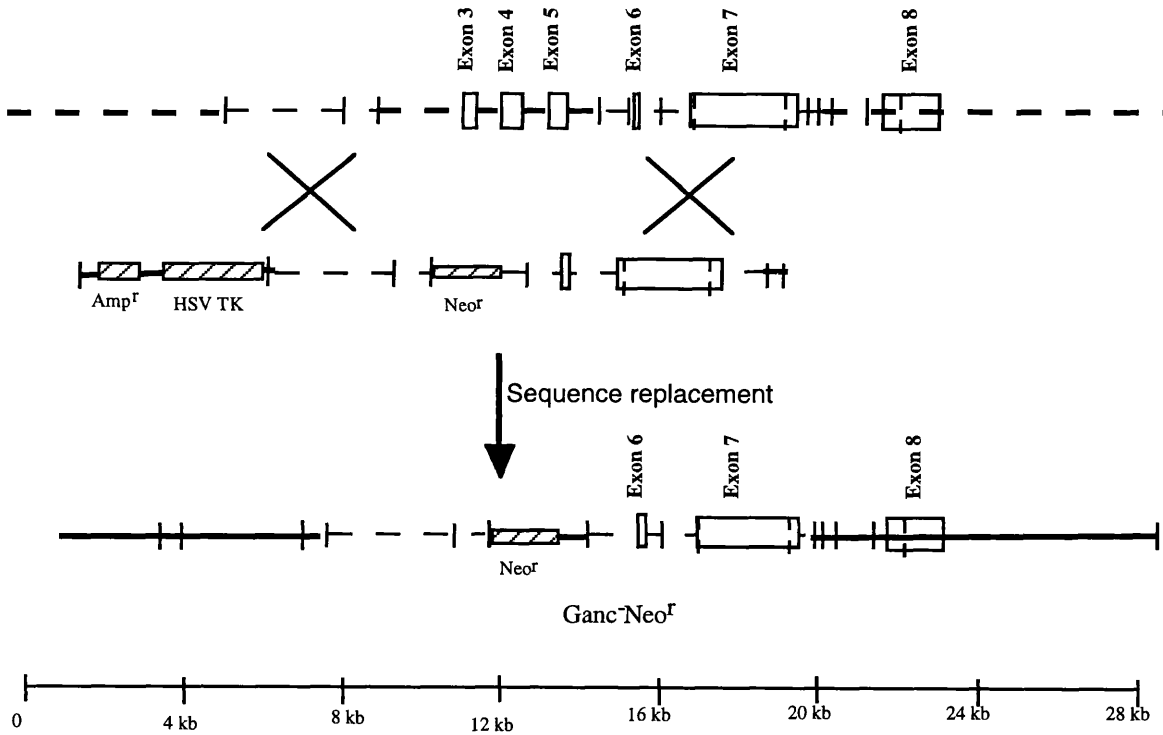


Figure 2.6.5 Disruption of the *Mobp* gene by gene targeting using the sequence replacement vector p*Mobp*-N/TK. This vector deletes exons 3, 4 and 5 of the *Mobp* gene. Figure 2.6.5 A corresponds to the wild-type *Mobp* locus. Figure 2.6.5 B corresponds to the targeting vector p*Mobp*-N/TK. The homologous recombination event, indicated by crossed lines, replaces the genomic sequence with construct sequence containing the neo^r gene. Figure 2.6.5 C corresponds to the correctly targeted (disrupted) *Mobp* locus. Fine and bold dotted lines (Figure 2.6.5 A) correspond to *Mobp* sequences that are homologous to sequences in p*Mobp*-N/TK and non-homologous to sequences in p*Mobp*-N/TK respectively. Dotted lines and solid black lines (Figure 2.6.2C) correspond to genomic and vector DNA respectively. The targeting recombination event causes the deletion of 5.5 kb of genomic DNA sequence and replaces this sequence with 2.5 kb of pNT derived vector DNA sequence.

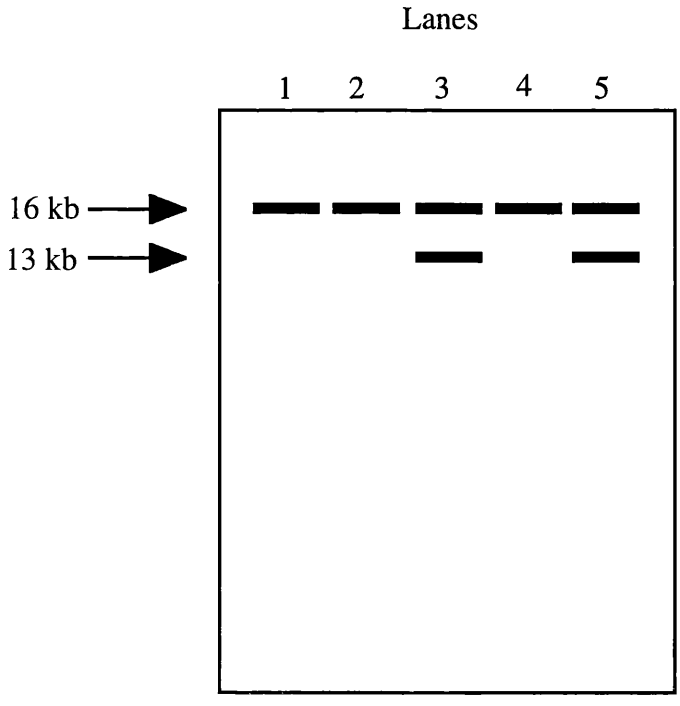


Figure 2.6.6 Schematic representation of predicted fragment size shift between DNA from wild-type ES cell lines and those carrying a correctly targeted *Mobp* allele The bands represent DNAs that have been cleaved by the restriction enzyme *Eco*RI (as described in 2.6.5) and hybridised with the probe DNA fragment (ASM 600). Fragments provided by DNA from wild-type ES cell lines are illustrated in lanes 1, 2 and 4. Fragments provided by DNA from ES cell lines carrying a correctly targeted *Mobp* allele are illustrated

2.6.6 Discussion

This study reports the generation of a gene disruption construct (p*Mobp*-N/TK) for the murine gene *Mobp*. p*Mobp*-N/TK was designed to delete the region encompassing exons 3-5 of the *Mobp* gene. Deletion of these exons will prohibit expression of all previously reported MOBP isoforms (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997) and the protein predicted to be produced by the novel transcript.

Though time constraint has prevented the completion of this study, it represents the first reported attempt to engineer a mutation in the *Mobp* gene. Time limitation has precluded completion of this work. However, evaluation of mice that are homozygous (-/-) for a null mutation in the *Mobp* gene will provide information regarding the basal function of the *Mobp* gene product (Deng *et al.*, 1993). It should be noted that some glia demonstrate a sensitivity to the dosage of certain myelin genes (Readhead *et al.*, 1994; Griffiths *et al.*, 1995; Zielasek *et al.*, 1996; refer also to Chapter 1, 1.7). Consequently, a study of mice heterozygous (*Mobp* $^{+/-}$) may also provide valuable information concerning the function of MOBP and its predicted role in stabilising the myelin sheath (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; Montague *et al.*, 1997a). The *Mobp* gene product is an interesting candidate gene for further study in relation to neuropathies affecting the CNS and specifically multiple sclerosis (refer also to Section 1, 2.2 and 2.7). By endeavouring to generate a loss of function mutation in the *Mobp* gene, this study provides a solid platform from which to begin assessing the function of MOBP. This represents the first steps in evaluating the possible role that *Mobp* may play in known neuropathies.

2.7 Discussion

2.7.1 Issues raised by experimental work

These studies (Section 2, 2.2-2.6) have raised a number of issues relating to the organisation and function of MOBP (refer also to Section 2, 2.3-2.6) and its putative association with known neuropathies in man and mouse (refer also to discussion in Section 2, 2.2). In an endeavour to address the issues raised by this work they are discussed below.

2.7.2 Multiple Sclerosis

Multiple Sclerosis (MS) is an autoimmune disorder affecting CNS myelin (Martyn, 1991; Allen and Brankin, 1993; Steinman, 1996). It is believed that MS develops when T-lymphocytes escape negative selection (deletion) in the thymus (Roder and Hickey, 1996). Autoimmunity would then develop because the inciting autoantigen was unavailable and thus remained undetected in the thymus. The autoantigen(s) responsible for inciting the autoimmune response in MS are, as yet, unknown (Martyn, 1991; Allen and Brankin, 1993; Steinman, 1996). Myelin proteins are acknowledged to be of particular interest in this context. Experimental allergenic encephalomyelitis (EAE) is a progressive demyelinating disease similar to MS (Miller and Rodriguez, 1996). It has been reported (Brocke *et al.*, 1996) that a single peptide analogue, comprising 13 amino acids (aa), of MBP (p87-99) is sufficient to induce EAE in mice. Unlike MBP, expression of MOBP is restricted to the CNS. In this context, the *Mobp* gene product is a strong candidate for further study in connection with MS.

2.7.3 Association with known neuropathies

2.7.3.1 Linkage of the *Mobp* gene locus to known neuropathies

The possible association of *Mobp* with any locus implicated in a known neuropathy, and in particular any locus associated with Multiple Sclerosis, must be among the most immediate issues to be addressed. The map position assigned to *Mobp* by this study (Section 2, 2.2) suggests that it is closely linked to a number of known neurological mutations in mouse. The published genetic mapping data available for these mutants was inadequate to permit exclusion of allelism with *Mobp*. I suggest that further work is required for the evaluation of the possibility that the *Mobp* gene product may be associated with any pathologies in the future. In order to evaluate the possible linkage of the *Mobp*

gene locus to any locus associated with MS the human *Mobp* gene locus must be identified. This may be readily performed using human × rodent radiation hybrid mapping panels. The human locus must then be evaluated for possible association with MS by comparison with loci previously implicated in the neuropathy. The evaluation of the possible association of *Mobp* with known murine mutants can be performed by genotyping a marker adjacent to a disease locus in question, using the same panel of DNAs as *Mobp*. This will permit direct comparison of the loci in question (refer also to discussion in Section 2, 2.2).

2.7.3.2 Association of the *Mobp* locus with a novel transgene insertion mutation demonstrating demyelination

A recent study (Orian *et al.*, 1994) reported the random insertion of a transgene, in distal mouse chromosome 9, resulting in a phenotype demonstrating dysmyelination. *Mobp* is a good candidate locus for this mutational event. Initial evaluation of the involvement of the *Mobp* gene in this pathology may be performed by assaying for the presence of the *Mobp* transcript in affected mice.

2.7.4 Complexity of the *Mobp* gene encoding myelin-associated oligodendrocytic basic protein

These studies demonstrate the degree of complexity involved in the generation of splice variants from the *Mobp* gene transcript (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; refer also to discussion in Section 2, 2.3). Generation of the previously identified splice variants of the *Mobp* transcript requires a complex range of alternative splicing events (refer also to Section 2; 2.3, 2.4 and 2.5). Elucidation of the genomic region encompassing all coding and 3' exons of the murine *Mobp* gene has provided greater insight into these splicing events. These studies demonstrate that the complex *Mobp* transcriptional unit, which comprises 8 exons, can also utilise a number of splice donor/acceptor sites within exons 7 and 8, in the generation of previously identified splice variants (refer also to Section 2; 2.3 and 2.4). This is the first report of the genomic organisation of the *Mobp* gene in any species. The selection of the mouse, as the species in which to examine this gene, has also permitted the initiation of a study to assess MOBP function via gene targeting (refer also to Section 2; 2.6).

2.7.4.1 Further study of the novel transcriptional unit within *Mobp*

To date it is not known whether the novel transcriptional unit, identified within the murine *Mobp* gene, produces a protein. Further study of this transcriptional unit requires confirmation of the production of a transcript e.g. by northern blot analysis, and subsequent assay for the presence of a protein corresponding to the predicted amino acid sequence. This will require antibodies to be raised against a synthetic peptide, designed to a unique portion of the predicted protein, and subsequent performance of a Western blot assay.

2.7.4.2 Functional analysis of the murine *Mobp* gene and the associated novel transcriptional unit

Assessment of the basal function of the *Mobp* gene product begins with the generation of mice homozygous for a null mutation (-/-) within the *Mobp* gene (refer also to Section 2, 2.6). Thus the primary aim of future study of *Mobp* gene function must be to complete the generation of $Mobp^{-/-}$ mice. Secondarily, the functions of the *Mobp* gene and the novel transcriptional unit must be dissected. In order to assess the basal function of the protein encoded by the novel transcriptional unit, sequence used exclusively in its production must be disrupted. Thus mice homozygous for a null mutation (-/-), specific to the novel transcriptional unit, may be generated by deleting the intron 3 of the *Mobp* gene. Sequence within this intron encodes amino acids 1-53 of the protein predicted to be encoded by the novel transcript. Elucidation of the genomic organisation of the *Mobp* gene and the complexity of the splicing events required to generate the splice variants encoding the family of related myelin-associated oligodendrocytic basic proteins (MOBP) has also provided a platform for the establishment of more sophisticated functional assays. It has been suggested that the isoform MOBP 170 may have a function alternative or additional to involvement in myelination (Montague *et al.*, 1997). This suggestion is based primarily on the detection of the Mobp 170 splice variant in the embryo (E12). Evaluation of the function of this isoform will require deletion of exon 4 of *Mobp*. To date exon 4 has only been demonstrated to be used in the production of MOBP 170 (refer also to Section 2, 2.4). The number of gene targeting strategies which may be envisaged for the *Mobp* gene is multiplied by the elucidation of its structure and complexity. Further studies may involve the deletion of exons containing sequences encoding amino acids used exclusively in the production of individual MOBP isoforms e.g. exon 6 is used exclusively in the production of MOBP 71 and exon 7a is used exclusively in the production of MOBP 81. Subsequently, animals lacking expression of specific MOBP

isoforms may be mated in an endeavour to create animals lacking expression of multiple isoforms. Another possible approach to studying the function of MOBP is the generation of mice carrying mutations in the *Mobp* gene which result in the production of an aberrant protein. However, such a study requires a greater understanding of the function of MOBP before it can be undertaken.

2.7.5 Future analysis of the *Mobp* promoter region

The organisation of the genomic region encompassing exons 1 and 2 of the *Mobp* gene is, as yet, incompletely understood. However, these studies do provide a solid basis for identification and analysis of the *Mobp* promoter region. Confirmation of the identity of the 1.2 kb PCR amplified genomic DNA fragment (refer to Section 2, 2.5) will permit its use as a probe to for the presence of sequences corresponding to exons 1 and 2 of the *Mobp* gene. It may be used to screen the bacteriophage clones that have been isolated, or the genomic libraries that were utilised, in the course of these studies.

2.7.6 Identification of mRNA localisation signals

It has been demonstrated that the different 3'UTRs of the splice variants Mobp 81A and Mobp 81B are generated as a consequence of the mutually exclusive use of exons 8 and 7, of the murine *Mobp* gene, respectively. This process of exon selection is reported to involve a fine balance between various processing events, including: transcription termination, cleavage and polyadenylation and splicing (Smith *et al.*, 1989). These splice variants have been suggested to localise to different parts of the oligodendrocyte (Holz *et al.*, 1996; Montague *et al.*, 1997) and to be present at different developmental stages (Montague *et al.*, 1997). It has further been suggested that the corresponding 3'UTRs of these splice variants contain as yet unidentified localisation signals (Holz *et al.*, 1996; Montague *et al.*, 1997). A number of recent reports (Smith *et al.*, 1989; Chabot, 1996) demonstrate a correlation between alternative splicing, molecular interaction e.g. protein localisation and biological function. Transgenic technologies may be used, *in vitro* and *in vivo* in the elucidation of localisation signals present in the corresponding exonic sequences (Haas *et al.*, 1994; Griffiths *et al.*, 1995; Ikenaka and Kagawa, 1995).

2.7.8 Concluding remarks

These studies represent a considerable contribution to the understanding of the map location, structure and complexity of the murine gene encoding myelin-associated

oligodendrocytic basic protein. They strengthen the suggestion that *Mobp* is a gene worthy of study in connection with neurodegenerative disease (Yamamoto *et al.*, 1994; Holz *et al.*, 1996; McCallion *et al.*, 1996; Montague *et al.*, 1997). Furthermore, these studies provide the basis for the construction of more complex and detailed hypotheses about the function of the *Mobp* gene product, permitting more incisive questions to be asked in the near future. Finally these studies have identified a novel transcriptional unit within the *Mobp* gene, increasing our understanding of the *Mobp* gene structure and adding to an already complex story.

Bibliography

Adams, M. D., Dubrick, M., Kerlavage, A. R., Moreno, R., Kelly, J. M., Utterback, T. R., Nagle, J. W., Fields, C., and Venter, J. C. (1992). Sequence identification of 2,375 human brain genes. *Nature* **355**, 632-634.

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropolous, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991). Complementary-DNA sequencing: Expressed sequence tags and the human genome project. *Science* **252**, 1651-1656.

Adlkofer, K., Martini, R., Aguzzi, A., Zielasek, J., Toyka, K. V., and Suter, U. (1995). Hypermyelination and demyelinating peripheral neuropathy in *Pmp*22 deficient mice. *Nature Genet.* **11**, 274-280.

Aguzzi, A., Brander, S., Marino, S., and Steinbach, J. P. (1996). Transgenic and knockout mice in the study of neurodegenerative disease. *J. Mol. Med.* **74**, 111-126.

Aguzzi, A., Brander, S., Sure, U., Ruedi, D., and Isenman, S. (1994). Transgenic and knock-out mice: models of neurological disease. *Brain Pathol.* **4**, 3-20.

Akowitz, A. A., Barbese, E., Scheld, K., and Carson, J. H. (1987). Structure and expression of myelin basic protein gene sequences in the *mld* mutant mouse: reitteration and rearrangement of the *Mbp* gene. *Genetics* **116**, 447-464.

Allen, A., and Brankin, B. (1993). Pathogenesis of multiple sclerosis: The immune diathesis and the role of viruses. *J. Neuropath. Exp. Neurol.* **52**, 95-105.

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.

Antonarakis, S. E., Boehm, C. D., Giardina, P. J., and Kazazian, H. H. (1982). Nonrandom association of polymorphic restriction sites in the beta-globin gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 137-141.

Avner, P., Amar, L., Dandolo, L., and Guénet, J. L. (1988). Genetic analysis of the mouse using interspecific crosses. *Trends Genet* . **4**, 18-23.

Barres, B. A., Hart, I. K., Coles, H. S., Burne, J. F., Voyvodic, J. T., Richardson, W. D., and Raff, M. C. (1992). Cell death and control of cell survival in the oligodendrocyte lineage. *Cell* **70**, 31-46.

Bartsch, U. (1996). Myelination and axonal regeneration in the central nervous system of mice deficient in the myelin-associated glycoprotein. *J. Neurocytol.* **25**, 303-313.

Beier, D. R. (1993). Single strand conformation polymorphism (SSCP) analysis as a tool for genetic mapping. *Mamm. Genome.* **4**, 627-631.

Beier, D. R., Her, H, Stevens, T.J., Sikela, J., Brady, K.P. Generation of a expressed sequence map of the mouse using single-strand conformation polymorphism (SSCP) mapping of cDNAs. *Amer. J. Hum. Genet.* **57**, 4, 1070

Beier, D. R., Dushkin, H., and Sussman, D. J. (1992). Mapping genes in the mouse using single stranded conformation polmorphism analysis of recombinant inbred strains and interspecific crosses. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9102-9106.

Botstein, D., White, R. L., Skolnick, M., and R.W., D. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.* **69**, 201-205.

Bradley, A. (1993). Gene targeting in embryonic stem cells. *Methods Enzymol.* **225**, 855-879.

Breen, M., Deakin, L., Macdonald, B., Miller, S., Sibson, R., Tarttelin, E., Avner, P., Bourgade, F., Guenet, J. L., Montagutelli, X., Poirier, C., Simon, D., Tailor, D., Bishop, M., Kelly, M., Rysavy, F., Rastan, S., Norris, D., Shepherd, D., Abbott, C., Pilz, A., Hodge, S., Jackson, I., Boyd, Y., Blair, H., Maslen, G., Todd, J. A., Reed, P. W., Stoye, J., Ashworth, A., McCarthy, L., Cox, R., Schalkwyk, L., Lehrach, H., Klose, J., Gangadharan, U., and Brown, S. (1994). Towards high-resolution maps of the mouse and human genomes : a facility for ordering markers to 0.1 cm resolution. *Hum. Mol. Genet.* **3**, 621-627.

Brocke, S., Koenraad, G., Allegratta, M., Ferber, I., Piercy, C., Blankenstein, T., Martin, R., Utz, U., Karin, N., Mitchell, D., Veromaa, T., Waisman, A., Gaur, A., Conlon, P., Ling, N., Fairchild, P. J., Wraith, D. C., O'Garra, A., Fathman, C. G., and Steinman, L. (1996). Treatment of experimental encephalomyelitis with a peptide analogue of myelin basic protein. *Nature* **379**, 343-346.

Brophy, P. J., Boccaccio, G. L., and Colman, D. R. (1993). The distribution of myelin basic protein mRNAs within myelinating oligodendrocytes. *Trends Neurosci.* **16**, 515-521.

Burke, D. T., Carle, G. F., and Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812.

Campagnoni, A. T. (1988). Molecular biology of proteins from the central nervous system. *J. Neurochem.* **51**, 1-14.

Capecchi, M. R. (1989). The new mouse genetics: Altering the mouse genome by gene targeting. *Trends Genet.* **5**, 70-76.

Chabot, B. (1996). Directing alternative splicing: cast and scenarios. *Trends Genet.* **12**, 472-478.

Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M., and Gruss, P. (1991) The molecular basis of the undulated *Pax-1* mutation. *Cell*, **66**, 5, 873-884.

Colman, D. R., Kreibach, G., Frey, A. B., and Sabatini, D. D. (1982). Synthesis and incorporation of myelin polypeptides into CNS myelin. *J. Cell. Biol.* **95**, 598-608.

Conner, B. J., Reyes, A. A., Morin, C., Itakura, K., Teplitz, R. L., and Wallace, R. B. (1983). Detection of sickle cell beta-globin allele by hybridisation with synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 278-282.

Copeland, N. G., and Jenkins, N. A. (1991). Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* **7**, 113-118.

Cotton, R. G. S., Rodrigues, N. R., and Campbell, R. D. (1988). Reactivty of cytosine and Thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4397-4401.

Coulson, A, Sulston, J., Brenner, S., and Karn, J. (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans. Proc. Natl. Acad. Sci. U.S.A.* **83**, 7821-7825.

Craft, C. M., Whitmore, D. H., and Wiechmann, A. F. (1994). Cone arrestin identified by targeting expression of a functional family. *J. Biol. Chem.* **269**, 4613-4619.

D'Allessio, J. M., Gruber, C. E., Cain, C., and Noon, M. C. (1990). Construction of directional libraries using the Superscript plasmid system. *Focus* **12**, 47-50.

Davies, R. W., Roberts, A. B., Morris, A. J., Griffith, G. W., Jerecic, J., Ghandi, S., Kaiser, K., and Savioz, A. (1994). Enhanced access to rare brain cDNAs by prescreening libraries: 207 new mouse brain ESTs. *Genomics* **24**, 456-463.

Deng, C., and Capecchi, M. R. (1992). Reexamination of targeting frequency as a function of the extent of homology between the targeting vector and target locus. *Mol. Cell. Biol.* **12**, 3365-3371.

Deng, C., Thomas, K. R., and Capecchi, M. R. (1993). Location of crossovers during gene targeting with insertion and replacement. *Mol. Cell. Biol.* **13**, 2134-2140.

Duncan, I. D., Hammang, J. P., Goda, S., and Quarles, R. H. (1989). Myelination in the *jimpy* mouse in the absence of proteolipid protein. *Glia* **2**, 148-154.

Duncan, I. D., Hammang, J. P., and Trapp, B. D. (1987). Abnormal compact myelin in the myelin deficient rat: Absence of proteolipid protein correlates with a deficit in the intraperiod line. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6287-6291.

el Aoumari, A., Fromaget, C., Dupont, E., Reggio, H., Durbec, P., Briand, J. P., Boller, K., Kreitman, B., and Gros, D. (1990). Conservation of a cytoplasmic carboxy-terminal domain of connexin-43, a gap junctional protein, in the mammal heart and brain. *J. Membr. Biol.* **115**, 229-240.

Feinberg, A.P., and Vogelstein, B. (1983) A technique for radiolabeling restriction fragments to high specific activity. *Anal. Biochem.* **301**, 5895, 6-13

Fischer, S. C., and Lerman, L. S. (1980). Separation of random fragments of DNA according to properties of their sequence. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4420-4424.

Folger, K. R., Wong, E. A., Wahl, G., and Cappechi, M. R. (1982). Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol. Cell. Biol.* **2**, 1372-1387.

Fremeau, R. T., and Popko, B. (1990). *In situ* analysis of myelin basic protein gene expression in myelin deficient oligodendrocytes: anitisense hnRNA and read through transcription. *E.M.B.O. Journal* **9**, 3533-3538.

Furuichi, Y., and Miura, K. (1975). A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* **253**, 374-375.

Gardiner, K., and Brennan, M. (1992). First international workshop on the identification of transcribed sequences. *Human Genome News* **3**, 7-9.

Gencic, S., and Hudson, L. D. (1990). Conservative amino acid substitution in the myelin proteolipid protein of *jimpy-msd* mice. *J. Neurosci.* **10**, 117-124.

Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K. A., Antonio, M., Beisel, K. W., Steel, K. P., and Brown, S. D. M. (1995). A type-VII myosin encoded by the mouse deafness gene shaker-1. *Nature* **374**, 62-64.

Giese, K. P., Martini, R., Lemke, G., Soriano, P., and Schachner, M. (1992). Disruption of the *P0* gene in mice leads to normal expression of recognition, and degeneration of myelin and axons. *Cell* **71**, 565-576.

Glavac, D., and Dean, M. (1993). Optimisation of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. *Human Mutation* **2**, 404-414.

Griffiths, I. R., Schneider, A., Anderson, J., and Nave, K. A. (1995). Transgenic and natural mouse models of proteolipid protein (*Plp*) related dysmyelination and demyelination. *Brain Path.* **5**, 275-281.

Guénet, J. L. (1996). The Mouse Genome. Genomes, molecular biology and drug discovery. (New York: Academic Press).

Haas, S., Haque, N. S., Beggs, A. H., Khalili, K., Knobler, R. L., and Small, J. (1994). Expression of the myelin basic-protein gene in transgenic mice expressing human neurotropic virus, JCV, early protein. *Virology* **202**, 89-96.

Hasty, P. J., Rivera-Pérez, J., and Bradley, A. (1991). The length of homology required for gene targeting in embryonic stem cells. *Mol. Cell. Biol.* **11**, 5586-5591.

Hawkins, J. D. (1988). A survey on intron and exon lengths. *Nucleic Acids Res.* **16**, 9893-9908.

Hayasaka, K., Himoro, M., Sato, W., Takada, G., Uyemura, K., Shimuzu, N., Bird, T. D., Conneally, P. M., and Chance, P. F. (1993). Charcot-Marie-Tooth neuropathy type 1B is associated with mutations of myelin *P0* gene. *Nature. Genet.* **5**, 31-34.

Hayashi, K., and Yandell, D. W. (1993). How sensitive is PCR-SSCP? *Human Mutation* **2**, 338-346.

Higgins, L. S., and Cordell, B. (1995). Transgenic mice and modelling Alzheimer's disease. *Rev. Neurosci.* **6**, 87-96.

Hochgeschwander, U. (1992). Towards a transcriptional map of the human genome. *Trends Genet.* **8**, 41-44.

Holz, A., Schaerenwiemers, N., Schaefer, C., Pott, U., Colello, R. J., and Schwab, M. E. (1996). Molecular and developmental characterization of novel cDNAs of the myelin-associated oligodendrocytic basic-protein. *J. Neurosci.* **16**, 467-477.

Hoog, C. (1991). Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucleic Acids Res.* **19**, 6123-6127.

Hudson, L. D., Berndt, J. A., Puckett, C., Kozak, C. A., and Lazzarini, R. A. (1987). Aberrant splicing of proteolipid protein mRNA in the dysmyelination *jimpy* mutant mouse. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1454-1458.

Ikenaka, K., and Kagawa, T. (1995). Transgenic systems in studying myelin gene-expression. *Dev. Neurosci.* **17**, 127-136.

Jacobson, M. (1991). Neuroglial ontogeny. in developmental neurobiology, M. Jacobson, ed. (New York: Plenum Press), pp. 95-142.

Jiang, W., Kahn, S. M., Guillem, J. G., Lu, S.-H., and Weinstein, I. B. (1989). Rapid detection of ras oncogenes in tumors: applications to colon, esophageal and gastric cancer. *Oncogene* **4**, 923-928.

Kandel, E.R., Schwartz, J.H., and Jessel, T.M. (1991) Principles of neural science. Third edition (London, Prentice Hall), pp44.

Kapfhamer, D., Sweet, H. O., Sufalko, D., Warren, S., Johnson, K. R., and Burmeister, M. (1996). The neurological mouse mutations *jittery* and *hesitant* are allelic and map to the region of mouse chromosome 10 homologous to 19p13.3. *Genomics* **35**, 533-538.

Kimura, M., Sato, M., Akatsuka, A., Nozawa-Kimura, S., Takahashi, R., Yokoyama, M., Nomura, T., and Katsuki, M. (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.

Kirstensson, K., Zeller, N. K., Dubois-Daleq, M., and Lazzarini, R. A. (1986). Expression of myelin basic protein in the developing rat brain as revealed by *in situ* hybridisation. *J Histochem. Cytochem.* **34**, 467-473.

Ko, M. S. H., Wang, X., Horton, J. H., Hagen, M. D., Takashi, N., Maeziki, Y., and Nadeau, J. H. (1994). Genetic mapping of 40 cDNA clones on the mouse genome by PCR. *Mamm. Genome* **5**, 349-355.

Kupryjanczyk, J., Beauchamp, R., Thor, A. D., Merrit, V., Edgerton, S. M., Bell, D. A., and Yandell, D. W. (1993). p53 gene alterations and protein accumulation in human ovarian cancer. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4961-4965.

Kurihara, T., Kataoka, K., Hong, D., Shioda, S., Sugano, S., Mitamura, K., Maruyama, K., and Yamaguchi, N. (1997). Genomic structure and promoter analysis of the gene encoding MM3, a member of transmembrane 4 superfamily. *Gene* **185**, 277-283.

Lachapelle, F., Lapie, P., and Gumpel (1992). Oligodendrocytes from *jimpy* and normal mature tissue can be "activated" when transplanted in a newborn environment. *Dev. Neurosci.* **14**, 105-113.

Langnaese, K., Beesley, P. W., and Gundelfinger, E. D. (1997). Synaptic membrane glycoproteins gp65 and gp55 are new members of the immunoglobulin superfamily. *J. Biol. Chem.* **272**, 821-827.

Lees, M. B., and Brostoff, S. W. (1984). Proteins of myelin. Myelin, P. Morell, ed. (New York: Plenum Press), pp. 197-224.

Lees, M. K., Borchelt, D. R., Wong, P. C., Sisodia, S. S., and Price, D. L. (1996). Transgenic models of neurodegenrative disease. *Curr. Opin. Neurobiol.* **6**, 651-660.

Lemke, G. (1992). Myelin and myelination. An introduction to molecular neurobiology, Z. Hall, ed. (Sunderland, MA: Sinauer), pp. 281-309.

Lemke, G. (1988). Unwrapping the genes of myelin. *Neuron* **1**, 525-543.

Lerman, L. S., and Silverman, K. (1987). Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Methods Enzymol.* **155**, 482-501.

Letsou, A., and Liskay, R. M. (1987). Effect of the molecular nature of mutation on the efficiency of intrachromosomal gene conversion in mouse cells. *Genetics* **117**, 759-769.

Levitt, R. C. (1991). Minireview. Polymorphisms in the transcribed 3' untranslated region of eukaryotic genes. *Genomics* **11**, 484-489.

Levitt, R. C. (1991). Polymorphisms in the transcribed 3'untranslated region of eukaryotic genes. *Genomics* **11**, 484-489.

Mansour, S. L., Thomas, K. R., and Capecchi, M. R. (1988). Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general selection strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348-352.

Martini, R., Mohajeri, M. H., Kasper, S., Giese, K. P., and Schachner, M. (1995). Mice doubly deficient in the genes for P0 and myelin basic-protein show that both proteins contribute to the formation of the major dense line in peripheral-nerve myelin. *J. Neurosci.* **15**, 4488-4495.

Martini, R., Zielasek, J., Toyka, K. V., Giese, K. P., and Schachner, M. (1995). Protein zero (P0)-deficient mice show myelin degeneration in peripheral-nerves characteristic of inherited human neuropathies. *Nature Genet.* **11**, 281-286.

Martyn, C. (1991). Epidemeology. McAlpine's Multiple Sclerosis, W. B. Matthews, ed. (Edinburgh: Churchill Livingstone).

Maruyama, K., and Sugano, S. (1994). Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138**, 171-174.

McCallion, A. S., Guénet, J. L., Montague, P., Griffiths, I. R., Savioz, A., and Davies, R. W. (1996). The mouse gene (*Mobp*) encoding myelin-associated oligodendrocytic basic-protein maps to distal chromosome-9. *Mamm. Genome* **7**, 847-849.

McCormick, M. K., Shero, J. H., Cheung, M. C., Kan, Y. W., Hieter, P. A., and Antonarakis, S. E. (1989). Construction of human chromosome 21-specific yeast artificial chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9991-9995.

Michaud, J., Brody, L. C., Steel, G., Fontaine, G., Martin, L. S., Valle, D., and Mitchell, G. (1991). Strand-separationg conformational polymorphism analysis: efficacy of detection of point mutations in the human ornithine delta-aminotransferase gene. *Genomics* **13**, 389-394.

Mikoshiba, K., Okano, H., Tamura, T. A., and Ikenaka, K. (1991). Structure and function of myelin protein genes. *Ann. Rev. Neurosci.* **14**, 201-217.

Miller, C. C. J., McPheat, J.C. and Potts, W.J. (1992). Targeted integration of the *Ren-1D* locus in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5020-5024.

Miller, D. J., and Rodriguez, M. (1996). Spontaneous central nervous system remyelination in strain A mice after infection with the Daniel's (DA) strain of Theiller's virus. *Acta. Neuropathol.* **91**, 559-565.

Montague, P., Dickinson, P. J., McCallion, A. S., Stewart, G. J., Savioz, A., Davies, R. W., Kennedy, P. G. E., and Griffiths, I. R. (1997). Developmental expression of the murine *Mobp* gene. *J. Neurosci. Res.* **49**, 133-143.

Montague, P., Kirkham, D., McCallion, A. S., Davies, R. W., Kennedy, P. G. E., Klugmann, M., Groome, N. P., and Griffiths, I. R. (1997). Expression of the *Mobp* gene in the *shiverer* mouse. *Submitted.*

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459-472.

Murray, V. (1989). Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17**, 8889.

Myers, R. M., Maniatis, T., and Lerman, L.S. (1987). Detection and localisation of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* **155**, 501-527.

Myers, R. M., Lumelsky, N., Lerman, L., and Maniatis, T. (1985). Detection of single base substitutions in total genomic DNA. *Nature* **313**, 495-498.

Nakajima, K., Ikenaka, K., Kagawa, T., Aruga, J., Nakao, J., Nakahira, K., Shiota, C., Kim, S. U., and Mikoshiba, K. (1993). Novel isoforms of mouse myelin basic protein predominantly expressed in embryonic stage. *J. Neurochem.* **60**, 1554-1563.

Nave, K.A., Lai, C., Bloom, F. E., and Milner, R. J. (1986). *Jimpy* mutant mouse: A 74 bp deletion in the mRNA for myelin proteolipid protein and evidence for a primary defect in RNA splicing. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9264-9268.

Nave, K. A. (1994). Neurological mouse mutants and the genes of myelin. *J. Neurosci. Res.* **38**, 607-612.

Nehls, M., Messerle, M., Sirulnik, A., Smith, A. J. H., and Boehm, T. (1994). 2 Large insert vectors, Lambda-PS and Lambda-KO, facilitate rapid mapping and targeted disruption of mammalian genes. *Biotechniques* **17**, 770-775.

Newman, S., Kitamura, K., and Campagnoni, A. T. (1987). Identification of a cDNA coding for a fifth form of myelin basic protein in mouse. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 886-890.

Noda, Y., Sato-Yoshitake, R., Kondo, S., Nangaku, M., and Hirokawa, N. (1995). KIF2 is a new microtubule-based anterograde motor that transports membranous organelles distinct from those carried by kinesin heavy chain or KIF3A/B. *J. Cell. Biol.* **129**, 157-167.

Noll, W. W., and Collins, M. (1987). Detection of human DNA polymorphisms with a simplified denaturing gradient gel electrophoresis technique. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3339-3343.

Okamato, A., Sameshima, Y., Yokoyama, S., Terashima, Y., Sugimura, T., Terada, M., and Yokota, J. (1991). Frequent allelic losses and mutations of the p53 gene in human ovarian cancer. *Cancer Res.* **51**, 5171-5176.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., and Matsubara, K. (1992). Large scale cDNA sequencing analysis of quantitive and qualitative aspects of gene expression. *Nature Genet.* **2**, 173-179.

Orian, J. M., Mitchell, A. W. S., Marshman, W. E., Webb, G. C., Ayers, M. M., Grail, D., Ford, J. H., Kaye, A. H., and Gonzales, M. F. (1994). Insertional mutagenesis inducing hypomyelination in transgenic mice. *J. Neurosci. Res.* **39**, 604-612.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2766-2770.

Orita, M., Suzuki, Y., Kanazawa, H., Sekiya, T., and Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874-879.

Pham-Dinh, D., Mattei, M. G., Nussbaum, J. L., Roussel, G., Pontarotti, P., Roeckel, N., Mather, I. H., Artzt, K., Lindahl, K. F., and Dautigny, A. (1993). Myelin/oligodendrocyte glycoprotein is a member of a subset of the immunoglobulin superfamily encoded within the major histocompatibility complex. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7990-7994.

Podsulo, S. E., Dean, M., Kolch, U., and O'Brien, S. J. (1991). Detecting high-resolution polymorphisms in human coding loci by combining PCR and single-strand conformation polymorphism (SSCP) analysis. *Am. J. Hum. Genet.* **49**, 106-111.

Polymeropoulos, M. H., Xiao, H., Glodek, A., Gorski, M., Adams, M. D., Moreno, R. F., Fitzgerald, M. G., Venter, J. C., and Merril, C. R. (1992). Chromosomal assignment of 46 brain cDNAs. *Genomics* **12**, 492-496.

Polymeropoulos, M. H., Xiao, H., Sikela, J. M., Adams, M., Venter, J. C., and Merril, C. R. (1993). Chromosomal distribution of 320 genes from a brain cDNA library. *Nature Genet.* **4**, 381-386.

Price, D. L., Becher, M. W., Wong, P. C., Borchelt, D. R., Lee, M. K., and Sisodia, S. S. (1996). Inherited neurodegenerative diseases and transgenic models. *Brain Pathol.* **6**, 467-480.

Privat, A. (1980). Morphological aspects of myelination. Neurological muations affecting myelination, N. Baumann, ed.: Elsevier), pp. 71-86.

Readhead, C., Schneider, A., Griffiths, I. R., and Nave, K-A. (1994). Premature arrest of myelination in transgenic mice with increased proteolipid protein gene dosage. *Neuron* **12**, 583-595.

Roach, A., Takahashi, N., Pravtcheva, D., Ruddlle, F., and Hood, L. (1985). Chromosomal mapping of mouse myelin basic protein and structure and transcription of the partially deleted gene in *shiverer*. *Cell* **42**, 149-155.

Roder, J., and Hickey, W. F. (1996). Mouse models, immunology, multiple sclerosis and myelination. *Nature Genet.* **12**, 6-8.

Rogers, S. D., Demaster, E., Catton, M., Ghilardi, J. R., Levin, L. A., Maggio, J. E., and Mantyh, P. W. (1997). Expression of endothelin-B receptors by glia in vivo is increased after CNS injury in rats, rabbits, and humans. *Exp. Neurol.* **145**, 180-195.

Rosenbluth, J. (1980). Central myelin in the mouse mutant *shiverer*. *J. Comp. Neurol.* **194**, 639-648.

Rothstein, R. (1991). Targeting, disruption, replacement and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.* **194**, 281-301.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Salzer, J. L., Homes, W. P., and Colman, D. R. (1987). The amino acid sequences of the myelin-associated glycoproteins: Homology to the immunoglobulin super-family. *J. Cell. Biol.* **104**, 957-965.

Savioz, A., and Davies, R. W. (1995). Discovering genes with localised expression in the mouse brain: cDNAs specific to the substantia nigra. *Gene* **154**, 225-230.

Scherer, J. B., and Adair, G. M. (1994). Homology dependence of targeted recombination at the chinese hamster *APRT* locus. *Mol. Cell. Biol.* **14**, 6663-6673.

Scherer, S. S., and Chance, P. F. (1995). Myelin genes: getting the dosage right. *Nature Genet.* **11**, 226-228.

Schneider, A., Montague, P., Griffiths, I. R., Fanarraga, M., Kennedy, P., P., B., and Nave, K-A. (1992). Uncoupling of hypomyelination and glial cell death by a mutation in the proteolipid protein gene. *Nature* **358**, 758-761.

Sheffield, V. C., Cox, R., Lerman, L. R., and Meyers, R. M. (1989). Attachment of a 40 base-pair G+C rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single base changes. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 232-236.

Shirozu, M., Tada, H., Tashiro, K., Nakamura, T., Lopez, N. D., Nazarea, M., Hamada, T., Sato, T., Nakano, T., and Honjo, T. (1996). Characterisation of the novel secreted membrane proteins isolated by the signal sequence trap method. *Genomics* **37**, 273-280.

Silver, L. M. (1995). Mouse genetics: Concepts and applications (Oxford: Oxford University Press).

Smith, C. W. J., Patton, J. G., and Nadal-Ginard, B. (1989). Alternative splicing in the control of gene expression. *Ann. Rev. Genet.* **23**, 527-577.

Spinardi, K., Mazards, R., and Theillet, C. (1991). Protocols for improved detection of point mutations by SSCP. *Nucleic Acids Res.* **19**, 4009.

Staugaitis, S. M., Colman, D. R., and Pedraza, L. (1996). Membrane adhesion and other functions for the myelin basic proteins. *Bioessays* **18**, 13-18.

Stegmann, A. P. (1993). Assignment of the human deoxycytidine kinase (DCK) gene to chromosome 4 band q13.3-q21.1. *Genomics.* **17**, 2, 528-529.

Steinman, L. (1996). Multiple sclerosis: A coordinated immunological attack against myelin in the central nervous system. *Cell* **85**, 299-302.

Stewart, a. (1995). Human gene map workshops held. *Human Genome Digest* **2**, 1-4, 6-9.

Stewart, G. J., Savioz, A., and Davies, R. W. (1996). Sequence analysis of 497 mouse brain ESTs expressed in the substantia nigra. *Genomics* **39**, 147-153.

Strachan, T., Abitbol, M., Davidson, D., and Beckmann, J. S. (1997). A new dimension for the human genome project: towards comprehensive expression maps. *Nature Genet.* **16**, 126-132.

Suter, U., Welcher, A. A., and Snipes, G. J. (1993). Progress in the molecular understanding of hereditary peripheral neuropathies reveals new insights into the biology of the peripheral nervous sytem. *Trends. Neurosci.* **16**, 50-56.

Suzuki, Y., Orita, M., Shiraishi, M., Hayashi, K., and Sekiya, T. (1990). Detection of Ras gene-mutations in human lung cancers by single strand conformation polymorphism analysis of polymerase chain reaction products. *Oncogene* **5**, 1037-1043.

Takahashi, N., and Ko, M. S. H. (1993). The short 3'-end region of complementary DNAs as PCR-based polymorphic markers for an expression map of the mouse genome. *Genomics* **16**, 161-168.

te Reile, H., Maandag, E. R., and Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5128-5132.

Thomas, K. R., and Capecchi, M. R. (1987). Site directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-512.

van Deursen, J., and Wieringa, B. (1992). Targeting of the creatine Kinase M gene in embryonic stem cells using isogenic and non-isogenic vectors. *Nucleic Acids Res.* **20**, 3815-3820.

Vermeesch, M. K., Knapp, P. E., Skoff, R. P., Studzinski, D. M., and Benjamins, J. A. (1990). Death of individual oligodendrocytes in *jimpy* brain preceeds expression of proteolipid protein. *Dev. Neurosci.* **12**, 303-315.

Waldman, A. S., and Liskay, R. M. (1987). Dependence of intrachromosomal recombination in mammalian cells on uninterupted homology. *Mol. Cell. Biol.* **8**, 5350-5357.

Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., and Itakura, K. (1981). The use of synthetic oligonucleotides as hybridisation probes: Hybridisation of oligonucleotides of mixed sequence to rabbit B-globin DNA. *Nucleic Acids Res.* **9**, 879-894.

Warren, A. J., Colledge, W. H., Carlton, M. B. L., Evans, M. J., Smith, A. J. H., and Rabbitts, T. H. (1994). The oncogenic cysteine-rich LIM domain protein Rbtn2 is essential for erythroid development. *Cell* **78**, 45-57.

Weil, D., Blanchard, S., Kaplan, J., Guilford, P., Gibson, F., Walsh, J., Mburu, P., Varela, A., Levillers, J., Weston, M. D., Kelley, P. M., Kimberling, W. J., Wagenaar, M., Levi-Acobas, F., Larget-Piet, D., Munnich, A., Steel, K. P., Brown, S. D. M., and Petit, C. (1995). Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature* **374**, 60-61.

White, M. B., Carvalho, M., Derse, D., Obrien, S. J., and Dean, M. (1992). Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12**, 301-306.

Wilcox, A. S., Khan, A. S., Hopkins, J. A., and Sikela, J. M. (1991). Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: Implications for an expression map of the genome. *Nucleic Acids Res.* **19**, 1837-1843.

Winter, E., Yamamoto, F., Almoguera, C., and Perucho, M. (1985). A method to detect annd characterise point mutations in transcribed genes: amplification and overexpression of the mutant c-K-ras allele in human tumor cells. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7575-7579.

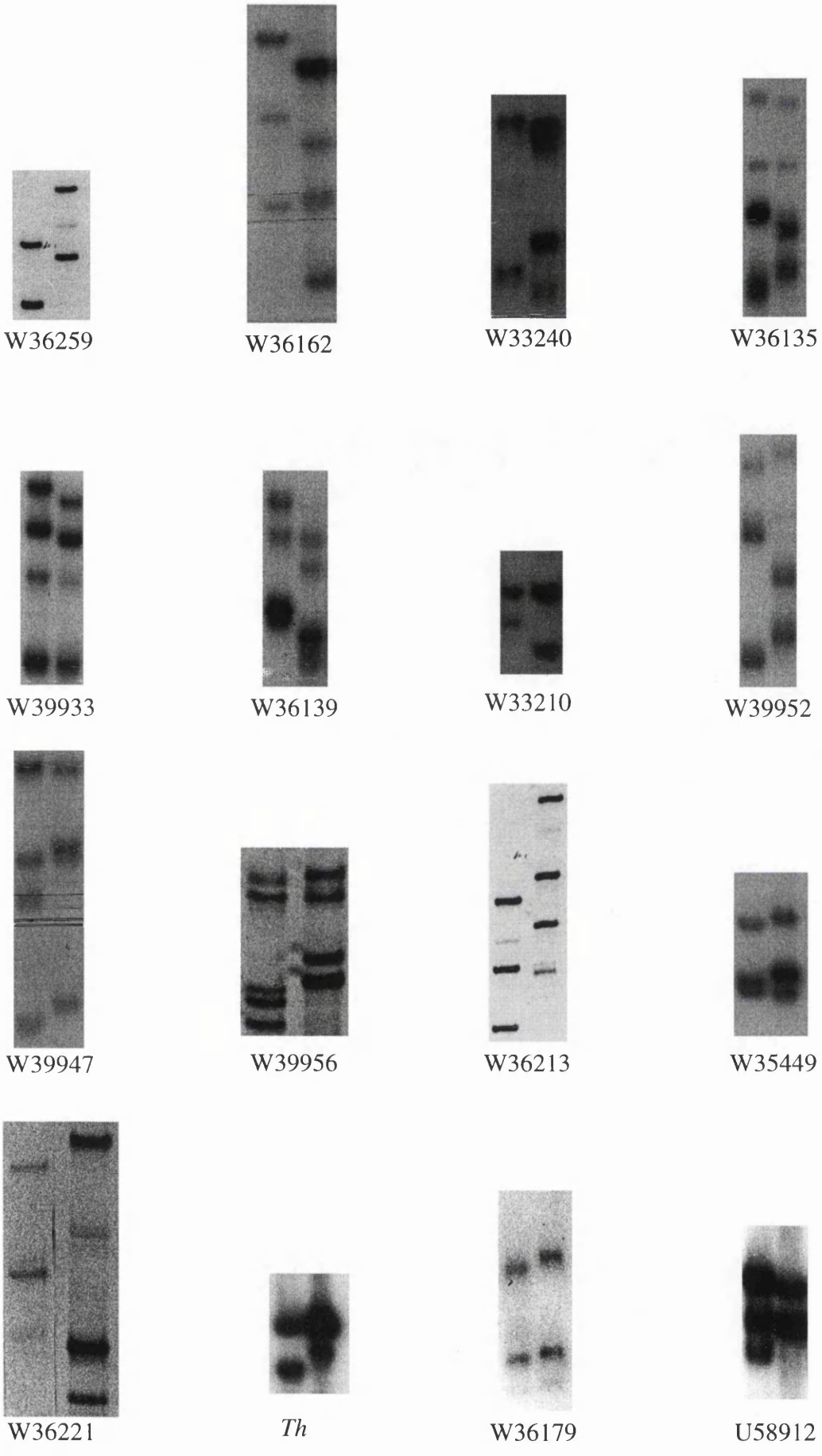
Yamamoto, Y., Mizuno, R., Nishimura, T., Ogawa, Y., Yoshikawa, H., Fujimura, H., Adachi, E., Kishimoto, T., Yanagihara, T., and Sakoda, S. (1994). Cloning and expression of myelin-associated oligodendrocytic basic protein. A novel basic protein constituting the central nervous system myelin. *J. Biol. Chem.* **269**, 31725-31730.

Yang, X. Y., and Skoff, R. P. (1997). Proteolipid protein regulates the survival and differentiation of oligodendrocytes. *J. Neurosci.* **17**, 2056-2070.

Zeng, Z., Su, K., Kyaw, H., and Li, Y. (1997). A novel endothelin receptor type-B-like gene enriched in the brain. *Biochem. Biophys. Res. Commun.* **233**, 559-567.

Zielasek, J., Martini, R., and Toyka, K. V. (1996). Functional abnormalities in P0 deficient mice resemble human hereditary neuropathies linked to P0 gene mutations. *Muscle Nerve* **19**, 946-952.

Appendices



Appendix I Examples of SSCPs demonstrated, between *M.spretus* and C57BL/6, by PCR amplified sequences from within ESTs utilised in these studies.

Chromos	Primers		101	391	421	2	5	8	10	12
Chr.1	D1Nds4	7cM	B	S	S	B	B	B	B	B
Chr.1	D1Mit24	42,3cM	B	S	B	B	B	B	B	B
Chr.1	D1Mit17	111cM	B	B	B	B	B	B	B	B
Chr.2	D2Mit1	1cM	B	B	B	B	B	B	B	B
Chr.2	D2Mit11	41cM	B	B	S	B	S	B	B	B
Chr.2	D2Mit55	67cM	S	B	S	S	S	B	B	B
Chr.2	D2Mit52	77cM	S	B	B	S	S	B	B	B
Chr.3	D3Mit54	0cM	B	B	S	B	B	S	S	S
Chr.3	D3Mit22	35cM	B	S	S	B	B	S	S	S
Chr.3	D3Mit9	38.6cM	B	S	S	B	B	S	S	S
Chr.3	D3Mit45	75.3cM	B	B	B	B	B	S	S	S
Chr.3	D3Mit19	81cM	B	B	B	B	B	S	S	S
Chr.4	D4Mit19	8cM	S	S	B	B	B	B	B	B
Chr.4	D4Mit38	18.8cM	S	S	B	B	B	B	B	B
Chr.4	ORM 1	38cM	S	S	B	B	B	B	B	B
Chr.4	D4Mit37	47.4cM	S	B	B	B	B	B	B	B
Chr.4	D4Mit33	67cM	S	S	B	B	B	B	B	B
Chr. 5	D5Mit48	0cM	B	S	S	S	B	B	S	B
Chr. 5	D5Nds2	28cM	B	B	B	B	B	B	S	S
Chr. 5	Gus	63cM	B	B	B	B	B	B	B	B
Chr.6	D6Mit47	11cM	S	S	B	B	B	B	B	B
Chr.6	D6Mit8	31cm	S	S	B	B	B	B	B	B
Chr.6	D6Nds5	41cM	S	S	B	B	B	B	B	B
Chr.6	D6Mit201	67cM	S	S	B	S	S	B	B	B
Chr.7	D7Mit21	3cM	B	B	B	B	B	B	S	B
Chr.7	D7Mit54	15cM	B	B	B	B	B	B	S	B
Chr.7	D7Mit40	51cM	B	B	B	B	B	B	B	B
Chr.7	Fgf-3	74cM	B	B	B	B	B	B	B	B
Chr.8	D8Mit141	4,6cM	S	B	B	B	B	B	B	B
Chr.8	D8Mit40	35cM	B	B	B	B	B	B	B	B
Chr.8	D8Mit35	52cM	B	B	B	B	B	B	B	B
Chr.8	D8Mit13	68,4cM	B	B	B	B	B	B	B	B
Chr.9	D9Mit160	2,2cM	B	B	B	B	B	B	B	B
Chr.9	D9Mit2	13,2cM	B	B	B	B	B	B	B	B
Chr.9	33(Cyp1a2)	31cM	B	B	B	B	B	B	B	B
Chr.9	D9Mit137	64,2cM	B	B	B	B	B	B	B	B
Chr.10	D10Mit77	2,2cM	S	B	B	B	B	S	B	S
Chr.10	D10Mit16	10,2cM	S	S	B	B	B	S	B	B
Chr.10	D10Mit15	23cM	B	S	B	B	S	S	B	B
Chr.10	D10Mit22	38,7cM	S	B	B	B	S	B	B	B
Chr.10	D10Mit67	46,4cM		B	B		S		B	B
Chr.10	D10Mit180	64cM	S	B	S	B	B	B	B	B

Genotyping data corresponding to the anchor markers used in the typing of DNAs from the N4 generation of the breeding programme to generate consomic mice.

Chromos	Primers		101	391	421	2	5	8	10	12
Chr.11	D11Mit1	2,2cM	B	B	B	B	B	B	B	B
Chr.11	D11Mit31	37cM	B	B	S	S	S	B	S	S
Chr.11	D11Mit159	43,5cM	B	B	S	S	S	S	S	S
Chr.11	D11Mit48	82cM	B	B	S	B	B	S	S	S
Chr.12	MODC	1cM	B	S	B	S	B	B	B	B
Chr.12	D12Mit65	22cM	B	S	B	B	B	B	B	B
Chr.12	D12Mit4	33cM	B	S	B	B	B	B	B	B
Chr.12	D12Mit27	50cM	B	S	B	B	B	B	B	B
Chr.12	1(Igh-V)	62cM	B	S	B	B	B	B	B	B
Chr.13	D13Mit80	6.4	S	S	B	B	S	B	B	B
Chr.13	D13 Mit 3	11cM	S	S	B	B	S	B	B	B
Chr.13	D13 Mit 13	30cM	S	S	B	B	S	B	B	B
Chr.13	D13 Mit 28	46cM	S	B	B	S	S	B	B	B
Chr.13	D13 Mit 47	62cM	S	B	B	S	B	B	B	B
Chr.14	D14 Mit1	2 cM	S	S	B	B	B	B	B	B
Chr.14	D14 Mit5	15 cM	B	B	B	B	B	B	B	B
Chr.14	120	26 cM	B	B	B	B	B	B	B	B
Chr.14	88	42 cM	B	B	B	B	B	B	B	B
Chr.15	D15Mit11	4,8cM	B	B	S	S	S	B	B	B
Chr.15	D15Mit29	33.9cM	S	B	S	B	B	B	B	B
Chr.15	D15Mit34	53.7cM	S	B	S	B	B	B	B	B
Chr.15	HOX		S	B	S	B	B	B	B	B
Chr.16	D16Mit8	3cM	S	B	B	B	B	B	B	B
Chr.16	D16Mit5	39cM	B	B	B	B	B	B	B	B
Chr.16	D16Mit5	51cM	B	B	B	S	S	B	B	B
Chr.17	D17Mit61	6,1cM	B	B	S	S	S	B	B	B
Chr.17	D17Mit52	14cM	B	B	S	S	S	B	B	B
Chr.17	D17Mit9	26cM	B	B	S	S	S	B	B	B
Chr.17	D17Mit185	35.2	S	S	B	S	S	B	B	B
Chr.17	D17Mit39	40cM	B	S	B	S	S	B	B	B
Chr.18	D18Mit18	0cM	S	B	S	B	B	B	B	B
Chr.18	D18Mit53	24cM	S	B	S	B	B	B	B	S
Chr.18	D18Mit45	37,4cM	S	B	S	B	B	B	B	B
Chr.18	D18Mit42	45,6cM	S	B	S	B	B		B	B
Chr.19	D19Mit32	1cM	S	B	B	B	B	B	B	B
Chr.19	D19Mit13	20cM	S	B	S	B	B		S	B
Chr.19	D19Mit92	56,6cM	S			B		B		
Chr.X	DXMit53	9,6cM	B	B	B	B	B	B	B	B
Chr.X	DXMit50	22,4cM	B	B	B	B	B	B	B	B
Chr.X	DXMit93	34,2cM	B	S	B	B	B	B	B	B
Chr.X	DXMit116	45,7cM	B	S	S	B	B	B	B	B
Chr.X	DXMit34	60cM	B	B	S	B	B	B	B	B
Chr.X	DXMit31	70,7cM	B	B	S	B	B	B	B	B

Genotyping data corresponding to the anchor markers used in the typing of DNAs from the N4 generation of the breeding programme to generate consomic mice.

Chromos	Primers		15	17	18	32	37	38	41	42
Chr.1	D1Nds4	7cM	B	S	S	B	S	B	B	S
Chr.1	D1Mit24	42,3cM	S	B	B	B	B	B	B	B
Chr.1	D1Mit17	111cM	B	B	B	B	B	B	B	B
Chr.2	D2Mit1	1cM	B	B	B	B	B	B	B	S
Chr.2	D2Mit11	41cM	B	S	S	B	B	B	S	B
Chr.2	D2Mit55	67cM	S	B	S	B	S	S	S	B
Chr.2	D2Mit52	77cM	S	B	B	B	S	B	B	
Chr.3	D3Mit54	0cM	B	S	B	B	S	B	B	B
Chr.3	D3Mit22	35cM	B	B	B	B	B	B	B	B
Chr.3	D3Mit9	38.6cM	B	B	B	B	B	B	B	B
Chr.3	D3Mit45	75.3cM	B	B	B	B	B	B	S	S
Chr.3	D3Mit19	81cM	B	B	B	B	B	B	S	S
Chr.4	D4Mit19	8cM	B	B	B	B	B	B	B	B
Chr.4	D4Mit38	18.8cM	B	B	B	B	B	B	B	B
Chr.4	ORM 1	38cM	B	B	B	B	B	B	B	B
Chr.4	D4Mit37	47.4cM	S	B	B	B	B	B	B	S
Chr.4	D4Mit33	67cM	B	B	B	B	S	B	B	S
Chr. 5	D5Mit48	0cM	B	B	S	B	S	S	B	B
Chr. 5	D5Nds2	28cM	B	B	B	B	B	B	S	B
Chr. 5	Gus	63cM	B	B	B	B	B	B	S	S
Chr.6	D6Mit47	11cM	S	B	B	B		B	B	B
Chr.6	D6Mit8	31cm	S	B	B	B	B	B	B	B
Chr.6	D6Nds5	41cM	S	B	B	B	S	B	B	B
Chr.6	D6Mit201	67cM	B	B	B	B	S	S	S	S
Chr.7	D7Mit21	3cM	B	B	B	B	B	B	B	S
Chr.7	D7Mit54	15cM	B	B	B	B	B	B	B	S
Chr.7	D7Mit40	51cM	B	B	B	B	B	B	B	B
Chr.7	Fgf-3	74cM	B	B	B	B	B	B	B	B
Chr.8	D8Mit141	4,6cM	B	B	B	S	B	B	B	B
Chr.8	D8Mit40	35cM	B	B	B	B	B	B	B	S
Chr.8	D8Mit35	52cM	B	B	B	B	B	B	B	S
Chr.8	D8Mit13	68,4cM	B	B	B	B	B	B	B	S
Chr.9	D9Mit160	2,2cM	B	B	B	B	B	B	B	B
Chr.9	D9Mit2	13,2cM	B	B	B	B	B	S	B	B
Chr.9	33(Cyp1a2)	31cM	B	B	B	B	B	S	B	B
Chr.9	D9Mit137	64,2cM	B	B	B	B	B	B	B	B
Chr.10	D10Mit77	2,2cM	B	B	B	B	S	B	B	B
Chr.10	D10Mit16	10,2cM	B	B	B	B	S	B	B	B
Chr.10	D10Mit15	23cM	B	B	B	S	B	B	B	B
Chr.10	D10Mit22	38,7cM	B	B	B	S	B	B	B	B
Chr.10	D10Mit67	46,4cM	B			S	B	B		B
Chr.10	D10Mit180	64cM	B	B	B	B	B	B	B	B

Genotyping data corresponding to the anchor markers used in the typing of DNAs from the N4 generation of the breeding programme to generate consomic mice.

Appendix II

Chromos	Primers		15	17	18	32	37	38	41	42
Chr.11	D11Mit1	2,2cM	B	B	B	S	B	B	B	B
Chr.11	D11Mit31	37cM	B	B	S	B	B	B	B	B
Chr.11	D11Mit159	43,5cM	B	S	S	B	B	B	B	B
Chr.11	D11Mit48	82cM	B	S	S	B	B	B	S	B
Chr.12	MODC	1cM	B	B	B	B	S	S	B	B
Chr.12	D12Mit65	22cM	B	B	B	B	S	S	B	B
Chr.12	D12Mit4	33cM	B	B	B	B	S	S	S	B
Chr.12	D12Mit27	50cM	B	B	B	B	S	S	S	B
Chr.12	1(Igh-V)	62cM	B	B	B	B	S	S	S	B
Chr.13	D13Mit80	6.4	S	B	B	B	S	B	S	B
Chr.13	D13 Mit 3	11cM	S	B	B	B	B	B	S	B
Chr.13	D13 Mit 13	30cM	S	B	B	B	B	B	S	B
Chr.13	D13 Mit 28	46cM	B	B	B	B	B	B	B	B
Chr.13	D13 Mit 47	62cM	B	B	B	B	B	B	B	B
Chr.14	D14 Mit1	2 cM	B	B	B	B	B	B	S	B
Chr.14	D14 Mit5	15 cM	B	B	B	B	B	B	S	B
Chr.14	120	26 cM	B	B	B	B	B	B	S	B
Chr.14	88	42 cM	B	B	B	B	B	B	S	B
Chr.15	D15Mit11	4,8cM	B	S	S	B	B	B	B	S
Chr.15	D15Mit29	33.9cM	B	S	S	B	B	B	B	B
Chr.15	D15Mit34	53.7cM	B	S	S	B	B	B	B	B
Chr.15	HOX		B	S	S	B	B	B	B	B
Chr.16	D16Mit8	3cM	B	B	B	S	B	B	B	B
Chr.16	D16Mit5	39cM	B	B	B	S	B	B	B	B
Chr.16	D16Mit5	51cM	B	B	B	S	B	B	B	B
Chr.17	D17Mit61	6,1cM	B	B	S	B	B	B	S	S
Chr.17	D17Mit52	14cM	B	B	S	B	B	B	S	S
Chr.17	D17Mit9	26cM	B	B	S	B	B	B	B	B
Chr.17	D17Mit185	35.2	B	B	B	B	B	B	B	B
Chr.17	D17Mit39	40cM	B	B	B	B	B	B	B	B
Chr.18	D18Mit18	0cM	S	B	S	B	S	S	B	S
Chr.18	D18Mit53	24cM	S	B	S	B	B	B	B	S
Chr.18	D18Mit45	37,4cM	S	B	S	B	S	S	B	S
Chr.18	D18Mit42	45,6cM	S	B	S	B	S	S	B	S
Chr.19	D19Mit32	1cM	B	B	B	B	B	B	B	B
Chr.19	D19Mit13	20cM	B	B	S	B	B	B	B	B
Chr.19	D19Mit92	56,6cM		B	S				B	
Chr.X	DXMit53	9,6cM	B	B	B	B	B	B	B	B
Chr.X	DXMit50	22,4cM	B	B	B	B	B	B	B	B
Chr.X	DXMit93	34,2cM	B	B	B	B	B	B	B	B
Chr.X	DXMit116	45,7cM	B	B	B	B	B	B	B	B
Chr.X	DXMit34	60cM	B	B	S	B	B	B	B	B
Chr.X	DXMit31	70,7cM	B	S	S	B	B	B	B	B

Genotyping data corresponding to the anchor markers used in the typing of DNAs from the N4 generation of the breeding programme to generate consomic mice.

Chromos	Primers		44	46	49
Chr.1	D1Nds4	7cM	S	S	B
Chr.1	D1Mit24	42,3cM	S	B	B
Chr.1	D1Mit17	111cM	S	B	S
Chr.2	D2Mit1	1cM	S	B	B
Chr.2	D2Mit11	41cM	S	S	S
Chr.2	D2Mit55	67cM	S	B	S
Chr.2	D2Mit52	77cM	B	B	B
Chr.3	D3Mit54	0cM	B	B	B
Chr.3	D3Mit22	35cM	B	B	B
Chr.3	D3Mit9	38.6cM	B	B	B
Chr.3	D3Mit45	75.3cM	B	B	B
Chr.3	D3Mit19	81cM	B	B	B
Chr.4	D4Mit19	8cM	B	S	B
Chr.4	D4Mit38	18.8cM	B	S	B
Chr.4	ORM 1	38cM	B	S	B
Chr.4	D4Mit37	47.4cM	S	S	B
Chr.4	D4Mit33	67cM	S	S	S
Chr. 5	D5Mit48	0cM	B	S	S
Chr. 5	D5Nds2	28cM	B	S	S
Chr. 5	Gus	63cM	B	B	S
Chr.6	D6Mit47	11cM	B	B	B
Chr.6	D6Mit8	31cm	B	B	B
Chr.6	D6Nds5	41cM	B	B	B
Chr.6	D6Mit201	67cM	B	S	B
Chr.7	D7Mit21	3cM	B	B	B
Chr.7	D7Mit54	15cM	B	B	B
Chr.7	D7Mit40	51cM	S	B	B
Chr.7	Fgf-3	74cM	S	B	B
Chr.8	D8Mit141	4,6cM	B	B	B
Chr.8	D8Mit40	35cM	S	S	S
Chr.8	D8Mit35	52cM	B	S	B
Chr.8	D8Mit13	68,4cM	B	S	B
Chr.9	D9Mit160	2,2cM	B	S	B
Chr.9	D9Mit2	13,2cM	B	S	B
Chr.9	33(Cyp1a2)	31cM	B	B	B
Chr.9	D9Mit137	64,2cM	B	B	B
Chr.10	D10Mit77	2,2cM	B	S	B
Chr.10	D10Mit16	10,2cM	B	S	S
Chr.10	D10Mit15	23cM	B	B	B
Chr.10	D10Mit22	38,7cM	B	B	B
Chr.10	D10Mit67	46,4cM		B	
Chr.10	D10Mit180	64cM	B	B	S

Genotyping data corresponding to the anchor markers used in the typing of DNAs from the N4 generation of the breeding programme to generate consomic mice.

Chromos	Primers		44	46	49
Chr.11	D11Mit1	2,2cM	B	B	B
Chr.11	D11Mit31	37cM	B	B	B
Chr.11	D11Mit159	43,5cM	B	B	B
Chr.11	D11Mit48	82cM	B	B	B
Chr.12	MODC	1cM	B	S	B
Chr.12	D12Mit65	22cM	B	S	B
Chr.12	D12Mit4	33cM	S	S	B
Chr.12	D12Mit27	50cM	S	S	B
Chr.12	1(Igh-V)	62cM	S	S	S
Chr.13	D13Mit80	6.4	S	S	B
Chr.13	D13 Mit 3	11cM	S	B	B
Chr.13	D13 Mit 13	30cM	S	B	B
Chr.13	D13 Mit 28	46cM	B	B	B
Chr.13	D13 Mit 47	62cM	B	B	B
Chr.14	D14 Mit1	2 cM	B	B	S
Chr.14	D14 Mit5	15 cM	B	B	S
Chr.14	120	26 cM	B	B	S
Chr.14	88	42 cM	B	B	S
Chr.15	D15Mit11	4,8cM	S	S	S
Chr.15	D15Mit29	33.9cM	B	B	B
Chr.15	D15Mit34	53.7cM	B	B	B
Chr.15	HOX		B	B	B
Chr.16	D16Mit8	3cM	B	B	B
Chr.16	D16Mit5	39cM	B	B	B
Chr.16	D16Mit5	51cM	B	B	B
Chr.17	D17Mit61	6,1cM	B	B	B
Chr.17	D17Mit52	14cM	B	B	B
Chr.17	D17Mit9	26cM	B	B	B
Chr.17	D17Mit185	35.2	B	B	B
Chr.17	D17Mit39	40cM	B	B	B
Chr.18	D18Mit18	0cM	S	S	S
Chr.18	D18Mit53	24cM	S	S	S
Chr.18	D18Mit45	37,4cM	S	S	S
Chr.18	D18Mit42	45,6cM	S	S	S
Chr.19	D19Mit32	1cM	B	B	S
Chr.19	D19Mit13	20cM	B	B	B
Chr.19	D19Mit92	56,6cM	B		B
Chr.X	DXMit53	9,6cM	B	B	B
Chr.X	DXMit50	22,4cM	B	B	B
Chr.X	DXMit93	34,2cM	B	B	B
Chr.X	DXMit116	45,7cM	B	B	B
Chr.X	DXMit34	60cM	B	B	B
Chr.X	DXMit31	70,7cM	B	B	B

Genotyping data corresponding to the anchor markers used in the typing of DNAs from the N4 generation of the breeding programme to generate consomic mice.

Appendix III

Chromosome (SEG/Pas)	DNA	Backcross Generation
1	59	N7
3	8, 10	N7
4	52 (F)	N7
6	87	N7
7	76	N6
12	40	N7
13	78, 80	N6
14	98, 99	N7
15	64, 66	N7
16	18	N7
18	2, 3, 5, 6 (F)	N7

DNAs from backcross generation N6/7 containing a single *M.spretus* (SEG/Pas) chromosome in a C57BL/6 genome backgraound (refer also to Section 1, 1.1). F, female.

Appendix IV

DNA (N4)	Genotype
101	1
391	1
421	2
2	/
5	2
8	/
10	1
12	1
15	1
17	1
18	2
32	1
37	2
38	/
41	2
42	2
44	1
46	1
49	1

W36162

DNA (N4)	Genotype
101	1
391	1
421	2
2	/
5	2
8	/
10	1
12	1
15	1
17	2
18	2
32	1
37	1
38	1
41	2
42	1
44	2
46	2
49	2

W33240

DNA (N4)	Genotype
101	2
391	1
421	1
2	/
5	1
8	/
10	1
12	1
15	2
17	1
18	1
32	1
37	1
38	1
41	1
42	2
44	2
46	2
49	2

W36235

DNA (N4)	Genotype
101	1
391	1
421	1
2	1
5	1
8	1
10	2
12	2
15	1
17	1
18	1
32	1
37	1
38	1
41	2
42	1
44	1
46	2
49	2

W39933

DNA (N4)	Genotype
101	2
391	2
421	1
2	/
5	1
8	/
10	1
12	1
15	2
17	1
18	1
32	1
37	2
38	1
41	1
42	1
44	1
46	1
49	1

W36139

DNA (N4)	Genotype
101	1
391	1
421	1
2	1
5	1
8	1
10	1
12	1
15	1
17	1
18	1
32	1
37	1
38	2
41	1
42	1
44	1
46	1
49	1

W33210

DNA (N4)	Genotype
101	1
391	2
421	1
2	/
5	2
8	2
10	1
12	1
15	1
17	1
18	1
32	1
37	/
38	1
41	1
42	1
44	1
46	1
49	1

W39952

DNA (N4)	Genotype
101	1
391	2
421	1
2	1
5	2
8	2
10	1
12	1
15	1
17	1
18	1
32	1
37	1
38	1
41	1
42	1
44	1
46	1
49	1

W39947

DNA (N4)	Genotype
101	2
391	2
421	1
2	1
5	2
8	1
10	1
12	1
15	2
17	1
18	1
32	1
37	1
38	1
41	1
42	1
44	1
46	1
49	1

W39956

DNA (N4)	Genotype
101	2
391	1
421	2
2	1
5	1
8	1
10	1
12	/
15	2
17	1
18	2
32	1
37	2
38	2
41	1
42	2
44	2
46	2
49	2

W35449

DNA (N4)	Genotype
101	1
391	2
421	1
2	/
5	1
8	/
10	1
12	1
15	1
17	1
18	1
32	1
37	1
38	1
41	1
42	1
44	1
46	1
49	1

U58912

DNA (N4)	Genotype
101	1
391	1
421	1
2	1
5	1
8	1
10	1
12	1
15	1
17	1
18	1
32	1
37	1
38	1
41	1
42	2
44	2
46	1
49	1

Th

DNA (N2)	Genotype
2	1
3	1
4	1
5	2
6	1
11	1
14	1
15	1
16	2
17	1
18	2
19	2
20	2
21	2
22	1
23	1
24	2
27	1
28	2
W36179	

DNA (N2)	Genotype
2	2
3	2
4	1
5	1
6	1
11	2
14	2
15	2
16	2
17	2
18	2
19	1
20	2
21	1
22	2
23	1
24	1
27	1
28	1
W36213	

Appendix IV Allocation of ESTs to chromosomes using DNAs derived from intermediate generations of the backcross programme to generate consomic strains (refer also to Section 1, 1.1 and 1.3). $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. N2 and N4 correspond to DNAs derived from the first and third backcross generations respectively.

Appendix V

DNA (EUCIB)	Genotype
LB138	[a]1
LB157	[a]2
LB167	2
LB161	1
LB139	2
LB290	1
LB234	2
LB238	2

W36162

DNA (EUCIB)	Genotype
LB064	1
LB142	2
LB163	2
LB161	2
LB145	1
LB111	2
LB158	1
LB238	2

W33240

DNA (EUCIB)	Genotype
PB171	1
PB160	2
PB187	2
PB179	1
PB191	1
PB201	2
PB230	1
PB186	2

W36235

DNA (EUCIB)	Genotype
PB066	1
PB065	2
PB081	2
PB070	1
LB417	2
LB242	2
LB237	1
LB254	2

W39933

DNA (EUCIB)	Genotype
LB410	1
LB348	2
LB520	2
LB391	1
LB352	2
LB401	2
LB408	1
LB399	[a]/
LB060	1
LB305	2

W36139

DNA (EUCIB)	Genotype
PB048	1
PB204	2
PB026	1
PB078	2
PB074	2
PB054	1
PB067	2
PB053	1
PB083	2
PB066	1

W33210

DNA (EUCIB)	Genotype
LS217	1
LS191	2
LS186	2
LS198	1
LS226	1
LS223	1
LS182	2

W39952

DNA (EUCIB)	Genotype
LB165	1
LB167	2
LB246	1
LB304	2
LB169	1
LB299	2
LB253	1
LB237	2

W39947

DNA (EUCIB)	Genotype
LS455	1
LS468	2
LS572	2
LS321	1
LS379	1
LS549	2
LS576	1
LS362	2

W39956

DNA (EUCIB)	Genotype
LB345	1
LB355	2
LB301	2
LB293	1
LB357	2
LB300	1
LB245	2
LB256	1

W35449

DNA (EUCIB)	Genotype
PB021	1
PB041	2
PB043	2
PB047	1
PB050	2
PB062	1
PB078	2
PB056	2

U58912

DNA (EUCIB)	Genotype
LS154	1
LS261	2
LS186	2
LS272	1
LS192	2
LS260	1
LS274	2
LS220	1

Th

DNA (EUCIB)	Genotype	DNA (EUCIB)	Genotype	DNA (EUCIB)	Genotype
LB175	1	LB300	1	LS175	1
LB162	2	LB309	2	LS176	2
LB168	1	LB257	2	LS190	1
LB160	2	LB303	1	LS186	2
LB164	1	LB295	2	LS192	1
LB200	2	LB254	1	LS174	2
LB201	1	LB301	2	LS191	2
LB240	2	LB249	1	LS193	2
		LB252	2		

W36221 W36213 W36259

Appendix V Allocation of ESTs to a position between anchor markers, on mouse chromosomes, using subset panels of EUCIB DNAs (refer also to Section 1, 1.1 and 1.3). $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively.

Appendix VI

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LB 158	1	[a]No	LB 393	2	[a]Yes	LB 213	/	[a]/
LB 166	1	No	LS 453	1	No	LB 347	2	No
LS 336	1	No	LS 539	1	No	LB 355	1	Yes
LB242	1	No	LS 246	1	No	LB 407	1	Yes
LB 251	1	No	LS 159	1	Yes	LS 464	1	Yes
LB 309	2	Yes	LS 318	1	Yes	LS 488	/	/
LB 384	1	No	LB 201	1	Yes	LB 441	1	Yes
LB 386	2	Yes	LS 379	/	/	LB 508	1	Yes

Localisation of W36179 between EUCIB anchor markers *D19Nds1 and D19Mit1.* The data displayed above are for the PCR-SSCP* typing of W36179 on DNAs demonstrating recombination events in the interval *D19Nds1 and D19Mit1.* [a]1, [a]2 and [a]/ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LB232	1	No	LB445	1	No	LB399	2	No
LB257	1	No	LB247	2	No	LB406	2	No
LB301	1	No	LB251	2	No	LB407	2	No
LB307	1	No	LB255	2	No	LB422	2	No
LB349	1	No	LB290	2	No			
LB396	1	No	LB293	2	No			
LB398	1	No	LB384	2	No			
LB408	1	No	LB386	2	No			

Localisation of W36221 between EUCIB anchor markers *D18Mit20 and D18Mit24.* The data displayed above are for the PCR-SSCP* typing of W36221 on DNAs demonstrating recombination events in the interval *D18Mit20 and D18Mit24.* [a]1, [a]2 and [a]/ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LB233	1	No	LB441	1	No	LB254	2	No
LB245	2	Yes	LB500	1	No	LB256	1	Yes
LB302	1	No	LB531	1	No	LB309	1	Yes
LB344	1	No	LB136	1	Yes	LB339	2	No
LB346	2	Yes	LB139	1	Yes	LB341	1	Yes
LB432	1	No	LB170	1	Yes	LB438	2	No
LB433	2	Yes	LB240	2	No	LB528	2	No

Localisation of W35449 between EUCIB anchor markers *D18Mit8 and D18Mit25.* The data displayed above are for the PCR-SSCP* typing of W35449 on DNAs demonstrating recombination events in the interval *D18Mit8 and D18Mit25.* $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LS175			LS331			LS329		
LS197			LS336			LS330		
LS230			LS361			LS376		
LS321			LS418			LS455		
LS322			LS127			LS481		
LS326			LS152			LS556		
LS327			LS323			LS582		
						LS 587		

Localisation of W36213 between EUCIB anchor markers *D17Mit39 and D17Mit123.* The data displayed above are for the PCR-SSCP* typing of W36213 on DNAs demonstrating recombination events in the interval *D17Mit39 and D17Mit123..* $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LS186	2	Yes	LS469	1	No	LS326	2	No
LS269	2	Yes	LS553	1	No	LS336	1	Yes
LS272	1	No	LS566	1	No	LS451	1	Yes
LS280	1	No	LS155	1	Yes	LS452	1	Yes
LS318	1	No	LS198	2	No	LS459	1	Yes
LS325	1	No	LS219	2	No	LS484	1	Yes
LS334	2	Yes	LS230	2	No			
LS404	2	Yes	LS321	2	No			

Localisation of W39956 between EUCIB anchor markers *D13Mit17* and *D13Mit61*. The data displayed above are for the PCR-SSCP* typing of W39956 on DNAs demonstrating recombination events in the interval *D13Mit17* and *D13Mit61*. $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LB140	1	No	LB292	1	No	LB357	1	Yes
LB202	1	No	LB302	2	Yes	LB384	1	Yes
LB232	1	No	LB307	1	No	LB387	1	Yes
LB242	1	No	LB350	2	Yes	LB389	1	Yes
LB246	1	No	LB247	2	No	LB392	1	Yes
LB257	2	Yes	LB304	1	Yes	LB411	1	Yes
LB291	1	No	LB341	2	No	LB345	1	Yes

Localisation of W39947 between EUCIB anchor markers *D10Mit16 and D10Mit20*. The data displayed above are for the PCR-SSCP* typing of W39947 on DNAs demonstrating recombination events in the interval *D10Mit16 and D10Mit20* $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LS087	/	/	LS583	2	Yes	LS223	1	No
LS192	2	Yes	LS587	2	Yes	LS319	2	Yes
LS226	2	Yes	LS739	2	Yes	LS692	2	Yes
LS230	2	Yes	LS088	/	/	LS697	2	Yes
LS269	2	Yes	LS152	2	No	LS777	2	Yes
LS334	2	Yes	LS173	2	No	LS791	2	Yes
LS455	2	Yes	LS183	1	Yes			
LS487	/	/	LS184	2	No			

Localisation of W39952 between EUCIB anchor markers *D10Mit20 and D10Mit7.* The data displayed above are for the PCR-SSCP* typing of W39952 on DNAs demonstrating recombination events in the interval *D10Mit20 and D10Mit7.* $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
PB030	/	/	PB113	/	/	PB137	1	Yes
PB034	2	Yes	PB125	2	Yes	PB138	1	Yes
PB036	2	Yes	PB134	1	No	PB145	1	Yes
PB044	2	Yes	PB109	1	Yes	PB155	1	Yes
PB064	2	Yes	PB123	2	No	PB181	1	Yes
PB066	2	Yes	PB124	1	Yes	PB190	1	Yes
PB087	2	Yes	PB127	1	Yes			
PB104	2	Yes	PB130	1	Yes			

Localisation of W36139 between EUCIB anchor markers *D6Nds4* and *D6Mit5.* The data displayed above are for the PCR-SSCP* typing of W36139 on DNAs demonstrating recombination events in the interval *D6Nds4* and *D6Mit.* $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
PB029	1	No	PB060	1	No	PB051	1	Yes
PB030	1	No	PB063	/	/	PB064	2	No
PB035	1	No	PB072	1	No	PB070	2	No
PB040	1	No	PB039	2	No	PB102	1	Yes
PB050	1	No	PB041	2	No	PB113	1	Yes
PB053	2	Yes	PB042	2	No	PB183	2	No
PB055	1	No	PB047	2	No			
PB056	1	No	PB049	/	/			

Localisation of W39933 between EUCIB anchor markers *D5Nds8* and *D5Nds2*. The data displayed above are for the PCR-SSCP* typing of W39933 on DNAs demonstrating recombination events in the interval *D5Nds8* and *D5Nds*. $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LS094	2	Yes	LS258	2	Yes	LS338	2	No
LS186	1	No	LS460	2	Yes	LS362	2	No
LS193	1	No	LS483	2	Yes	LS367	2	No
LS214	1	No	LS100	1	Yes	LS495	2	No
LS217	1	No	LS184	2	No	LS548	1	Yes
LS220	2	Yes	LS221	2	No	LS731	1	Yes
LS224	2	Yes	LS327	1	Yes	LS733	2	No

Localisation of W36135 between EUCIB anchor markers *Tryp1* and *D4Mit52*. The data displayed above are for the PCR-SSCP* typing of W36135 on DNAs demonstrating recombination events in the interval *Tryp1* and *D4Mit52*. $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LS188	2	Yes	LS324	2	Yes	LS272	2	No
LS189	2	Yes	LS325	2	Yes	LS333	1	Yes
LS192		/	LS328	2	Yes	LS334	1	Yes
LS193	1	No	LS181	1	Yes	LS465	1	Yes
LS229		/	LS183	2	No	LS376	1	Yes
LS230	2	Yes	LS191	1	Yes	LS404	1	Yes
LS265	2	Yes	LS194	1	Yes			
LS279	2	Yes	LS219	1	Yes			

Localisation of W33240 between EUCIB anchor markers *D2Mit1 and D2Mit11*. The data displayed above are for the PCR-SSCP* typing of W33240 on DNAs demonstrating recombination events in the interval *D2Mit1 and D2Mit11*. $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
PB021	2	Yes	PB053	1	No	PB153	2	No
PB041	2	Yes	PB056	1	No	PB186	1	Yes
PB043	1	No	PB062	1	No	PB203	2	No
PB044	2	Yes	PB031	1	Yes	PB225	1	Yes
PB045	2	Yes	PB078	1	Yes	PB246	2	No
PB047	1	No	PB139	1	Yes	LB062	1	Yes
PB049	/	/	PB142	1	Yes			
PB050	2	Yes	PB148	2	No			

Localisation of W36162 between EUCIB anchor markers *D2Mit11 and D2Nds3*. The data displayed above are for the PCR-SSCP* typing of W36162 on DNAs demonstrating recombination events in the interval *D2Mit11 and D2Nds3*. $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over	EUCIB DNA	Genotype	X-over
LS122	1	No	LS230	1	No	LS215	2	No
LS125	1	No	LS265	1	No	LS222	2	No
LS153	1	No	LS273	1	No	LS316	2	No
LS154	1	No	LS121	2	No	LS329	2	No
LS155	1	No	LS173	2	No	LS372	2	No
LS183	1	No	LS184	2	No	LS404	2	No
LS190	1	No	LS185	2	No			
LS226	2	Yes	LS186	2	No			

Localisation of W36259 between EUCIB anchor markers *Col3a1* and *D1Mit12*. The data displayed above are for the PCR-SSCP* typing of W36259 on DNAs demonstrating recombination events in the interval *Col3a1* and *D1Mit12*. $^{a}1$, $^{a}2$ and $^{a}/$ correspond to homozygote, heterozygote and non-informative SSCP patterns respectively. X-over, recombination event.

Appendix VII

```
   1 ACTAGTCAAT CATTTATTTT CTGTCCTTGC ACCTACAATA ATTTGTCCAT TCCCAATGTA
  61 TGATCTTTAG TTACTGAGAC AGTGGCCTCA TTCCTAATTA GAAGGAATGT TTCCTTGATC
 121 AAAAATTTGT TCCAACCCTG TTTTATTTTC CTAGTGGAAT TAGCCTGTGA CACATGAAGG
 181 CTCTATTTGC AGCTTTTATT CTATAGGGGG CATGGATTTA CTTTTCTTAA TTCAGGAGTT
 241 CTATAAAATC ACTATCAGGA ATTAGGAAGT CATCAATTTG TTTAAACTGC ACTATTTCAT
 301 GAAAGTTCAT CTTCATATTG GTTCTGTAGA AGATGTGCCC AATAGGGTGG GCTGATACAC
 361 AGAAATCATT AGGCTATGTA ACAGAGGCAG GAAAGGCAGG TCAGCAGCGA GAAGCAATCT
 421 TAGGATCAAG TCTCTTTGGA TATTGCCTCT CAGAGTGCAC CCACTGCAGC CATGCAAAAC
 481 GGTGGGTCAT CTCCAGGAAA CTCACTTACT TGCCATAGCC GTTCCCACAT GGCTCTTGTC
 541 ATAAAACCAT GCATAACACA CACACACACA CACACACACA CACACACACA CAGAGAGAGA
 601 GAGAGAGAGA GAGAGAGAGA GAGAGAGAGA GAGAGAGAGA GAGAACCAAG CTCTGCACCT
 661 TTGTTCTGCA GAGCAGTGTC ACCTGTGGAC GCTGAGCAGA AGGGTTCCTC CCTCACATGC
 721 TGTTTTGTGT CTCCCCTCCC CTCCCCTTCT TCATCAATGG TTTCCAGTGA GATGAGTCAA
 781 AAAGTGGCCA AGGAGGGCCC CAGGCTCTCC AAGAACCAGA AGTTCTCCGA GCACTTCAGC
 841 ATCCACTGCT GCCCACCCTT CACCTTCCTC AACTCCAAGC GTGAGATCGT GGACCGCAAG
 901 TACAGCATCT GCAAGAGCGG TTGCTTTTAC CAGAAGAAGG AGGAGGACTG GATCTGCTGT
 961 GCCTGCCAGA AGACTAGGTG AGTGCGGGCT CTGGGCCAGG TCTACACCAC ATTCCCTTTG
1021 CTCCATCCAT GACTGTCAGC TCCCTGTTAG GAAGGTTCTG GAGGGCGAGC AGAAGGTGGG
1081 AGGGAGGGCC GGGCCTTCCT GCTGTTTCCA GTTCCTGCCA GCATCCCTTA AGCGACGGTG
1141 GAGAGCCCAG GCTGCAGCCT CCAGCTTCTT TTGGCCCTCT CTGCACCAGC CGCCGTGCCA
1201 CGTCCCCTCA GAGGCCCAAG CACCAGCCAG CTGCKTCCCC AGTGGTGGTC AGAGCGCCGC
1261 CAGCCAAGCC AAGTCCCCTC TGATGCCAGC CAAGCCAAGG TCCCCACCGA GGCCAGCCAA
1321 GCCAAGTCCC CTTCAAGGAC TGAGCGCCAG CCGCGTCCCC GCCCAGAGGT CCGACCACCA
1381 CCAGCCAAGC AGAAGCCCCC TCAGAAGTCT AAGCAGCCAG CACGCAGCAG CCCCCTCAGA
1441 GGGCCAGGCA CCAGCCGCGG GGGGTCTCCC ACCAGAGCTC CTAGGTTCTG GTAACACCAT
1501 CTCTTGCCTT TTGTCCCCTC TAGCCTAAGG TCAGTAGCTG CTCTCTGCAA GTACTAACAT
1561 GGGGGATCTG TCCACAGTAC CAAGCCTGTG TAATCCACTC CCTGCATTAA ACCCCCTCTG
1621 TTTGAAAGAC CTGGCGTGGC TTCTGTTTTT CCAGATGTTA CCGGTCCTCA CTTTAGATGG
1681 CTGTGTCTGT GCCAATCCAT AGGACTGGAA GTCACTAATT TACCAGCATA GTATTGGCCA
1741 CAGTGGCAGG CTGAGTGTAT GAGCCTCCTT AACCACCTTG TGCCTCAGAG AGACATAGTG
1801 GTAACTGATG AGTTGTATCT GTGCATGTGT GATGGGAGGA CCCGATGCTC GATTTGCTGA
1861 CCCTCCTGAC GCATCTCATC AGATATCTGT TTGCTCATAC AGCCACTCAG CAAACTCTAC
1921 TGCTTTATAG ATAATAAGGT TCATATGCTG TACTGGGGAG AGGAAAATAA TTTGTGATTT
1981 CTGAGTACCA AAAAGAGACA AGCTGGCCCT GATGTACTGT AATCACTCCA GTACTGCCAT
2041 GCTCAGGGTG TTTATGCCCA TGGGAGCCCC AACTGGGAAC TTTCCAGCTC TGCCGTGGGA
2101 ACCTGAGGCT TCACAGAGAA AATTAGGTTG GAGTGAGCAA ACTGGAGAAG GATGGCCAGA
```

2161 GAGGAACTTA CCCATCTGCT CTACCTCGGG TTAGACCAGT TTCCTTTCTT CTTTTTGAGA

2221 CAGACACTCC CACTGTAGTT CATGCTATCC CATATCTTAC TCCGTATCTA ATGTTTACTT

2281 CAAACTTATA AATATCTTAC TGCTTTAAAT TCCCAAATGC TTGGATTAAA CGTGTTAACT

2341 GCAATGCTTG GTGGTGACTT CCTCTTCTTT CTCTTCCTCC TCCTCTCCCT CCTTCCCCCC

2401 TCTGTCTCCC CCCCTCTTCT TCTTTATAGA CTTTATTTTT TGGTTCAGTT TTAGGTCCAC

2461 AGAGAAACTG AGGGAAAGAT AGAAACTTCT ATTCCCTGCC CCGCACACAG AGTAGGCAGC

2521 ATTTTACTCC TCGGAGGTAG ATCTGTGTAG TGGGTCCACA GCGAGCACTG GTTCACATCC

2581 TTTTCACTCT GCCACTCAGC CCCTGCGCAG CTTTAGAGAA CCCCCGTGAG AATCTCACGC

2641 ATGGTTCTTT CAGACAGCAA AGATGCAAGA ACAATTGATT TACCCAGCCA AGAACTAACT

2701 GGCTGCTGCT CACTAGACCT CAGTACCTTA GAGAGGTGTA TCCAGGCTGA CTGATCACTA

2761 AGTCCCTCCA AGGGACTGGA GTAAAAAAGC TGGCTCAGCT GGCTAACTTC TGACATACAG

2821 AAGTGCCCAG AAAATGGAGG GGGCTGAGGA TATTCATTTT GTACAGAAGA GGCAGTCCTT

2881 CATAGGAAGC TAAGCGATCG GCTGGGGTGG TTAGGTTCAA GCTGTAACTG GACGTTCTCT

2941 CAGTCACTCT TTGGGTATCA GCTAAGGTGT TTCCGCTGGC TGTGCCCTAT GTCCCTGGGA

3001 ACGTAGGACC AGCTTCCTAC CAGCCTGCAC CTCTCATGCA CCCCGAGTCC CACCCAATAC

3061 ACAAGCTCCC TTCTCAGAAT CTCAGAGGTC CTAAGGTAAT ATACTAGAAT AGATATTTTA

3121 CCACTGCATC GTTTATTTCG TTTATTGATC TGACTATTTC AAATCAGAGG CGTAGGAGGC

3181 ACATTAATGG GAGCAGTTTA ATCAGGGCAG CTAGTCAGAC ATTTAGAAAG GGTTAAAAAG

3241 GTTTTCAGAA GCTGCCCGAA GTCAGATCTG ACGCTGTTCC TCAGGTATTG TGTTTGACGA

3301 ATCAAACGGT TCAGACAAAG GTCTTTTTTT TGGATCAATA ATAACTCCTT AGAAAAGGCC

3361 CTCTTCCTGG ACCGGAGAAT ATCTGTCTTA GAAATGAATG GGCTTGGTTT TCAGCATCCC

3421 AGTGGA

Appendix VII Contiguous sequence 1

This sequence encompasses exons 3, 4 and 5 of the *Mobp* gene.

Appendix VIII

```
   1 ATTTAGGGCG CTTGTTTTTT CCAAGCCTTA ACAACAGTCT TAGGGCTGTT TTGTTTCTTT
  61 CCGATTTCCT AGGGGGGAAA TTTTTGTAAA TTTCAAAACT TTCAAAAGCT AACTTTCAAA
 121 AGGAGTGAGA TCAGGGTGGA TTACAAGATG ATTGAACCCT CGTGGGCTTC CATCCTTTCA
 181 AACCCTCCTG TTCGGTTAGG ACAATGGTGT TTGATATCTT CCTGGAAGCC AAACCTTCTT
 241 CTCTTGCATC TAAGGAGTTG CCATTATCCT CAGTTTAGGG AGTTCTGCTC CCTAAACTGT
 301 GCTAGCTGAC CCCACATTTC TTTCTGAGCA GGTGCTGCCC TCTGCTGGGC ACGGATTAAT
 361 AGACCTGTTT GTCTTCCTCC CAGGACTGTG CGTAAATAAG GCTCAGGCAA TCTCACACCC
 421 ATCTCTCTAG CTTACACTCA GCTCAGGGCT TGACTGATGT TGATGGCTGA ATGGGACCAA
 481 ACAATTTACC TCACAGTCTG ACTCAGCAGG AAGGTGGACG GCTTCAGCAT CGGGTTTGAG
 541 TCCACACTTT TTGATAAACT TGTAACCAGT GCGTGTCTGG GCCCACCTAT TAGTAATACA
 601 TTTACTAAAT ATTGCTTATC TTTTTAAATG TAGTTCATTC CCTAACAAAT TAGAGACATG
 661 TCAACCTGGT CAGGATTTTA TAAGAGTTCA TATATACAGA TGTGCGCACG GTTGCATTTA
 721 AGTCTGCGGT ATCAGTCACT GCTCTCACTG ATGTGACGGG CTATCTGGGC AAACACGTTA
 781 GGAAGGCTCA TTCTGTTTAA TGTATACAGG AGCTCTGTTT TTATGCATAC CAGACGAGGG
 841 AATCAGATTC TATTACAGAT GGTTGTGAGA CACCATGTGG CTGTTGGGAA TTGAACTCAG
 901 GACCTCTGGA AGAGCAGTAG GTGTCCCTAA CCACTGAGCC ATCTCTCCAG TCCTGGAAGG
 961 ATCATTTTGA CTCACAGTTT GAAGGGACAC AGGCCACCAT GGTGAGCAAA GAGTGGTGGC
1021 AAGAACCTGA GGTGGCAGGT GATATGGCAC CCGCAGTCAG GCAGCAGAGA GGTGACCGCT
1081 GGTGCTCAGC TAGCTTTCTC CTTCCTATTC AGTCCCTGGC CCTGAGACCA GGGTCTGTTG
1141 TCCCTCACAT TCCAGCCGAG TCTTCTCCGC TCACTTAAAC CTTTCTGCGA ACTTGTCCAC
1201 AAACACAGAG ATGTGTTTCC TTGGTGACCC TAAATCCAGG CAAGTTGACA ATAAAGAGCA
1261 AGAGAAAGTG AGCTGAACAC CAGCGCTTAC CTCTCTGCTG CCTGACTGTG GATGGCATGT
1321 GACCTGCAGA GCCTCAGGTC CTTGTCACAA GTAATAAAAT GTTTTTGTTT CTGTTCTAAG
1381 AGCAGGATAT ACTTATGCTA GCTTTGTGTT CTTCTGGGTC CCTTACCACC AGTAGCCAGA
1441 AAACCAAATA TTATTATGCA TGGAAACAGG GATATATATA TATATATATC TCCTTTCCTA
1501 TGCTCAGTTC AGGAAAAAAA GAAATCAACA TATCAGTTTT GCCCAGAGAT TAAAAGGAGT
1561 TGAGAAAGCT GGATTTCTAA GGAGAAAAAA CAAAACAAAA AAACAAAACA AAACAAAAAA
1621 ATCATCTCAA GTGGTTCACT CAGAGCTCAG CAGTTAAAAG CCCTTCTTGC TCTTGTAGGA
1681 CCTGAGTTCA GTTCCCAGCA CCTACATGGT GGCTTACAAC TCTATAACTT TAGTTTTAGC
1741 GCCCTCTTCT GGAACCTGTG AGAACTGCAT GCACATGGTG CACATAGACA CACACAGACA
1801 CACAGACACA CACAGACACA CAGACACACA CAGACACACA GACACACACA GACACAGACA
1861 CACACACACA CACACACACA CACACACACA CACACACACA CACACACACA CACACACACA
1921 GAGTAAACAC ATCTTTTTAA AATGCCCGTG AAGTCACTGT TGTTCTAAGT GCTCTGCAGG
1981 AAGTGGACAC CACCTACCAA TTCATAAATC TTTCTTTTTC TCTCTATAAA TCAGATTGAG
2041 GAGGAGGAGC AGGTCAAACC CCAAGGAAGA AGTGACCGAG AAGGGACTTC TTGTGAATGG
2101 ACAGCCTCTG CCTGTGGACT CACTGCTTCC CGCAACCCAT CTGCTCTCAG ACGAATTATC
```

2161 TGACCTCAAG TATGATGCAG GAGTGGACGC CTGCTGATGT AACAACCACA GATGCCCTCA

2221 GTCCCCATGG TGTGGGAGCC TTAGGGCAGC CTTCCTGGAG GTGAGAAGGG GGCTCTTCGT

2281 GTCGTACTCT AGCCACGGAG TCCTGGAAGG ATCTCTTGTG CTCCAGAGCT CAGGGAACGG

2341 GTCTTTTGGT TTCATCTGGA GCCCCTTTCT CTCTGGCAGG TGCACGCTTC CTCAGGGTCT

2401 TTGGCAAAGG TGATTGCAGA GCTAATCCAC ACACATTCCT CCTTCTTCCT CTCCCCTCCC

2461 TCCCTGCATC TGGTCAGCAA ACATGACTCT TGATACTTTT TTTTCTCTTC CTGGCTCTCT

2521 GTCCGGCCTC TCTTGCTCGC ACACCATTTG CTTCCCATTC ACCTCTTAGC CAGCAGTACT

2581 CAGAGCCAGA GTCGAGTCTG GTGAAAGGGC CATGGAATTG TAAGATACTA AGTTGTTGTA

2641 TGATCCTTAA ATGCGGCGAT GTTCCGATAA GTTAGGGGAT CGCGCTAAAT TCCCATGTGA

2701 ACTGTAGGTA ACTTGTGTCA GAAACAGAGG AAATGGAGGC ATCCTCCTTC CTCCCCCTCC

2761 GCGGGCCCAG TCTCCAGCCT CACACAGGGC TTTGAGAGAA TCAGTTCCTG CCCCAGTCTG

2821 GGGAGAACCA CCCTCTAAAG CCTTGCCTTT GAGCTTGTCA CCTGCACAGG CGCCCGGCGC

2881 AGCTCTGACT GCTTTCAAAT GATATCATTT AGGAAGGTCA TATGGCTTCA GCCCTGTCCC

2941 CCAACCCTCC CAATCACTTC CCTCGCAGCA TCCTGGACAA GTCCCCATGC GCACTTGTAT

3001 AGGTCTAGAA AGAGTGCCAG GAGCGTTTCT GTCTCCAAAT CAAAACCCCT CAAATTCCTG

3061 AGAGGCCTGA TGCTCCAGTG CTTCCCCTGG CTGGCCCTCC GCGGCTATGT CTGTGGCAAG

3121 GAGAGGTGAA GGAAAACAGG AAGGGAACTT GGCCTTCTGG AAAGGAAGGC AAAGTGCGTG

3181 TGTGTGTGCG TGTGTGTGTT TGCGCGTGCG CGTACGCGTG CGCGTGTGAT GTGTGCGTGT

3241 GTTAGAGAGT TATCAGTCCA AGATCCCATC CCTGAGTAGG AGGCCTTCCA CGGGATGAAG

3301 GGGTCTGTGC TCACTGTGGA GGTGGAGTAG CGCGGGGCTG AGTTTCCGTC TCCTCGCTCT

3361 GGTCTGCCTC CACAGTCCGT AGCCTCCAGA CTCCCGGGGA AGAATGAGCC CGAGCATGGG

3421 CTCTTTTTGT TCTAGACCAC AGGAGCCAGC CCTCCCCCAT GGCGTGGCAC TTGCTTGTAC

3481 CCATCTGTAC ACTTGCTGTG GGGACAGAGA TTGGGCCAGA TGGAAGCGCC TCGTACTGCC

3541 TTGTACTCGT GTGGACTGCT GTGTGGGTGT TGGCCTTGGG AGTAGCTTGG GGAGAATATG

3601 TGCCAGCTCT CAGTTTTCTT GCCCAAGTGC AAATGCTAGA GGCCCTTCCA AGTCCCAGGT

3661 TCATTCTTGT CCCCCAAAGG CCTCAACTGC AGGAGGTCGT CATGAACTGC AGGAGGTCGT

3721 CATGAAAAGG CCACAACGGT ACATGCAACA CTTCAGAGTC CCAAAGCCTT CTCAGCAGGG

3781 GGCACATGGT GATCGGCTTG GGCCTCTTTT CCCTAGTCAG CCCAGATGGG CGGGGGCGGA

3841 GGGTGGAAGG AGAGGAGTTG GATAGATCCA TCTCTAGTCA GGCTCCATCT GGGTCGGCCC

3901 CACAACTTCA GCTTGCTTTG AGCCCAAACT TTGGTAAGTT ATAAACACAA AGAATAGAGA

3961 AAGAGAGATA GAGAGAGACA CACACAAACA CATACACATT ACACAGATAT ATACACACAC

4021 ACATATACCA CACACACGCA CGCACGCACG CACGCACACG CACACGCGCA CACACACACA

4081 CACAAACCTT GTCACTCCTC ACCCCTCCTA ACACTTTTCT CCTTAGGAGG CCAGTGGACG

4141 GGAAGGAGGC CATGTCTCTA TTCACAAGGA AACCAGGCTC AGGGCCTGAC TCAGTGGTTA

4201 AAAAATTCTT GCCTTGCAAG CATGAGAGCT AGAGTTCAAA TCCCCATGAC CCACACGAAT

4261 GCTGGGTAGG CCTGACAACG TACCTCTAAA TCCAAACTGA GAAGCTGGTG ACAGGAGATC

4321 CCAGAGCAAG CAAGAGTGGC CATATCTGTG TGCTTTGGGA CTGAGTGGGA AACCCTTCTT

4381 CTGTGACTAT GGTAGAAGAT ATAGGACTGA CTCCCAATGT CTAACTCTGG GACTCCATGC

4441 ACATGCACCC CTGTATGCTT GTGCACATAC ACTTGCACAC ACACACACAC ACACACACAC

```
4501 ACACACACAC ACACCACACC ACACCACACC ACACACCACA CATCCACATG GAAATGGGGA
4561 AAAAGGAAAC AAGAGAGTAC CAGACTCATG GTTCTTCGAG GATGGGTGCA TCCTAGCCAG
4621 GACTTAGAAG CCTTTCCCCA CCCAGGAAGA CCTCACCATT TCTTCCTCCT GTTCCCTCTC
4681 TTGTACGTGT GAGCTGCTAA TGGGGTGCTT CATTCACCAT GGGGCTGTGT GCATGCTCCT
4741 CAGAGCCTGT CACGCTCTGG CCTACGGCTC CTACACAGGT GCCCTGAGCG CAGATATCTT
4801 CTCAGAGATT CCTCCACGGT CACCAGACTC CATCATTGTC TTCCAGGATC ATCTGCTCAC
4861 TTATTT
```

Appendix VIII Contiguous sequence 2

This sequence encompasses exons 6 and 7 of the *Mobp* gene.

Appendix IX

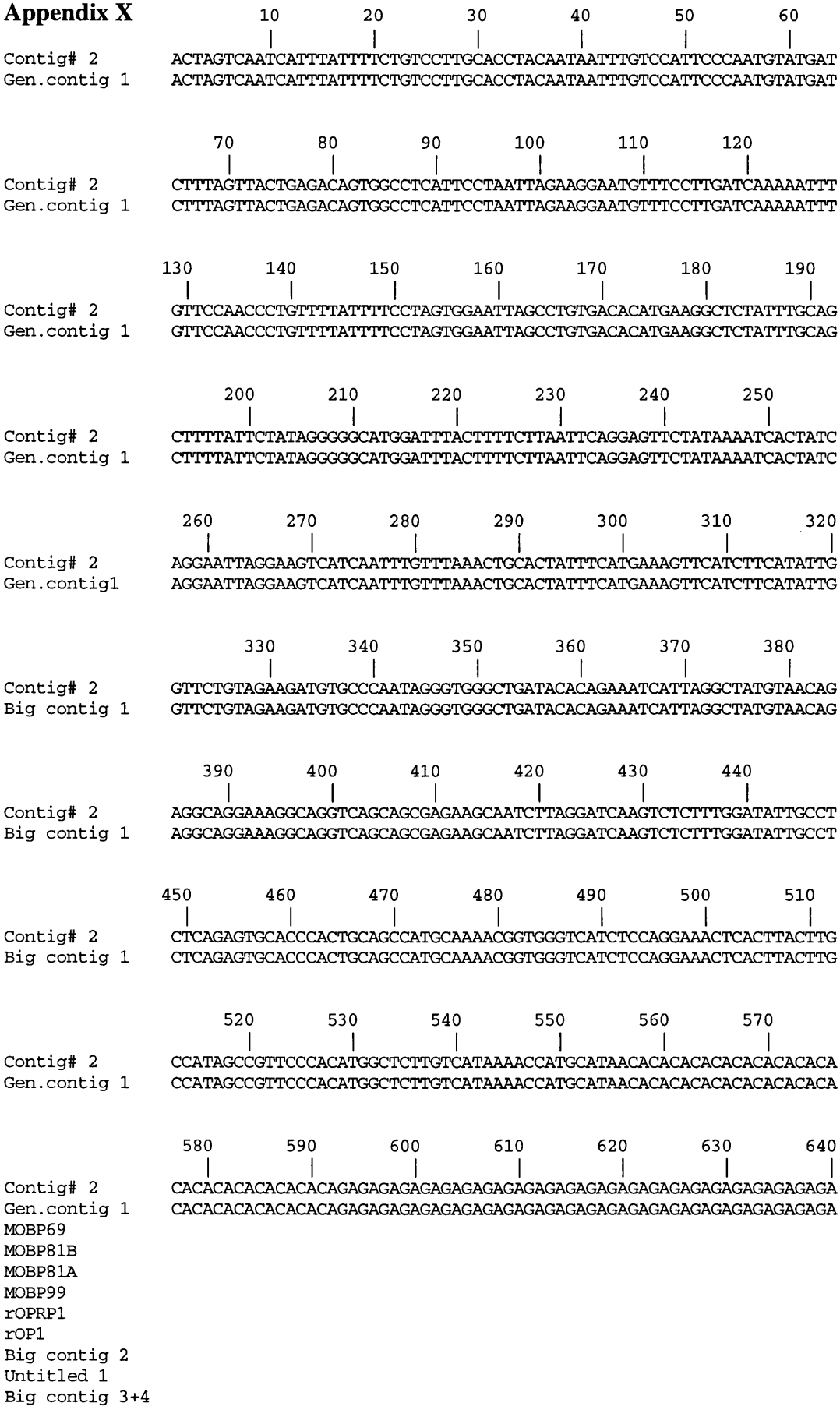
```
   1 TGGATCTCCT TTGGGAAACT TTTTTTTGGT TTATTCTAGG AATAGGTTTT TTTTGGGAGG
  61 TAGAAAAAGA TTCAGAACAA CAAACAAGGT ACATTCCGCA GGGCGGGAGA GGGGCCCTTC
 121 CAGCTTACTG ACCGGAGTTG TTGGTTGAGC TGGTTTCAGG GGTGTGGCGT TCCATGGGTC
 181 TTGCCCCTCT CTGCCCTCCC TATCCTCAGG GCCAAGTGTG TGACCAGAGC TCGGTGTGTG
 241 CTTCCCTTCA CAGATATGTC TTCTTCTTCA TGAGAGCGGG TACCTGCTTC TGAAGGATGT
 301 AGGAGGCCAA CGGGAACGCA GGGGTCCAAC CGCCAGGAGA ACCAGGGGAG GAACAGCCCC
 361 ATTTAGTTCT GATGTTTTTT CTCTCCCCTC AGCATGTGGA CGATGCCATT CCATCACCTC
 421 CCTCTATAAC TGCCAGCGAG ATATCTCTGT CAACTGTCTC GACTCTCCGA GGTAAGATTC
 481 TTTTTCACTT TCTGTTATTT TTTAACAAGG TTTTTCCTCT CTGTCACTGA CATTCCAAAT
 541 GCTGGCTGTG ATGTCACTAA GTACAGATTA TATGTGCTTG TTCCGCCCAG GATTCGGGTT
 601 CTTCAGGCTG AAGATTCTTG TGTCTTCTCG GATGTGCAAA ACAAGGTCAT GGTGCACAAC
 661 CATGCAAACC GCAAATGCAA AAGAGAAATT GTCAAGGTTG GCGGTGACGT GAGAAAAATC
 721 CTAAGCAGCT ACCGATGGGA GTCTAACATT GTCACCACTG AGCTGTACTC TGGGGACACT
 781 GTCAAAGGCA AAATATCTTA AAACCTCTAA TCAAGTGACT GCAATTTTAG GAATCCATCC
 841 TGGAGAAAAC TCACACCCCT TTAGAGAAGG AGAAATGCGC AGATTTGTCT AAAAGATCGT
 901 GGTTTGTAGT TAAGAGAAGT GGAATGAAAC TAGCCAAAAT GTCTGTCCAA GGGAATGAGA
 961 GAATCAATTT TGGGACGTTC ACAAATAAAG TCCTAACAGC GGTTAAAAGC CAATCTGTCA
1021 GATAGACCAT GTCATGTAGG TAAATCTAGA AAGCAAACTG AGAACTATAA AAATCAAGTA
1081 CCGGGCAGTA AATGCAGTAT GACGTTCTGG GCCTGAATAT TGACATCCCA CAGTCCTGTA
1141 TTGGTTATTG CTTAATAAGG AGAAGCCACA AGATCTCACA CTTTTTCTAT AATGCAATAG
1201 GATTTAAAAA TGACAGCAGA ACACAAAGAC ACTAGAAATT GTCCTGTCAG AAATAAAAGT
1261 ATAAGAAATA ATTCCTGGCT TGAGGAAATA ACCCAGACTA AAAGCTACAT AGAAACGGCT
1321 TAACAGTACT GGCAGGGCAT ATATTTGTTT GGAGAGGGTC GATGTGAAGT CTGTGAATAG
1381 TCTGGGATGT TTACAAGGTC GCATAAATTT TAAACACCAA AAATATTTGT TTTGAATCAA
1441 TTTTGGGTCT GGGCTATAGA GTCTGCCTTG TGATGGTTCC TGACACGAGA GAAAAGGGAA
1501 GAAATGTTAA TACAGCTCTC GAATTTTCCC TTTAAAGACC AGGTAGGTAG GAAAGGACAA
1561 CAGGATGTTG GGGAATGTGG AGGGATGTAA AACAGACAGT ATTACGGTGG CGAAAACTCA
1621 GTGTAGGGCG ATAAGAACAT CAGATGGCCT TGTGCTGGGA GCCGGGGTGG CCAAGCACAG
1681 GCGGATGCTT GAAGACCCCA AAGGTCACCA TGGCGTCCTC ACAGCCAGAC CCGGCAAAGT
1741 TCAGATCAAG TGCACGGATG GGGGCCATGT TAGTGGAAAC TGTCCAAGGT GTGCGGAGGG
1801 GGTGAGGAAG ACAGCCAGGA GGAAGAGATG GCTCAGAATG CTGCACTACA GCAAAGAAGA
1861 AGCAGGCCAA ATATCGTCCT AAGGCAGCCG AGGGGTCCCG GCTTCCTCTC CCGAGGGCAC
1921 TGTGGCTCCG TGGAAGTGCC GAGGCAGAGG CAGGGGAATT TAACTGCACA AAAACTACAC
1981 GCAGAAACTC AGGTGCTAAA ATGGACATGT ATGTGACCAC CTTTCTGAGG ACCAGTTCCA
2041 GTGAGTGCCC AGGATGTTTC CTTCTCGACT ACGGTTACTG AGCTTAGAAG AATGGGAACG
2101 GGTTTAGGAG CACACAGTTA GCCCCATTGG TGAGACAAAT TTCCAGGCTA GCTTCATCCA
```

```
2161 ATCCATTAAG GAGACTTTAG GCTTTTTCAT CTGGTTTTCA TTGGGGCAAA GCAAAGCATA
2221 TCCTGAGTGT TGGTTTGTTT CCGCCCTCCA CGCCACCCCT GAGACACCGA GCATGCGCAT
2281 TAACCCAGAG CGCGCTAACC CATGCGCAGC GTTGGACGTT AGCTCATTAA TGTGATTAAC
2341 CAGAAACCAC TAAATTTTGT TTGTTGTAGA AATAAAACCT GGGGATGGAT CAAACTATAT
2401 CTGGAGATTG CTTGAAGCTG ATGGGTTACT AAGCCAAATG TGCTTTGGAA TTCCTCACTG
2461 AGTCGCGTTG TAATCACCAC GTACTGTTTT TCTACTTAAC ATCGACTCTC AGGAACATTC
2521 CCAAGTTAAT ACCGGAGACA TTTCATAAAA CTCCCAAAGC TGTTCTTCAT ACATTTCTTG
2581 AATTAGCAAT CCAAGCCCCC ATGTTCTGAC ATTATTCCCG AATGTTTAAT TAATCCTTAC
2641 ATTTTGGGGA AAAAACGGAA GGC
```

Appendix IX Contiguous sequence 3

This sequence encompasses exon 8 of the *Mobp* gene

Appendix X


10 20 30 40 50 60
Contig# 2 ACTAGTCAATCATTTATTTTCTGTCCTTGCACCTACAATAATTTGTCCATTCCCAATGTATGAT
Gen.contig 1 ACTAGTCAATCATTTATTTTCTGTCCTTGCACCTACAATAATTTGTCCATTCCCAATGTATGAT
70 80 90 100 110 120
Contig# 2 CTTTAGTTACTGAGACAGTGGCCTCATTCCTAATTAGAAGGAATGTTTCCTTGATCAAAAATTT
Gen.contig 1 CTTTAGTTACTGAGACAGTGGCCTCATTCCTAATTAGAAGGAATGTTTCCTTGATCAAAAATTT
130 140 150 160 170 180 190
Contig# 2 GTTCCAACCCTGTTTTATTTTCCTAGTGGAATTAGCCTGTGACACATGAAGGCTCTATTTGCAG
Gen.contig 1 GTTCCAACCCTGTTTTATTTTCCTAGTGGAATTAGCCTGTGACACATGAAGGCTCTATTTGCAG
200 210 220 230 240 250
Contig# 2 CTTTTATTCTATAGGGGGCATGGATTTACTTTTCTTAATTCAGGAGTTCTATAAAATCACTATC
Gen.contig 1 CTTTTATTCTATAGGGGGCATGGATTTACTTTTCTTAATTCAGGAGTTCTATAAAATCACTATC
260 270 280 290 300 310 320
Contig# 2 AGGAATTAGGAAGTCATCAATTTGTTTAAACTGCACTATTTCATGAAAGTTCATCTTCATATTG
Gen.contig1 AGGAATTAGGAAGTCATCAATTTGTTTAAACTGCACTATTTCATGAAAGTTCATCTTCATATTG
330 340 350 360 370 380
Contig# 2 GTTCTGTAGAAGATGTGCCCAATAGGGTGGGCTGATACACAGAAATCATTAGGCTATGTAACAG
Big contig 1 GTTCTGTAGAAGATGTGCCCAATAGGGTGGGCTGATACACAGAAATCATTAGGCTATGTAACAG
390 400 410 420 430 440
Contig# 2 AGGCAGGAAAGGCAGGTCAGCAGCGAGAAGCAATCTTAGGATCAAGTCTCTTTGGATATTGCCT
Big contig 1 AGGCAGGAAAGGCAGGTCAGCAGCGAGAAGCAATCTTAGGATCAAGTCTCTTTGGATATTGCCT
450 460 470 480 490 500 510
Contig# 2 CTCAGAGTGCACCCACTGCAGCCATGCAAAACGGTGGGTCATCTCCAGGAAACTCACTTACTTG
Big contig 1 CTCAGAGTGCACCCACTGCAGCCATGCAAAACGGTGGGTCATCTCCAGGAAACTCACTTACTTG
520 530 540 550 560 570
Contig# 2 CCATAGCCGTTCCCACATGGCTCTTGTCATAAAACCATGCATAACACACACACACACACACACA
Gen.contig 1 CCATAGCCGTTCCCACATGGCTCTTGTCATAAAACCATGCATAACACACACACACACACACACA
580 590 600 610 620 630 640
Contig# 2 CACACACACACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
Gen.contig 1 CACACACACACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Big contig 2
Untitled 1
Big contig 3+4

```
                          650        660        670        680        690        700
                           |          |          |          |          |          |
Contig# 2          GAGAACCAAGCTCTGCACCTTTGTTCTGCAGAGCAGTGTCACCTGTGGACGCTGAGCAGAAGGG
Gen.contig 1       GAGAACCAAGCTCTGCACCTTTGTTCTGCAGAGCAGTGTCACCTGTGGACGCTGAGCAGAAGGG
MOBP69              TATCCACA GGAACCTTTCACAGCAGCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGC
MOBP81B                    GGAACCTTTCACAGCAGCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGC
MOBP81A                         CTTTCACAGCAGCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGC
MOBP99                                  GCCAATACCTGCAGGGCAACAAAGAATCAAATGAGAGC
rOPRP1                                                                  TGAGAGC
rOP1                                                                    TGAGAGC
Big contig2a
Gen.contig2b
Gen contig 3+4

                              710        720        730        740        750        760
                               |          |          |          |          |          |
Contig# 2          TTCCTCCCTCACATGCTGTTTTGTGTCTCCCCTCCCCTCCCCTTCTTCATCAATGGTTTCCAGT
Big contig 1       TTCCTCCCTCACATGCTGTTTTGTGTCTCCCCTCCCCTCCCCTTCTTCATCAATGGTTTCCAGT
MOBP69             GAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAG     *     T
MOBP81B            GAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAG     *     T
MOBP81A            GAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAG     *     T
MOBP99             GAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAG     *     T
rOPRP1             GAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAG     *     T
rOP1               GAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGATGGGAGCTTGAAAACACAG     *     T
Big contig2a
Gen.contig2b
Gen contig 3+4

                         770        780        790        800        810        820        830
                          |          |          |          |          |          |          |
Contig# 2          GAGATGAGTCAAAAAGTGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCCGAGC
Gen contig 1       GAGATGAGTCAAAAAGTGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCCGAGC
MOBP69             AAGATGAGTCAAAAAGTGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGC
MOBP81B            AAGATGAGTCAAAAAGTGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGC
MOBP81A            AAGATGAGTCAAAAAGTGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGC
MOBP99             AAGATGAGTCAAAAAGTGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGC
rOPRP1             AAGATGAGTCAAAAAGTGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGC
rOP1               AAGATGAGTCAAAAAGTGGCCAAGGAGGGCCCCAGGCTCTCCAAGAACCAGAAGTTCTCAGAGC
Gen contig2a
Gen.contig2b
Gen contig 3+4

                              840        850        860        870        880        890
                               |          |          |          |          |          |
Contig# 2          ACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCG
Gen contig 1       ACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCG
MOBP69             ACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCG
MOBP81B            ACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCG
MOBP81A            ACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCG
MOBP99             ACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCG
rOPRP1             ACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCG
rOP1               ACTTCAGCATCCACTGCTGCCCACCCTTCACCTTCCTCAACTCCAAGCGTGAGATCGTGGACCG
Gen contig2a
Gen.contig2b
Gen contig 3+4

                         900        910        920        930        940        950        960
                          |          |          |          |          |          |          |
Contig# 2          CAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGGAGGACTGGATCTGCTGT
Gen contig 1       CAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGGAGGACTGGATCTGCTGT
MOBP69             CAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGGAGGACTGGATCTGCTGT
MOBP81B            CAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGGAGGACTGGATCTGCTGT
MOBP81A            CAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGGAGGACTGGATCTGCTGT
MOBP99             CAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGGAGGACTGGATCTGCTGT
rOPRP1             CAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGGAGGACTGGATCTGCTGT
rOP1               CAAGTACAGCATCTGCAAGAGCGGTTGCTTTTACCAGAAGAAGGAGGAGGACTGGATCTGCTGT
Gen contig2a
Gen.contig2b
Gen contig 3+4
```

```
                          970        980        990       1000       1010       1020
                           |          |          |          |          |          |
Contig# 2          GCCTGCCAGAAGACTAG|GTGAGTGCGGGCTCTGGGCCAGGTCTACACCACATTCCCTTTGCTCC
Gen contig 1       GCCTGCCAGAAGACTAG|GTGAGTGCGGGCTCTGGGCCAGGTCTACACCACATTCCCTTTGCTCC
MOBP69             GCCTGCCAGAAGACCAG|
MOBP81B            GCCTGCCAGAAGACCAG|
MOBP81A            GCCTGCCAGAAGACCAG|
MOBP99             GCCTGCCAGAAGACCAG|
rOPRP1             GCCTGCCAGAAGACCAG|
 rOP1              GCCTGCCAGAAGACCAG|
Gen contig2a
Gen.contig2b
Gen contig 3+4

                     1030       1040       1050       1060       1070       1080
                        |          |          |          |          |          |
Contig# 2          ATCCATGACTGTCAGCTCCCTGTTAGGAAGGTTCTGGAGGGCGAGCAGAAGGTGGGAGGGAGGG
Gen contig 1       ATCCATGACTGTCAGCTCCCTGTTAGGAAGGTTCTGGAGGGCGAGCAGAAGGTGGGAGGGAGGG
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen.contig2b
Gen contig 3+4

                1090       1100       1110       1120       1130       1140       1150
                   |          |          |          |          |          |          |
Contig# 2          CCGGGCCTTCCTGCTGTTTCCAGTTCCTGCCAGCATCCCTTAAGCGACGGTGGAGAGCCCAGGC
Gen contig 1       CCGGGCCTTCCTGCTGTTTCCAGTTCCTGCCAGCATCCCTTAAGCGACGGTGGAGAGCCCAGGC
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                    1160       1170       1180       1190       1200       1210
                       |          |          |          |          |          |
Contig# 2          TGCAGCCTCCAGCTTCTTTTGGCCCTCTCTGCACCAG|CCGCCGTGCCACGTCCCCTCAGAGGCC
Gen contig 1       TGCAGCCTCCAGCTTCTTTTGGCCCTCTCTGCACCAG|CCGCCGTGCCACGTCCCCTCAGAGGCC
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1                                                  |CCGCCGTGCCACATCCCCTCAGAAGCC
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                1220       1230       1240       1250       1260       1270       1280
                   |          |          |          |          |          |          |
Contig# 2          CAAGCACCAGCCAGCTGCKTCCCCAGTGGTGGTCAGAGCGCCGCCAGCCAAGCCAAAGTCCCCT
Gen contig 1       CAAGCACCAGCCAGCTGCKTCCCCAGTGGTGGTCAGAGCGCCGCCAGCCAAGCCAAAGTCCCCT
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1             CAAGCACCAGCCAGCTGCATCCCCTGTGGTGGTCAGAGCGCCGCCAGCCAAGCCAAAGTCCCCT
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4
```

```
                        1290       1300       1310       1320       1330       1340
                          |          |          |          |          |          |
Contig# 2          CTGATGCCAGCCAAGCCAAGGTCCCCACCGAGGCCAGCCAAGCCAAGGTCCCCTTCAAGGACTG
Gen contig 1       CTGATGCCAGCCAAGCCAAGGTCCCCACCGAGGCCAGCCAAGCCAAGGTCCCCTTCAAGGACTG
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1             CCGAGGCCAGCCAAGCCAAGGTCCCCTCCGATTCCAGCCAAGCCAAGGTCCCCTTCAAGGACTG
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                     1350       1360       1370       1380       1390       1400
                       |          |          |          |          |          |
Contig# 2          AGCGCCAGCCGCGTCCCCGCCCAGAGGTCCGACCACCACCAGCCAAGCAGAAGCCCCCTCAGAA
Gen contig 1       AGCGCCAGCCGCGTCCCCGCCCAGAGGTCCGACCACCACCAGCCAAGCAGAAGCCCCCTCAGAA
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1             AGCGCCAGCCGCGTCCCCGCCCAGAGGTCCGACCACCACCAGCCAAGCAGAAGCCCCCTCAGAA
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                  1410       1420       1430       1440       1450       1460       1470
                    |          |          |          |          |          |          |
Contig# 2          GTCTAAGCAGCCAGCACGCAGCAGCCCCCTCAGAGGGCCAGGCACCAGCCGCGGGGGGTCTCCC
Gen contig 1       GTCTAAGCAGCCAGCACGCAGCAGCCCCCTCAGAGGGCCAGGCACCAGCCGCGGGGGGTCTCCC
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1             GTCTAAACAGCCAGCACGAAGCAGCCCCCTCAGAGGGCCAGGCACCAGCCGCGGGGGGTCTCCC
 rOP1
Gen contig2a
Gen contig2b
Big contig 3+4

                       1480       1490       1500       1510       1520       1530
                         |          |          |          |          |          |
Contig# 2          ACCAGAGCTCCTAGGTTCTGGTAACACCATCTCTTGCCTTTTGTCCCCTCTAGCCTAAGGTCAG
Gen contig 1       ACCAGAGCTCCTAGGTTCTGGTAACACCATCTCTTGCCTTTTGTCCCCTCTAGCCTAAGGTCAG
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1             ACCAGAGCTCCTAGGTTCTGGTAACACCATCTCTTGCCTTTTGTCCCCCCTAGCCTAAGGTCAG
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                     1540       1550       1560       1570       1580       1590       1600
                       |          |          |          |          |          |          |
Contig# 2          TAGCTGCTCTCTGCAAGTACTAACATGGGGGATCTGTCCACAGTACCAAGCCTGTGTAATCCAC
Gen contig 1       TAGCTGCTCTCTGCAAGTACTAACATGGGGGATCTGTCCACAGTACCAAGCCTGTGTAATCCAC
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1             TAGCTGCTCTCTGCAAGTACTAACATGGGGGATCTCTCCACAGTACCAAGC TGTGTAATCTAC
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4
```

```
                        1610       1620       1630       1640       1650       1660
                          |          |          |          |          |          |
Contig# 2               TCCCTGCATTAAACCCCCTCTGTTTGAAAGACCTGGCGTGGCTTCTGTTTTTCCAGATGTTACC
Gen contig 1            TCCCTGCATTAAACCCCCTCTGTTTGAAAGACCTGGCGTGGCTTCTGTTTTTCCAGATGTTACC
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1                  TCC TGCATTAAACCCCCTCTGTTTG
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                      1670       1680       1690       1700       1710       1720
                        |          |          |          |          |          |
Contig# 2               GGTCCTCACTTTAGATGGCTGTGTCTGTGCCAATCCATAGGACTGGAAGTCACTAATTTACCAG
Gen contig 1            GGTCCTCACTTTAGATGGCTGTGTCTGTGCCAATCCATAGGACTGGAAGTCACTAATTTACCAG
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                    1730       1740       1750       1760       1770       1780       1790
                      |          |          |          |          |          |          |
Contig# 2               CATAGTATTGGCCACAGTGGCAGGCTGAGTGTATGAGCCTCCTTAACCACCTTGTGCCTCAGAG
Gen contig 1            CATAGTATTGGCCACAGTGGCAGGCTGAGTGTATGAGCCTCCTTAACCACCTTGTGCCTCAGAG
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                        1800       1810       1820       1830       1840       1850
                          |          |          |          |          |          |
Contig# 2               AGACATAGTGGTAACTGATGAGTTGTATCTGTGCATGTGTGATGGGAGGACCCGATGCTCGATT
Gen contig 1            AGACATAGTGGTAACTGATGAGTTGTATCTGTGCATGTGTGATGGGAGGACCCGATGCTCGATT
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                      1860       1870       1880       1890       1900       1910       1920
                        |          |          |          |          |          |          |
Contig# 2               TGCTGACCCTCCTGACGCATCTCATCAGATATCTGTTTGCTCATACAGCCACTCAGCAAACTCT
Gen contig 1            TGCTGACCCTCCTGACGCATCTCATCAGATATCTGTTTGCTCATACAGCCACTCAGCAAACTCT
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4
```

```
                          1930        1940        1950        1960        1970        1980
                           |           |           |           |           |           |
Contig# 2          ACTGCTTTATAG|ATAATAAGGTTCATATGCTGTACTGGGGAGAGGAAAATAATTTGTGATTTCT
Gen contig 1       ACTGCTTTATAG|ATAATAAGGTTCATATGCTGTACTGGGGAGAGGAAAATAATTTGTGATTTCT
MOBP69                         |ATAATAAGGTTCTATTCACTGTACTGGGGAGAGGAATATGATTTGTGATTCC
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                        1990        2000        2010        2020        2030        2040
                           |           |           |           |           |           |
Contig# 2          GAGTACCAAAAAGAGACAAGCTGGCCCTGATGTACTGTAATCACTCCAGTACTGCCATGCTCAG
Gen contig 1       GAGTACCAAAAAGAGACAAGCTGGCCCTGATGTACTGTAATCACTCCAGTACTGCCATGCTCAG
MOBP69             CGAGGACCAAGAAGAGAGAAGCTGGCCCTGGTGTACACAGTAATCACTCCAATACTGCCGTGCC
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                    2050        2060        2070        2080        2090        2100        2110
                      |           |           |           |           |           |           |
Contig# 2          GGTGTTTATGCCCATGGGAGCCCCAACTGGGAACTTTCCAGCTCTGCCGTGGGAACCTGAGGCT
Big contig 1       GGTGTTTATGCCCATGGGAGCCCCAACTGGGAACTTTCCAGCTCTGCCGTGGGAACCTGAGGCT
MOBP69             AGGATGTATGACCAAAGGAGCCCTGAATGGCGATTGTCCAGCTCTGCCATGGGAGCCTGAGGCT
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Big contig 2
Untitled 1
Big contig 3+4

                         2120        2130        2140        2150        2160        2170
                           |           |           |           |           |           |
Contig# 2          TCACAGAGAAAATTAGGTTGGAGTGAGCAAACTGGAGAAGGATGGCCAGAGAGGAACTTACCCA
Gen contig 1       TCACAGAGAAAATTAGGTTGGAGTGAGCAAACTGGAGAAGGATGGCCAGAGAGGAACTTACCCA
MOBP69             TCACAGAGAAAATTAGGTTGGAGCGGCGTA CAGGAGAAGGATAGACAGAGAGGCACTTACCAA
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                    2180        2190        2200        2210        2220        2230        2240
                      |           |           |           |           |           |           |
Contig# 2          TCTGCTCTACCTCGGGTTAG|ACCAGTTTCCTTTCTTCTTTTTGAGACAGACACTCCC  ***
Gen contig 1       TCTGCTCTACCTCGGGTTAG|ACCAGTTTCCTTTCTTCTTTTTGAGACAGACACTCCC  ***
MOBP69             TCTGCTCTGCCTTAG|
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
```

```
                      2250       2260       2270       2280       2290       2300
                       |          |          |          |          |          |
Contig# 2          ATTTAGGGCGCTTGTTTTTTCCAAGCCTTAACAACAGTCTTAGGGCTGTTTTGTTTCTTTC
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       ATTTAGGGCGCTTGTTTTTTCCAAGCCTTAACAACAGTCTTAGGGCTGTTTTGTTTCTTTC
Gen contig2b
Gen contig 3+4

                   2310       2320       2330       2340       2350       2360
                    |          |          |          |          |          |
Contig# 2       CGATTTCCTAGGGGGGAAATTTTTGTAAATTTCAAAACTTTCAAAAGCTAACTTTCAAAAGGAG
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a    CGATTTCCTAGGGGGGAAATTTTTGTAAATTTCAAAACTTTCAAAAGCTAACTTTCAAAAGGAG
Gen contig2b
Gen contig 3+4

                2370       2380       2390       2400       2410       2420       2430
                 |          |          |          |          |          |          |
Contig# 2       TGAGATCAGGGTGGATTACAAGATGATTGAACCCTCGTGGGCTTCCATCCTTTCAAACCCTCCT
Big contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Big contig2a    TGAGATCAGGGTGGATTACAAGATGATTGAACCCTCGTGGGCTTCCATCCTTTCAAACCCTCCT
Gen contig2b
Big contig 3+4

                    2440       2450       2460       2470       2480       2490
                     |          |          |          |          |          |
Contig# 2       GTTCGGTTAGGACAATGGTGTTTGATATCTTCCTGGAAGCCAAACCTTCTTCTCTTGCATCTAA
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a    GTTCGGTTAGGACAATGGTGTTTGATATCTTCCTGGAAGCCAAACCTTCTTCTCTTGCATCTAA
Gen contig2b
Gen contig 3+4

                2500       2510       2520       2530       2540       2550       2560
                 |          |          |          |          |          |          |
Contig# 2       GGAGTTGCCATTATCCTCAGTTTAGGGAGTTCTGCTCCCTAAACTGTGCTAGCTGACCCCACAT
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a     GGAGTTGCCATTATCCTCAGTTTAGGGAGTTCTGCTCCCTAAACTGTGCTAGCTGACCCCACAT
Gen contig2b
Gen contig 3+4
```

```
                        2570       2580       2590       2600       2610       2620
                          |          |          |          |          |          |
Contig# 2          TTCTTTCTGAGCAGGTGCTGCCCTCTGCTGGGCACGGATTAATAGACCTGTTTGTCTTCCTCCC
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       TTCTTTCTGAGCAGGTGCTGCCCTCTGCTGGGCACGGATTAATAGACCTGTTTGTCTTCCTCCC
Gen contig2b
Gen contig 3+4

                     2630       2640       2650       2660       2670       2680
                       |          |          |          |          |          |
Contig# 2          AGGACTGTGCGTAAATAAGGCTCAGGCAATCTCACACCCATCTCTCTAGCTTACACTCAGCTCA
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1                 |GACTGTGTGAATCAGGCTCAGGCAATCTTACACTCATCTCTCTAGAGCCCAGCTCAGGGTTT
Gen contig2a       AG|GACTGTGCGTAAATAAGGCTCAGGCAATCTCACACCCATCTCTCTAGCTTACACTCAGCTCA
Gen contig2b
Gen contig 3+4

                  2690       2700       2710       2720       2730       2740       2750
                    |          |          |          |          |          |          |
Contig# 2          GGCTGACGTCGGCGACTGAATGAGACCAGATAATTTACCTCACAGTCTGACCCAGCAAGAAAGT
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1               GGCTGACGTCGGCGACTGAATGAGACCAGATAATTTACCTCACAGTCTGACCCAGCAAGAAAGT
Gen contig2a       GGGCTTGACTGATGTTGATGGCTGAATGGGACCAAACAATTTACCTCACAGTCTGACTCAGCAG
Gen contig2b
Gen contig 3+4

                        2760       2770       2780       2790       2800       2810
                          |          |          |          |          |          |
Contig# 2          GAAGGTGGACGGCTTCAGCATCGGGTTTGAGTCC ***         AATCAG|ATTGAGAAGGA
Gen contig 1
MOBP69                                                              |ATTGAGAAGGA
MOBP81B                                                             |ATTGAGAAGGA
MOBP81A                                                             |ATTGAGAAGGA
MOBP99
rOPRP1
rOP1               CAG                                              |ATTGAGAAGGA
Gen contig2a       GAAGGTGGACGGCTTCAGCATCGGGTTTGAGTCC ***         AATCAG|ATTGAGGAGGA
Gen contig2b
Gen contig 3+4

                  2820       2830       2840       2850       2860       2870       2880
                    |          |          |          |          |          |          |
Contig# 2          GGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGACTTCTCGTGAATGGACAGCCTCTG
Gen contig 1
MOBP69             GGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGACTTCTCGTGAATGGACAGCCTCTG
MOBP81B            GGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGACTTCTCGTGAATGGACAGCCTCTG
MOBP81A            GGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGACTTCTCGTGAATGGACAGCCTCTG
MOBP99
rOPRP1
rOP1               GGAGCAGGTCAACCCCAAGAAAAAAGTGACCGAGAAGGGACTTCTCGTGAATGGACAGCCTCTG
Gen contig2a       GGAGCAGGTCAACCCCAAGGAAGAAGTGACCGAGAAGGGACTTCTTGTGAATGGACAGCCTCTG
Gen contig2b
```

```
                          2890       2900       2910       2920       2930       2940
                            |          |          |          |          |          |
Contig# 2          CCTGTGGACTCACTGCTTCCCGCAACCCATCTGCTCTCAGACGAATTATCTGACCTCAAGTATG
Gen contig 1
MOBP69             CTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTATCTGGCCTCGAGTATGAC
MOBP81B            CTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTATCTGGCCTCGAGTATGAC
MOBP81A            CTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTATCTGGCCTCGAGTATGAC
MOBP99
rOPRP1
rOP1               CTTGTGGATTTACTGCTTCACCGCAACCCATCTGCCCTCAGACTTATCTGGCCTCGAGTATGAC
Gen contig2a       CCTGTGGACTCACTGCTTCCCGCAACCCATCTGCTCTCAGACGAATTATCTGACCTCAAGTATG
Gen contig2b
Gen contig 3+4

                        2950       2960       2970       2980       2990       3000
                          |          |          |          |          |          |
Contig# 2          ATGCAGGAGTGGACGCCTGCTGATGTAACAACCACAGATGCCCTCAGTCCCCATGGTGTGGGAG
Gen contig 1
MOBP69             GCAGGGGTGGACACCTGCTGATGTAACAACTACAGATGCCCTCAGTCCCCATGGTGTGGGAGCC
MOBP81B            GCAGGGGTGGACACCTGCTGATGTAACAACTACAGATGCCCTCAGTCCCCATGGTGTGGGAGCC
MOBP81A            GCAGGGGTGGACACCTGCTGATGTAACAACTACAGATGCCCTCAGTCCCCATGGTGTGGGAGCC
MOBP99
rOPRP1
rOP1               GCAGGGGTGGACACCTGCTGATATAACAACTACAGATGCCCTCAGTCCCCATGGTGTGGGAGCC
Gen contig2a       ATGCAGGAGTGGACGCCTGCTGATGTAACAACCACAGATGCCCTCAGTCCCCATGGTGTGGGAG
Gen contig2b
Big contig 3+4

                    3010       3020       3030       3040       3050       3060       3070
                      |          |          |          |          |          |          |
Contig# 2          CCTTAGGGCAGCCTTCCTGGAG|GTGAGAAGGGGGCTCTTCGTGTCGTACTCTAGCCACGGAGTC
Big contig 1
MOBP69             TTAGGGCAGCCTGCCTGGAG|
MOBP81B            TTAGGGCAGCCTGCCTGGAGGTGAGGAGGGAGCTCTTCATGCCTTACTCTAGCCACAGAGTCCT
MOBP81A            TTAGGGCAGCCTGCCTGGAG|
MOBP99
rOPRP1
rOP1               TTAGGGCAGCCTGCCTGGAG|
Gen contig2a       CCTTAGGGCAGCCTTCCTGGAG|GTGAGAAGGGGGCTCTTCGTGTCGTACTCTAGCCACGGAGTC
Gen contig2b
Gen contig 3+4

                         3080       3090       3100       3110       3120       3130
                           |          |          |          |          |          |
Contig# 2          CTGGAAGGATCTCTTGTGCTCCAGAGCTCAGGGAACGGGTCTTTTGGTTTCATCTGGAGCCCCT
Gen contig 1
MOBP69
MOBP81B            GGAAGGACCCCTCGGGCTCCAGAGCTCAGGGAATGGGTCTTTTGGTTTCATCTGGAGCCCTTTT
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       CTGGAAGGATCTCTTGTGCTCCAGAGCTCAGGGAACGGGTCTTTTGGTTTCATCTGGAGCCCCT
Gen contig2b
Gen contig 3+4

                    3140       3150       3160       3170       3180       3190       3200
                      |          |          |          |          |          |          |
Contig# 2          TTCTCTCTGGCAGGTGCACGCTTCCTCAGGGTCTTTGGCAAAGGTGATTGCAGAGCTAATCCAC
Gen contig 1
MOBP69
MOBP81B            CTCTCTGGCAGGTGCACAGTTCCTCAGGGTCTTTGGCAAAGGTGATTACAGAGTCAATCCACAC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       TTCTCTCTGGCAGGTGCACGCTTCCTCAGGGTCTTTGGCAAAGGTGATTGCAGAGCTAATCCAC
Gen contig2b
Gen contig 3+4
```

```
                        3210       3220       3230       3240       3250       3260
                          |          |          |          |          |          |
Contig# 2          ACACATTCCTCCTTCTTCCTCTCCCCTCCCTCCCTGCATCTGGTCAGCAAACATGACTCTTGAT
Gen contig 1
MOBP69
MOBP81B            ACATTCCTCCTCCTTCCAATTCCCCTCTCTCCCTGGGTCAGGTTGACAACCTGGCTCTTTTTCC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       ACACATTCCTCCTTCTTCCTCTCCCCTCCCTCCCTGCATCTGGTCAGCAAACATGACTCTTGAT
Gen contig2b
Gen contig 3+4

                     3270       3280       3290       3300       3310       3320
                       |          |          |          |          |          |
Contig# 2          ACTTTTTTTTTCTCTTCCTGGCTCTCTGTCCGGCCTCTCTTGCTCGCACACCATTTGCTTCCCAT
Gen contig 1
MOBP69
MOBP81B            TTTCCTTCCTAATTCTCTGCCTGGCCTCTCTTGATCACGTGACATCTGTCTCCCATTGACCTTC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       ACTTTTTTTTTCTCTTCCTGGCTCTCTGTCCGGCCTCTCTTGCTCGCACACCATTTGCTTCCCAT
Gen contig2b
Gen contig 3+4

                  3330       3340       3350       3360       3370       3380       3390
                    |          |          |          |          |          |          |
Contig# 2          TCACCTCTTAGCCAGCAGTACTCAGAGCCAGAGTCGARTCTGGTGAAAGGGCCATGGAATTGTA
Gen contig 1
MOBP69
MOBP81B            TTAATCAGGAGTGGTCAGAGCCAACGTCCACACTGGTGAAGGCTCAGGTTATTTTTGGAGTAAG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       TCACCTCTTAGCCAGCAGTACTCAGAGCCAGAGTCGARTCTGGTGAAAGGGCCATGGAATTGTA
Gen contig2b
Gen contig 3+4

                      3400       3410       3420       3430       3440       3450
                        |          |          |          |          |          |
Contig# 2          AGATACTAAGTTGTTGTATGATCCTTAAATGCGGCGATGTTCCGATAAGTTAGGGGATCGCGCT
Gen contig 1
MOBP69
MOBP81B            ATACAGAGTTGCTGTATGATCCTTAAATGCAGCAATGTTCTCATAAGAATGGGATCCCGTTAAA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       AGATACTAAGTTGTTGTATGATCCTTAAATGCGGCGATGTTCCGATAAGTTAGGGGATCGCGCT
Gen contig2b
Gen contig 3+4

                    3460       3470       3480       3490       3500       3510       3520
                      |          |          |          |          |          |          |
Contig# 2          AAATTCCCATGTGAACTGTAGGTAACTTGTGTCAGAAACAGAGGAAATGGAGGCATCCTCCTTC
Gen contig 1
MOBP69
MOBP81B            TCTTCACGTCAACTGTAGGTAACTTGTGTCAGAAACAGAGAAAAGGGGGCATCTTCCTCCCTCC
MOBP81A
MOBP99
rOPRP1
rOP1
Big contig2a       AAATTCCCATGTGAACTGTAGGTAACTTGTGTCAGAAACAGAGGAAATGGAGGCATCCTCCTTC
Gen contig2b
Big contig 3+4
```

```
                        3530       3540       3550       3560       3570       3580
                         |          |          |          |          |          |
Contig# 2          CTCCCCCTCCGCGGGCCCAGTCTCCAGCCTCACACAGGGCTTTGAGAGAATCAGTTCCTGCCCC
Gen contig 1
MOBP69
MOBP81B            CCCTCTGCAGGCCCAGTCTCCAGCGTTACACAGGGCTTTGTGAGCACCAGCTCCTGCCCCAGTC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       CTCCCCCTCCGCGGGCCCAGTCTCCAGCCTCACACAGGGCTTTGAGAGAATCAGTTCCTGCCCC
Gen contig2b
Gen contig 3+4

                      3590       3600       3610       3620       3630       3640
                       |          |          |          |          |          |
Contig# 2          AGTCTGGGGAGAACCACCCTCTAAAGCCTTGCCTTTGAGCTTGTCACCTGCACAGGCGCCCGGC
Gen contig 1
MOBP69
MOBP81B            TGGGGATAACAACCCTCTAAACACTGACTTTGACTTGTCCCCAGTGCAGTTGGCACTCAGCCTC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       AGTCTGGGGAGAACCACCCTCTAAAGCCTTGCCTTTGAGCTTGTCACCTGCACAGGCGCCCGGC
Gen contig2b
Gen contig 3+4

                   3650       3660       3670       3680       3690       3700       3710
                    |          |          |          |          |          |          |
Contig# 2          GCAGCTCTGACTGCTTTCAAATGATATCATTTAGGAAGGTCATATGGCTTCAGCCCTGTCCCCC
Gen contig 1
MOBP69
MOBP81B            ACAGACTCTCAGTGCAGCTCTGACTGCTCTCAATTATATAATTTAGGAAGGTCATGTGGCTTCA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       GCAGCTCTGACTGCTTTCAAATGATATCATTTAGGAAGGTCATATGGCTTCAGCCCTGTCCCCC
Gen contig2b
Gen contig 3+4

                       3720       3730       3740       3750       3760       3770
                        |          |          |          |          |          |
Contig# 2          AACCCTCCCAATCACTTCCCTCGCAGCATCCTGGACAAGTCCCCATGCGCACTTGTATAGGTCT
Gen contig 1
MOBP69
MOBP81B            GCCCTGTCCCCCCATTCCCCATCACCTCCCTTGAGGCATCCTGGACAAGTCCCATGTGCACTTGT
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       AACCCTCCCAATCACTTCCCTCGCAGCATCCTGGACAAGTCCCCATGCGCACTTGTATAGGTCT
Gen contig2b
Gen contig 3+4

                   3780       3790       3800       3810       3820       3830       3840
                    |          |          |          |          |          |          |
Contig# 2          AGAAAGAGTGCCAGGAGCGTTTCTGTCTCCAAATCAAAACCCCTCAAATTCCTGAGAGGCCTGA
Gen contig 1
MOBP69
MOBP81B            ATAGGTCTAGAAAGAATGCTTGGAGAGTTTCTGTCTCCAGGTCAAAATTCCTCAAATTCCTGGG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       AGAAAGAGTGCCAGGAGCGTTTCTGTCTCCAAATCAAAACCCCTCAAATTCCTGAGAGGCCTGA
Gen contig2b
Gen contig 3+4
```

```
                        3850       3860       3870       3880       3890       3900
                         |          |          |          |          |          |
Contig# 2       TGCTCCAGTGCTTCCCCTGGCTGGCCCTCCGCGGCTATGTCTGTGGCAAGGAGAGGTGAAGGAA
Gen contig 1
MOBP69
MOBP81B         AAGGGCAGAGGCTTGATGCTCCAGTGTTTTCCGTGGCTGACCCTCCACATCTGGGTCTGTGGTA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a    TGCTCCAGTGCTTCCCCTGGCTGGCCCTCCGCGGCTATGTCTGTGGCAAGGAGAGGTGAAGGAA
Gen contig2b
Gen contig 3+4

                      3910       3920       3930       3940       3950       3960
                       |          |          |          |          |          |
Contig# 2       AACAGGAAGGGAACTTGGCCTTCTGGAAAGGAAGGCAAAGTGCGTGTGTGTGTGCGTGTGTGTG
Gen contig 1
MOBP69
MOBP81B         AGGGAGAAGTCAAGAAAAAGGGGAAAGGAACTTGACTTTCTATAAAGGAAGGTAGTGTGTGTGT
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a    AACAGGAAGGGAACTTGGCCTTCTGGAAAGGAAGGCAAAGTGCGTGTGTGTGTGCGTGTGTGTG
Gen contig2b
Gen contig 3+4

                   3970       3980       3990       4000       4010       4020       4030
                    |          |          |          |          |          |          |
Contig# 2       TTTGCGCGTGCGCGTACGCGTGCGCGTGTGATGTGTGCGTGTGTTAGAGAGTTATCAGTCCAAG
Gen contig 1
MOBP69
MOBP81B         GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTACGTGTGCACATGTGTGTGTTAGA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a    TTTGCGCGTGCGCGTACGCGTGCGCGTGTGATGTGTGCGTGTGTTAGAGAGTTATCAGTCCAAG
Gen contig2b
Gen contig 3+4

                        4040       4050       4060       4070       4080       4090
                         |          |          |          |          |          |
Contig# 2       ATCCCATCCCTGAGTAGGAGGCCTTCCACGGGATGAAGGGGTCTGTGCTCACTGTGGAGGTGGA
Gen contig 1
MOBP69
MOBP81B         GAGCTAGCAATCCAAGAGTCCTTCCCAGACTAGGAGGCCAATCCATGGTATGAAGGAGTTTGTG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a    ATCCCATCCCTGAGTAGGAGGCCTTCCACGGGATGAAGGGGTCTGTGCTCACTGTGGAGGTGGA
Gen contig2b
Gen contig 3+4

                   4100       4110       4120       4130       4140       4150       4160
                    |          |          |          |          |          |          |
Contig# 2       GTAGCGCGGGGCTGAGTTTCCGTCTCCTCGCTCTGGTCTGCCTCCACAGTCCGTAGCCTCCAGA
Gen contig 1
MOBP69
MOBP81B         CTCAACTGTCTATGTGGAGCAGGGTTGGGCTGAGTTTCTGTCTCCTTGCTCTAGTTTGCCTCCA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a    GTAGCGCGGGGCTGAGTTTCCGTCTCCTCGCTCTGGTCTGCCTCCACAGTCCGTAGCCTCCAGA
Gen contig2b
Gen contig 3+4
```

```
                        4170       4180       4190       4200       4210       4220
                          |          |          |          |          |          |
Contig# 2          CTCCCGGGGAAGAATGAGCCCGAGCATGGGCTCTTTTTGTTCTAGACCACAGGAGCCAGCCCTC
Gen contig 1
MOBP69
MOBP81B            CGATCTGTAGCTCAGACTCTTGGGAAAGAATGAGCTTCGGTATGGACACTTGTTCTAGACCATA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       CTCCCGGGGAAGAATGAGCCCGAGCATGGGCTCTTTTTGTTCTAGACCACAGGAGCCAGCCCTC
Gen contig2b
Gen contig 3+4

                     4230       4240       4250       4260       4270       4280
                       |          |          |          |          |          |
Contig# 2          CCCCATGGCGTGGCACTTGCTTGTACCCACCTGTACACTTGCTGTGGGGACAGAGATGGGGCCA
Gen contig 1
MOBP69
MOBP81B            GGAGCCAACCCTCCCCCATGGCACTTGCTTGTACCTCATCTGCACTCTTGCTGTGAGGACAGAG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       CCCCATGGCGTGGCACTTGCTTGTACCCACCTGTACACTTGCTGTGGGGACAGAGATGGGGCCA
Gen contig2b
Gen contig 3+4

                  4290       4300       4310       4320       4330       4340       4350
                    |          |          |          |          |          |          |
Contig# 2          ATGGAAACGCCTCCTACTGCCTTGTACTCCTGTAGAATACGTTGTAGGTGTTGGCCCCTGGGAA
Gen contig 1
MOBP69
MOBP81B            ATTGGGCCAGATGTAACCGCCTTGTACTGCCTTGTACTCGTGTGGACTGGCGTGTGGGTGTCGG
MOBP81A
MOBP99
rOPRP1
Gen contig2a       ATGGAAACGCCTCCTACTGCCTTGTACTCCTGTAGAATACGTTGTAGGTGTTGGCCCCTGGGAA
Gen contig2b
Gen contig 3+4

                      4360       4370       4380       4390       4400       4410
                        |          |          |          |          |          |
Contig# 2          TACCTCGGGAGAATAGT
Gen contig 1
MOBP69
MOBP81B            CGTTGGGATTAGCTTGGGGAGATATTTGCCAGCTCTCGGTTTTCTTGCCCAAGTGCAGATGCTA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a       TACCTCGGGAGAATAGT
Gen contig2b
Gen contig 3+4

                    4420       4430       4440       4450       4460       4470       4480
                      |          |          |          |          |          |          |
Contig# 2
Gen contig 1
MOBP69
MOBP81B            GAGAGCCCCTCCAAGTCCCAACTTAATTCTTGCCTCCCTGAAGACCTCAACCGTAGAAGGTCAT
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4
```

```
                      4490       4500       4510       4520       4530       4540
                         |          |          |          |          |          |
Contig# 2
Gen contig 1
MOBP69
MOBP81B         CATGAAAAGGCCACCACTTCAGAGTCCTGGCCTAAACTCAGAGTCCACATGCCTTCCCAGCAGG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                    4550       4560       4570       4580       4590       4600
                       |          |          |          |          |          |
Contig# 2
Gen contig 1
MOBP69
MOBP81B         GGCCACAGAGGTGATCCACTTGGGCCTCTTTCCCCACTCAGCCCAGATGGGTGGGGGTGGAGGG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                4610       4620       4630       4640       4650       4660       4670
                   |          |          |          |          |          |          |
Contig# 2
Gen contig 1
MOBP69
MOBP81B         TGGAGGGTGGAAGGAGAGGAGTTGGAGAAAATCCATGTCTATTCAGGCTCCGTCTGGGTCTGCT
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                      4680       4690       4700       4710       4720       4730
                         |          |          |          |          |          |
Contig# 2
Gen contig 1
MOBP69
MOBP81B         GCACAACAACTTCAGCTTGTTTGAGCCCAAACTTACATAAGTTATAATCACAAAGCTGGAACTA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b                                                              GCCCC
Gen contig 3+4

                  4740       4750       4760       4770       4780       4790       4800
                     |          |          |          |          |          |          |
Contig# 2       AAAAATTTAGGTTGTTTGGGGCCCAAAATTGGTAAGTTTTAAACACAAGGATTGGGGAAGGGGG
Gen contig 1
MOBP69
MOBP81B         GAGAAAGAGAGAAGGACACACACACACACACACACACACACACACACACATACACACACACTCA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b    AAAAATTTAGGTTGTTTGGGGCCCAAAATTGGTAAGTTTTAAACACAAGGATTGGGGAAGGGGG
Gen contig 3+4
```

```
                          4810       4820       4830       4840       4850       4860
                           |          |          |          |          |          |
Contig# 2          TGGGGGGGGACACACACCAACACTACACACTACACGGGTACTAACACACACACATATACCAACA
Gen contig 1
MOBP69
MOBP81B            TATATACACATACACACATATACACATATATACACATACACACGCATACACACATACACACATA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig 2
Gen contig2b       TGGGGGGGGACACACACCAACACTACACACTACACGGGTACTAACACACACACATATACCAACA
Gen contig 3+4

                     4870       4880       4890       4900       4910       4920
                      |          |          |          |          |          |
Contig# 2          CACACGCACGCACGCACGCACACGCACACGCACAACAACACACACACAAACCTTGTCACT
Gen contig 1
MOBP69
MOBP81B            TACACACATACATAACACACATACACATGCATACACACACTCATATATACACATACACACATAT
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       CACACGCACGCACGCACGCACACGCACACGCACAACAACACACACACAAACCTTGTCACT
Gen contig 3+4

                   4930       4940       4950       4960       4970       4980       4990
                    |          |          |          |          |          |          |
Contig# 2          CCTCACCCCTCCTAACACTTTTCTCCTTAGGAGGCCAGTGGACGGGAAGGAGGCCATGTCTCTA
Gen contig 1
MOBP69
MOBP81B            ACACATATATATACGCATACACACATACACACATATACACACATACATAACACACATACACATG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       CCTCACCCCTCCTAACACTTTTCTCCTTAGGAGGCCAGTGGACGGGAAGGAGGCCATGTCTCTA
Gen contig 3+4

                        5000       5010       5020       5030       5040       5050
                         |          |          |          |          |          |
Contig# 2          TTCACAAGGAAACCAGGCTCAGGGCCTGACTCAGTGGTTAAAAAATTCTTGCCTTGCAAGCATG
Big contig 1
MOBP69
MOBP81B            CATACACACACACACACACACACACACACACACACACACACACACACACCCTCGTCACTCCTCATC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       TTCACAAGGAAACCAGGCTCAGGGCCTGACTCAGTGGTTAAAAAATTCTTGCCTTGCAAGCATG
Gen contig 3+4

                   5060       5070       5080       5090       5100       5110       5120
                    |          |          |          |          |          |          |
Contig# 2          AGAGCTAGAGTTCAAATCCCCATGACCCACACGAATGCTGGGTAGGCCTGACAACGTACCTCTA
Gen contig 1
MOBP69
MOBP81B            CCTCCTAACATTTTCTCCTTAGGAGGCTAGTAGAAGGGAAGGAAGCCGTTTCTATTTACAAGGA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       AGAGCTAGAGTTCAAATCCCCATGACCCACACGAATGCTGGGTAGGCCTGACAACGTACCTCTA
Gen contig 3+4
```

```
                          5130       5140       5150       5160       5170       5180
                           |          |          |          |          |          |
Contig# 2          AATCCAAACTGAGAAGCTGGTGACAGGAGATCCCAGAGCAAGCAAGAGTGGCCATATCTGTGTG
Gen contig 1
MOBP69
MOBP81B            AGTCCGGCTCACGGCCCGGGAGATGACTGCCTTGCAGGCATGGGGGATAGAGTTCGGATCCCCG
MOBP81A
MOBP99
rOPRP1
rOP1
Big contig2a
Gen contig2b       AATCCAAACTGAGAAGCTGGTGACAGGAGATCCCAGAGCAAGCAAGAGTGGCCATATCTGTGTG
Big contig 3+4

                        5190       5200       5210       5220       5230       5240
                         |          |          |          |          |          |
Contig# 2          CTTTGGGACTGAGTGGGAAACCCTTCTTCTGTGACTATGGTAGAAGATATAGGACTGACTCCCA
Gen contig 1
MOBP69
MOBP81B            TGACCCACGTGAAGGCTGGGTAGGCCTGGCAGTGGACCTCTAAATCCAAACTCAGAAAGTGGAG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       CTTTGGGACTGAGTGGGAAACCCTTCTTCTGTGACTATGGTAGAAGATATAGGACTGACTCCCA
Gen contig 3+4

                     5250       5260       5270       5280       5290       5300       5310
                      |          |          |          |          |          |          |
Contig# 2          ATGTCTAACTCTGGGACTCCATGCACATGCACCCCTGTATGCTTGTGCACATACACTTGCACAC
Gen contig 1
MOBP69
MOBP81B            ACAGGAGATGCCAGAGCAAGTCAAGAGGCCATATCCATGAACTCTGGGACTGATGGAGAAACCC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       ATGTCTAACTCTGGGACTCCATGCACATGCACCCCTGTATGCTTGTGCACATACACTTGCACAC
Gen contig 3+4

                          5320       5330       5340       5350       5360       5370
                           |          |          |          |          |          |
Contig# 2          ACACACACACACACACACACACACACACACACCACACCACACCACACCACACACCACACATC
Gen contig 1
MOBP69
MOBP81B            TTCCTTCCTCCAGTGAATATGATTAGAAGACATAGGCTGACTCCCAGTGTCAAACTCAGGGACT
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig 2
Gen contig2b       ACACACACACACACACACACACACACACACACCACACCACACCACACCACACACCACACATC
Gen contig 3+4

                        5380       5390       5400       5410       5420       5430       5440
                         |          |          |          |          |          |          |
Contig# 2          CACATGGAAATGGGGAAAAAGGAAACAAGAGAGTACCAGACTCATGGTTCTTCGAGGATGGGTG
Gen contig 1
MOBP69
MOBP81B            CCATGCACACACACCCCTGCATGCTCCTGCACATACACTTGCAAACATGAGCACACACACTTCA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig 2
Gen contig2b       CACATGGAAATGGGGAAAAAGGAAACAAGAGAGTACCAGACTCATGGTTCTTCGAGGATGGGTG
Gen contig 3+4
```

```
                          5450       5460       5470       5480       5490       5500
                           |          |          |          |          |          |
Contig# 2          CATCCTAGCCAGGACTTAGAAGCCTTTCCCCACCCAGGAAGACCTCACCATTTCTTCCTCCTGT
Gen contig 1
MOBP69
MOBP81B            TACATCCACATGGAAATGGGGGTAAGAGAAACGAGAGAGTACCTGGCTCATGGTTCCTCTAGGA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       CATCCTAGCCAGGACTTAGAAGCCTTTCCCCACCCAGGAAGACCTCACCATTTCTTCCTCCTGT
Gen contig 3+4

                     5510       5520       5530       5540       5550       5560
                      |          |          |          |          |          |
Contig# 2          TCCCTCTCTTGTACGTGTGAGCTGCTAATGGGGTGCTTCATTCACCATGGGGCTGTGTGCATGC
Gen contig 1
MOBP69
MOBP81B            TGGGCCCACCCTGCCCAGGACTTAGAAGCCCTTTCCACACCTAGCAGGACCACACGATTTCTTC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       TCCCTCTCTTGTACGTGTGAGCTGCTAATGGGGTGCTTCATTCACCATGGGGCTGTGTGCATGC
Gen contig 3+4

                  5570       5580       5590       5600       5610       5620       5630
                   |          |          |          |          |          |          |
Contig# 2          TCCTCAG|AGCCTGTCACGCTCTGGCCTACGGCTCCTACACAGGTGCCCTGAGCGCAGATATCTT
Gen contig 1
MOBP69
MOBP81B            TGCCTGTTCTCTCTCTTGTACGTGTGAACAAACGGGATGTGTCATCCACCATGGGGCTGTGTGC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       TCCTCAG|AGCCTGTCACGCTCTGGCCTACGGCTCCTACACAGGTGCCCTGAGCGCAGATATCTT
Gen contig 3+4

                       5640       5650       5660       5670       5680       5690
                        |          |          |          |          |          |
Contig# 2          CTCAGAGATTCCTCCACGGTCACCAGACTCCATCATTGTCTTCCAGGATCATCTGCTCACTTAT
Gen contig 1
MOBP69                    |GGCCTCTTGTGCTCTGACCTCCCTGTCCTACATGGATGCCCTCGGCTCAAATAT
MOBP81B            GCACTCCTAG|GGCCTCTTGTGCTCTGACCTCCCTGTCCTACATGGATGCCCTCGGCTCAAATAT
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b       CTCAGAGATTCCTCCACGGTCACCAGACTCCATCATTGTCTTCCAGGATCATCTGCTCACTTAT
Gen contig 3+4

                  5700       5710       5720       5730       5740       5750       5760
                   |          |          |          |          |          |          |
Contig# 2
Gen contig 1
MOBP69             CACCCCAGAGATTCTTCCACGGTCACCGACTCCGTCATTCTAGTTTGTCTTCAGGAATCATCAC
MOBP81B            CACCCCAGAGATTCTTCCACGGTCACCGACTCCGTCATTCTAGTTTGTCTTCAGGAATCATCAC
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4
```

```
                          5770       5780       5790       5800       5810       5820
                            |          |          |          |          |          |
Contig# 2
Gen contig 1
MOBP69          TCACTTGTTTAATTATTGTTTAATTCTCGCTCCAATGCTTTGCTTAATGTAAATCCCATGGCTG
MOBP81B         TCACTTGTTTAATTATTGTTTAATTCTCGCTCCAATGCTTTGCTTAATGTAAATCCCATGGCTG
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                    5830       5840       5850       5860       5870       5880
                      |          |          |          |          |          |
Contig# 2
Gen contig 1
MOBP69          CCCCTTTGATTT
MOBP81B         CCCCTTTGATTTCTCCTGTCTTCCCTTGTCTGGCACTAACAAGACGTCAATAAATTCTTGCCGA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4

                5890       5900       5910       5920       5930       5940       5950
                  |          |          |          |          |          |          |
Contig# 2                    TGGATCTCCTTTGGGAAACTTTTTTTTGGTTTATTCTAGGAATAGGTTTT
Gen contig 1
MOBP69
MOBP81B         CTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4                TGGATCTCCTTTGGGAAACTTTTTTTTGGTTTATTCTAGGAATAGGTTTT

                     5960       5970       5980       5990       6000       6010
                       |          |          |          |          |          |
Contig# 2       TTTTGGGAGGTAGAAAAAGATTCAGAACAACAAACAAGGTACATTCCGCAGGGCGGGAGAGGGG
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4  TTTTGGGAGGTAGAAAAAGATTCAGAACAACAAACAAGGTACATTCCGCAGGGCGGGAGAGGGG

                6020       6030       6040       6050       6060       6070       6080
                  |          |          |          |          |          |          |
Contig# 2       CCCTTCCAGCTTACTGACCGGAGTTGTTGGTTGAGCTGGTTTCAGGGGTGTGGCGTTCCATGGG
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4  CCCTTCCAGCTTACTGACCGGAGTTGTTGGTTGAGCTGGTTTCAGGGGTGTGGCGTTCCATGGG
```

```
                          6090       6100       6110       6120       6130       6140
                            |          |          |          |          |          |
Contig# 2                 TCTTGCCCCTCTCTGCCCTCCCTATCCTCAGGGCCAAGTGTGTGACCAGAGCTCGGTGTGTGCT
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4            TCTTGCCCCTCTCTGCCCTCCCTATCCTCAGGGCCAAGTGTGTGACCAGAGCTCGGTGTGTGCT

                        6150       6160       6170       6180       6190       6200
                          |          |          |          |          |          |
Contig# 2                 TCCCTTCACAGATATGTCTTCTTCTTCATGAGAGCGGGTACCTGCTTCTGAAGGATGTAGGAGG
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4            TCCCTTCACAGATATGTCTTCTTCTTCATGAGAGCGGGTACCTGCTTCTGAAGGATGTAGGAGG

                    6210       6220       6230       6240       6250       6260       6270
                      |          |          |          |          |          |          |
Contig# 2                 CCAACGGGAACGCAGGGGTCCAACCGCCAGGAGAACCAGGGGAGGAACAGCCCCATTTAGTTCT
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4            CCAACGGGAACGCAGGGGTCCAACCGCCAGGAGAACCAGGGGAGGAACAGCCCCATTTAGTTCT

                         6280       6290       6300       6310       6320       6330
                           |          |          |          |          |          |
Contig# 2                 GATGTTTTTTCTCTCCCCTCAG|CATGTGGACGATGCCATTCCATCACCTCCCTCTATAACTGCC
Gen contig 1
MOBP69
MOBP81B
MOBP81A                            |CACTTGGACGATGCCACTCCATCACCTCCCTCTATAACCGCC
MOBP99
rOPRP1
rOP1                               |CACTTGGACGATGCCACTCCATCACCTCCCTCTATAACCGCC
Gen contig2a
Gen contig2b
Gen contig 3+4            GATGTTTTTTCTCTCCCCTCAG|CATGTGGACGATGCCATTCCATCACCTCCCTCTATAACTGCC

                        6340       6350       6360       6370       6380       6390       6400
                          |          |          |          |          |          |          |
Contig# 2                 AGCGAGATATCTCTGTCAACTGTCTCGACTCTCCGAGGTAAGATTCTTTTTCACTTTCTGTTAT
Gen contig 1
MOBP69
MOBP81B
MOBP81A                   GAAGAGGCATCTCTGTCCACCTTCTGACCCTCCATGGTTGTTCAGTCTGAAGATTCCTGTGTCT
MOBP99
rOPRP1
rOP1                      GAAGAGGCATCTCTGTCCACCTTCTGACCCTCCATGGTTGTTCAGTCTGAAGATTCCTGTGTCT
Gen contig2a
Gen contig2b
Gen contig 3+4            AGCGAGATATCTCTGTCAACTGTCTCGACTCTCCGAGGTAAGATTCTTTTTCACTTTCTGTTAT
```

```
                          6410       6420       6430       6440       6450       6460
                           |          |          |          |          |          |
Contig# 2          TTTTTAACAAGGTTTTTCCTCTCTGTCACTGACATTCCAAATGCTGGCTGTGATGTCACTAAGT
Gen contig 1
MOBP69
MOBP81B
MOBP81A            TCAGGGATGTGCAAAACAAGGTCATGGTGTGCGACCATCCAAATGGCAAATATAAAAAAGAAAT
MOBP99
rOPRP1
rOP1               TCAGGGATGTGCAAAACAAGGTCATGGTGTGCGACCATCCAAATGGCAAATATAAAAAAGAAAT
Gen contig2a
Gen contig2b
Gen contig 3+4     TTTTTAACAAGGTTTTTCCTCTCTGTCACTGACATTCCAAATGCTGGCTGTGATGTCACTAAGT

                     6470       6480       6490       6500       6510       6520
                      |          |          |          |          |          |
Contig# 2          ACAGATTATATGTGCTTGTTCCGCCCAGGATTCGGGTTCTTCAGGCTGAAGATTCTTGTGTCTT
Gen contig 1
MOBP69
MOBP81B
MOBP81A            TGTCAAGTGTTGGCGGTGACGCAAGAAAAATCCTAGAAGCCACAGATGGGGATTTATACACTGT
MOBP99
rOPRP1
rOP1               TGTCAAGTGTTGGCGGTGACGCAAGAAAAATCCTAGAAGCCACAGATGGGGATTTATACACTGT
Gen contig2a
Gen contig2b
Gen contig 3+4     ACAGATTATATGTGCTTGTTCCGCCCAGGATTCGGGTTCTTCAGGCTGAAGATTCTTGTGTCTT

                  6530       6540       6550       6560       6570       6580       6590
                   |          |          |          |          |          |          |
Contig# 2          CTCGGATGTGCAAAACAAGGTCATGGTGCACAACCATGCAAACCGCAAATGCAAAAGAGAAATT
Gen contig 1
MOBP69
MOBP81B
MOBP81A            CGCAACACCGTTGAGATGTACTCTGGGGACACTATCAAGGCAAAGTATCTTAAAACCTCTAATC
MOBP99
rOPRP1
rOP1               CGCAACACCGTTGAGATGTACTCTGGGGACACTATCAAGGCAAAGTATCTTAAAACCTCTAATC
Gen contig2a
Gen contig2b
Gen contig 3+4     CTCGGATGTGCAAAACAAGGTCATGGTGCACAACCATGCAAACCGCAAATGCAAAAGAGAAATT

                      6600       6610       6620       6630       6640       6650
                       |          |          |          |          |          |
Contig# 2          GTCAAGGTTGGCGGTGACGTGAGAAAAATCCTAAGCAGCTACCGATGGGAGTCTAACATTGTCA
Gen contig 1
MOBP69
MOBP81B
MOBP81A            AAGCGACTGTAATTTTAGGCATCCATCCTGGGGAAAACTCACACCCTTTTAGAGAAAGGAGAAA
MOBP99
rOPRP1
rOP1               AAGCGACTGTAATTTTAGGCATCCATCCTGGGGAAAACTCACACCCTTTTAGAGAAAGGAGAAA
Gen contig2a
Gen contig2b
Gen contig 3+4     GTCAAGGTTGGCGGTGACGTGAGAAAAATCCTAAGCAGCTACCGATGGGAGTCTAACATTGTCA

                   6660       6670       6680       6690       6700       6710       6720
                    |          |          |          |          |          |          |
Contig# 2          CCACTGAGCTGTACTCTGGGGACACTGTCAAAGGCAAAATATCTTAAAACCTCTAATCAAGTGA
Gen contig 1
MOBP69
MOBP81B
MOBP81A            TGCCCAAATTTGTTTAAAAGATCGTGGTTTGTAGTTAAGAGAGGTAGAATGAGATTAACCAAAA
MOBP99
rOPRP1
rOP1               TGCCCAAATTTGTTTAAAAGATCGTGGTTTGTAGTTAAGAGAGGTAGAATGAGATTAACCAAAA
Gen contig2a
Gen contig2b
Gen contig 3+4     CCACTGAGCTGTACTCTGGGGACACTGTCAAAGGCAAAATATCTTAAAACCTCTAATCAAGTGA
```

```
                          6730       6740       6750       6760       6770       6780
                            |          |          |          |          |          |
Contig# 2          CTGCAATTTTAGGAATCCATCCTGGAGAAAACTCACACCCCTTTAGAGAAGGAGAAATGCGCAG
Gen contig 1
MOBP69
MOBP81B
MOBP81A            TGTCGGTCCACGGGAGAGCAAGAGAGTCAATCTTGGTATGTCCACAAAGTCCTAAGAGCAGTTA
MOBP99
rOPRP1
rOP1               TGTCGGTCCACGGGAGAGCAAGAGAGTCAATCTTGGTATGTCCACAAAGTCCTAAGAGCAGTTA
Gen contig2a
Gen contig2b
Gen contig 3+4     CTGCAATTTTAGGAATCCATCCTGGAGAAAACTCACACCCCTTTAGAGAAGGAGAAATGCGCAG

                     6790       6800       6810       6820       6830       6840
                       |          |          |          |          |          |
Contig# 2          ATTTGTCTAAAAGATCGTGGTTTGTAGTTAAGAGAAGTGGAATGAAACTAGCCAAAATGTCTGT
Gen contig 1
MOBP69
MOBP81B
MOBP81A            AAAGCCAATCTGTCAGACAGACCCATGCCGAGCAGGTAAATCTAGAAAGCAAACCAAGGACTAT
MOBP99
rOPRP1
rOP1               AAAGCCAATCTGTCAGACAGACCCATGCCGAGCAGGTAAATCTAGAAAGCAAACCAAGGACTAT
Gen contig2a
Gen contig2b
Gen contig 3+4     ATTTGTCTAAAAGATCGTGGTTTGTAGTTAAGAGAAGTGGAATGAAACTAGCCAAAATGTCTGT

                   6850       6860       6870       6880       6890       6900       6910
                     |          |          |          |          |          |          |
Contig# 2          CCAAGGGAATGAGAGAATCAATTTTGGGACGTTCACAAATAAAGTCCTAACAGCGGTTAAAAGC
Gen contig 1
MOBP69
MOBP81B
MOBP81A            AAAAGTCAAGCACCAGGCAGTAAATCCAG|AATAAGAGCCTATGCATACATATTGACAGCTCACA
MOBP99                                         |AATAAGAGCCTATGCATACATATTGACAGCTCACA
rOPRP1
rOP1               AAAAGTCAAGCACCAGGCAGTAAATCCAG|AATAAGAGCCTATGCATACATATTGACAGCT ACA
Gen contig2a
Gen contig2b
Gen contig 3+4     CCAAGGGAATGAGAGAATCAATTTTGGGACGTTCACAAATAAAGTCCTAACAGCGGTTAAAAGC

                       6920       6930       6940       6950       6960       6970
                         |          |          |          |          |          |
Contig# 2          CAATCTGTCAGATAGACCATGTCATGTAGGTAAATCTAGAAAGCAAACTGAGAACTATAAAAAT
Big contig 1
MOBP69
MOBP81B
MOBP81A            ATCCTGTACTGGTTATTGCTTACATAAGGAGAAGCCACAGATCTCATTCTTTCTTCTAGAATGC
MOBP99             ATCCTGTACTGGTTATTGCTTACATAAGGAGAAGCCACAGATCTCATTCTTTCTTCTAGAATGC
rOPRP1
rOP1               ATCCTGTACTGGTTATTGCTTACATAAGGAGAAGCCACAGATCT ATTCTTTCTTCTAGAATGC
Gen contig2a
Gen contig2b
Gen contig 3+4     CAATCTGTCAGATAGACCATGTCATGTAGGTAAATCTAGAAAGCAAACTGAGAACTATAAAAAT

                   6980       6990       7000       7010       7020       7030       7040
                     |          |          |          |          |          |          |
Contig# 2          CAAGTACCGGGCAGTAAATGCAG|TATGACGTTCTGGGCCTGAATATTGACATCCCACAGTCCTG
Gen contig 1
MOBP69
MOBP81B
MOBP81A            AATAGAATTTAAAAATGACAGCAGAACAAAAGGACACTAGAAATTATCCTCTCAGAAATAAAAG
MOBP99             AATAGAATTTAAAAATGACAGCAGAACAAAAGGACACTAGAAATTATCCTCTCAGAAATAAAAG
rOPRP1
rOP1               AATAGAATTTAAAAATGACAGCAGAACAAAAGGACACTAGAAATTATCCTCTCAGAAATAAAAG
Gen contig2a
Gen contig2b
Gen contig 3+4     CAAGTACCGGGCAGTAAATGCAG|TATGACGTTCTGGGCCTGAATATTGACATCCCACAGTCCTG
```

```
                          7050       7060       7070       7080       7090       7100
                            |          |          |          |          |          |
Contig# 2          TATTGGTTATTGCTTAATAAGGAGAAGCCACAAGATCTCACACTTTTTCTATAATGCAATAGGA
Big contig 1
MOBP69
MOBP81B
MOBP81A            TATAAGAAATAATTTCTGGCTTGAGGAAATAACCCAGGCTAAAAGCTACACAGGCCTGGCTTAA
MOBP99             TATAAGAAATAATTTCTGGCTTGAGGAAATAACCCAGGCTAAAAGCTACACAGGCCTGGCTTAA
rOPRP1
rOP1               TATAAGAAATAATTTCTGGCTTGAGGAAATAACCCAGGCTAAAAGCTACACAGGCCTGGCTTAA
Gen contig2a
Gen contig2b
Gen contig 3+4     TATTGGTTATTGCTTAATAAGGAGAAGCCACAAGATCTCACACTTTTTCTATAATGCAATAGGA

                     7110       7120       7130       7140       7150       7160
                       |          |          |          |          |          |
Contig# 2          TTTAAAAATGACAGCAGAACACAAAGACACTAGAAATTGTCCTGTCAGAAATAAAAGTATAAGA
Gen contig 1
MOBP69
MOBP81B
MOBP81A            CCATACCAGCAGGGCATACCTTTGTTAGGAGAGGGTCAATTCGAAGTTTGTGAACAGCCTGGGC
MOBP99             CCATACCAGCAGGGCATACCTTTGTTAGGAGAGGGTCAATTCGAAGTTTGTGAACAGCCTGGGC
rOPRP1
rOP1               CCATACCAGCAGGGCATACCTTTGTTAGGAGAGGGTCAATTCGAAGTTTGTGAACAGCCTGGGC
Gen contig2a
Gen contig2b
Gen contig 3+4     TTTAAAAATGACAGCAGAACACAAAGACACTAGAAATTGTCCTGTCAGAAATAAAAGTATAAGA

                   7170       7180       7190       7200       7210       7220       7230
                     |          |          |          |          |          |          |
Contig# 2          AATAATTCCTGGCTTGAGGAAATAACCCAGACTAAAAGCTACATAGAAACGGCTTAACAGTACT
Gen contig 1
MOBP69
MOBP81B
MOBP81A            TGTTTCCGAGGCTGCATAAATCTTAAACACCAAAAATGTGTGTTTTGAATTAACTTTGGGTCTG
MOBP99             TGTTTCCGAGGCTGCATAAATCTTAAACACCAAAAATGTGTGTTTTGAATTAACTTTGGGTCTG
rOPRP1
rOP1               TGTTTCCGAGGCTGCATAAATCTTAAACACCAAAAATGTGTGTTTTGAATTAACTTTGGGTCTG
Gen contig2a
Gen contig2b
Gen contig 3+4     AATAATTCCTGGCTTGAGGAAATAACCCAGACTAAAAGCTACATAGAAACGGCTTAACAGTACT

                        7240       7250       7260       7270       7280       7290
                          |          |          |          |          |          |
Contig# 2          GGCAGGGCATATATTTGTTTGGAGAGGGTCGATGTGAAGTCTGTGAATAGTCTGGGATGTTTAC
Gen contig 1
MOBP69
MOBP81B
MOBP81A            GTCTACAGAGTCCGGCTTGTGATGGCTCCTGACACGGGAGAAAAGGGGATAAGTTAATATAAGG
MOBP99             GTCTACAGAGTCCGGCTTGTGATGGCTCCTGACACGGGAGAAAAGGGGATAAGTTAATATAAGG
rOPRP1
rOP1               GTCTACAGAGTCCGGCTTGTGATGGCTCCTGACACGGGAGAAAAGGGGATAAGTTAATATAAGG
Gen contig2a
Gen contig2b
Gen contig 3+4     GGCAGGGCATATATTTGTTTGGAGAGGGTCGATGTGAAGTCTGTGAATAGTCTGGGATGTTTAC

                   7300       7310       7320       7330       7340       7350       7360
                     |          |          |          |          |          |          |
Contig# 2          AAGGTCGCATAAATTTTAAACACCAAAAATATTTGTTTTGAATCAATTTTGGGTCTGGGCTATA
Gen contig 1
MOBP69
MOBP81B
MOBP81A            CTCTGGGATTTTCCTTTGAGGAACAGGCAGGTAGGAAAGGACACTAGAATGTTGGGGAATGTGG
MOBP99             CTCTGGGATTTTCCTTTGAGGAACAGGCAGGTAGGAAAGGACACTAGAATGTTGGGGAATGTGG
rOPRP1
rOP1               CTCTGGGATTTTCCTTTGAGGAACAGGCAGGTAGGAAAGGACACTAGAATGTTGGGGAATGTGG
Gen contig2a
Gen contig2b
Gen contig 3+4     AAGGTCGCATAAATTTTAAACACCAAAAATATTTGTTTTGAATCAATTTTGGGTCTGGGCTATA
```

```
                          7370       7380       7390       7400       7410       7420
                            |          |          |          |          |          |
Contig# 2                 GAGTCTGCCTTGTGATGGTTCCTGACACGAGAGAAAAGGGAAGAAATGTTAATACAGCTCTCGA
Gen contig 1
MOBP69
MOBP81B
MOBP81A                   AGGGATGTAAAACAGACAGTATTATGATGGGGGGAACTCAGGAAGGAACGGTAAGGACACCAGA
MOBP99                    AGGGATGTAAAACAGACAGTATTATGATGGGGGGAACTCAGGAAGGAACGGTAAGGACACCAGA
rOPRP1
rOP1                      AGGGATGTAAAACAGACAGTATTATGATGGGGGGAACTCAGGAAGGAACGGTAAGGACACCAGA
Gen contig2a
Gen contig2b
Gen contig 3+4            GAGTCTGCCTTGTGATGGTTCCTGACACGAGAGAAAAGGGAAGAAATGTTAATACAGCTCTCGA

                          7430       7440       7450       7460       7470       7480
                            |          |          |          |          |          |
Contig# 2                 ATTTTCCCTTTAAAGACCAGGTAGGTAGGAAAGGACAACAGGATGTTGGGGAATGTGGAGGGAT
Gen contig 1
MOBP69
MOBP81B
MOBP81A                   ATGGCCTCGTGCTGGGAGCCAGGGTGGCCAAGCCCAGGCAGATGCACAAAGACCCCAAAGGCCA
MOBP99                    ATGGCCTCGTGCTGGGAGCCAGGGTGGCCAAGCCCAGGCAGATGCACAAAGACCCCAAAGGCCA
rOPRP1
rOP1                      ATGGCCTCGTGCTGGGAGCCAGGGTGGCCAAGCCCAGGCAGATGCACAAAGACCCCAAAGGCCA
Gen contig2a
Gen contig2b
Gen contig 3+4            ATTTTCCCTTTAAAGACCAGGTAGGTAGGAAAGGACAACAGGATGTTGGGGAATGTGGAGGGAT

                        7490       7500       7510       7520       7530       7540       7550
                          |          |          |          |          |          |          |
Contig# 2                 GTAAAACAGACAGTATTACGGTGGCGAAAACTCAGTGTAGGGCGATAAGAACATCAGATGGCCT
Gen contig 1
 MOBP69
MOBP81B
MOBP81A                   CCATTGCCTCACAACCAGACCCAGCAAAGTTCAGATCAAGTGCATAAATGGGGGTCACGTCAGT
MOBP99                    CCATTGCCTCACAACCAGACCCAGCAAAGTTCAGATCAAGTGCATAAATGGGGGTCACGTCAGT
rOPRP1
rOP1                      CCATTGCCTCACAACCAGACCCAGCAAAGTTCAGATCAAGTGCATAAATGGGGGTCACGTCAGT
Gen contig2a
Gen contig2b
Gen contig 3+4            GTAAAACAGACAGTATTACGGTGGCGAAAACTCAGTGTAGGGCGATAAGAACATCAGATGGCCT

                          7560       7570       7580       7590       7600       7610
                            |          |          |          |          |          |
Contig# 2                 TGTGCTGGGAGCCGGGGTGGCCAAGCACAGGCGGATGCTTGAAGACCCCAAAGGTCACCATGGC
Gen contig 1
MOBP69
MOBP81B
MOBP81A                   GGAAACTGTCCGGAGCATGGAGCAGGGGGTGAGGAAGACAGTCAGAAGGAAGCAAGAGATGGCT
MOBP99                    GGAAACTGTCCGGAGCATGGAGCAGGGGGTGAGGAAGACAGTCAGAAGGAAGCAAGAGATGGCT
rOPRP1
rOP1                      GGAAACTGTCCGGAGCATGGAGCAGGGGGTGAGGAAGACAGTCAGAAGGAAGCAAGAGATGGCT
Gen contig2a
Gen contig2b
Gen contig 3+4            TGTGCTGGGAGCCGGGGTGGCCAAGCACAGGCGGATGCTTGAAGACCCCAAAGGTCACCATGGC

                        7620       7630       7640       7650       7660       7670       7680
                          |          |          |          |          |          |          |
Contig# 2                 GTCCTCACAGCCAGACCCGGCAAAGTTCAGATCAAGTGCACGGATGGGGGCCATGTTAGTGGAA
Gen contig 1
MOBP69
MOBP81B
MOBP81A                   TAGAGTGTTGCATGGCAGCAAGAAGAAGCAGTCCAAATATCATCCCAAGGCAGCCATGGCTTCC
MOBP99                    TAGAGTGTTGCATGGCAGCAAGAAGAAGCAGTCCAAATATCATCCCAAGGCAGCCATGGCTTCC
rOPRP1
rOP1                      TAGAGTGTTGCATGGCAGCAAGAAGAAGCAGTCCAAATATCATCCCAAGGCAGCCATGGCTTCC
Gen contig2a
Gen contig2b
Gen contig 3+4            GTCCTCACAGCCAGACCCGGCAAAGTTCAGATCAAGTGCACGGATGGGGGCCATGTTAGTGGAA
```

```
                          7690       7700       7710       7720       7730       7740
                            |          |          |          |          |          |
Contig# 2          ACTGTCCAAGGTGTGCGGAGGGGGTGAGGAAGACAGCCAGGAGGAAGAGATGGCTCAGAATGCT
Gen contig 1
MOBP69
MOBP81B
MOBP81A            TCTCCCGAGGACACTGTGGCCCTGTGGAAGTGCCAAGGCAGAGGCAGAGGAATTTAACCGCACA
MOBP99             TCTCCCGAGGACACTGTGGCCCTGTGGAAGTGCCAAGGCAGAGGCAGAGGAATTTAACCGCACA
rOPRP1
rOP1               TCTCCCGAGGACACTGTGGCCCTGTGGAAGTGCCAAGGCAGAGGCAGAGGAATTTAACCGCACA
Gen contig2a
Gen contig2b
Gen contig 3+4     ACTGTCCAAGGTGTGCGGAGGGGGTGAGGAAGACAGCCAGGAGGAAGAGATGGCTCAGAATGCT

                     7750       7760       7770       7780       7790       7800
                       |          |          |          |          |          |
Contig# 2          GCACTACAGCAAAGAAGAAGCAGGCCAAATATCGTCCTAAGGCAGCCGAGGGGTCCCGGCTTCC
Gen contig 1
MOBP69
MOBP81B
MOBP81A            CAAACTACATGCAGAAACTAAGGTGCTAAATGAACATCTACGTGACAACCTTTCTGAGGACCAG
MOBP99             CAAACTACATGCAGAAACTAAGGTGCTAAATGAACATCTACGTGACAACCTTTCTGAGGACCAG
rOPRP1
rOP1               CAAACTACATGCAGAAACTAAGGTGCTAAATGAACATCTACGTGACAACCTTTCTGAGGACCAG
Gen contig2a
Gen contig2b
Gen contig 3+4     GCACTACAGCAAAGAAGAAGCAGGCCAAATATCGTCCTAAGGCAGCCGAGGGGTCCCGGCTTCC

                7810       7820       7830       7840       7850       7860       7870
                  |          |          |          |          |          |          |
Contig# 2          TCTCCCGAGGGCACTGTGGCTCCGTGGAAGTGCCGAGGCAGAGGCAGGGGAATTTAACTGCACA
Gen contig 1
MOBP69
MOBP81B
MOBP81A            GTTCCAGTGAGCGGCCCAGGGCCGTTATCTTCGCGATGGTAACTGAGCTGAGAAGAATGGGAAC
MOBP99             GTTCCAGTGAGCGGCCCAGGGCCGTTATCTTCGCGATGGTAACTGAGCTGAGAAGAATGGGAAC
rOPRP1
rOP1               GTTCCAGTGAGCGGCCCAGGGCCGTTATCTTCGCGATGGTAACTGAGCTGAGAAGAATGGGAAC
Gen contig2a
Gen contig2b
Gen contig 3+4     TCTCCCGAGGGCACTGTGGCTCCGTGGAAGTGCCGAGGCAGAGGCAGGGGAATTTAACTGCACA

                     7880       7890       7900       7910       7920       7930
                       |          |          |          |          |          |
Contig# 2          AAAACTACACGCAGAAACTCAGGTGCTAAAATGGACATGTATGTGACCACCTTTCTGAGGACCA
Gen contig 1
MOBP69
MOBP81B
MOBP81A            CGGATAGGAGCACACAGTAGCCCACTGGTGAGACAAATCCACGGGCGAGCCTCAACCAACCCAC
MOBP99             CGGATAGGAGCACACAGTAGCCCACTGGTGAGACAAATCCACGGGCGAGCCTCAACCAACCCAC
rOPRP1
rOP1               CGGATAGGAGCACACAGTAGCCCACTGGTGAGACAAATCCACGGGCGAGCCTCAACCAACCCAC
Gen contig2a
Gen contig2b
Gen contig 3+4     AAAACTACACGCAGAAACTCAGGTGCTAAAATGGACATGTATGTGACCACCTTTCTGAGGACCA

                   7940       7950       7960       7970       7980       7990       8000
                     |          |          |          |          |          |          |
Contig# 2          GTTCCAGTGAGTGCCCAGGATGTTTCCTTCTCGACTACGGTTACTGAGCTTAGAAGAATGGGAA
Gen contig 1
MOBP69
MOBP81B
MOBP81A            TAAGGAAACTTCACGCCTTTTCATCTAGTTTTCATTTTGGCAAAGCAAAGCCATCCTGAGTGCT
MOBP99             TAAGGAAACTTCACGCCTTTTCATCTAGTTTTCATTTTGGCAAAGCAAAGCCATCCTGAGTGCT
rOPRP1
rOP1               TAAGGAAACTTCACGCCTTTTCATCTAGTTTTCATTTTGGCAAAGCAAAGCCATCCTGAGTGCT
Gen contig2a
Gen contig2b
Gen contig 3+4     GTTCCAGTGAGTGCCCAGGATGTTTCCTTCTCGACTACGGTTACTGAGCTTAGAAGAATGGGAA
```

```
                        8010       8020       8030       8040       8050       8060
                         |          |          |          |          |          |
Contig# 2               CGGGTTTAGGAGCACACAGTTAGCCCCATTGGTGAGACAAATTTCCAGGCTAGCTTCATCCAAT
Gen contig 1
MOBP69
MOBP81B
MOBP81A                 TGCTTGTCTCCGCCCTCCACGCCACCCCCGAGACACAGAGCATGCGCATTAACCCAGAGCGCGC
MOBP99                  TGCTTGTCTCCGCCCTCCACGCCACCCCCGAGACACAGAGCATGCGCATTAACCCAGAGCGCGC
rOPRP1
rOP1                    TGCTTGTCTCCGCCCTCCACGCCACCCCCGAGACACAGAGCATGCGCATTAACCCAGAGCGCGC
Gen contig2a
Gen contig2b
Gen contig 3+4          CGGGTTTAGGAGCACACAGTTAGCCCCATTGGTGAGACAAATTTCCAGGCTAGCTTCATCCAAT

                      8070       8080       8090       8100       8110       8120
                       |          |          |          |          |          |
Contig# 2               CCATTAAGGAGACTTTAGGCTTTTTCATCTGGTTTTCATTGGGGCAAAGCAAAGCATATCCTGA
Gen contig 1
MOBP69
MOBP81B
MOBP81A                 TAACACATGCACAGCGTTGGACATTAGCCTATTAATGGGAATAACCAGAAACTACTAAATTTTG
MOBP99                  TAACACATGCACAGCGTTGGACATTAGCCTATTAATGGGAATAACCAGAAACTACTAAATT
rOPRP1
rOP1                    TAACACATGCACAGCGTTGGACATTAGCCTATTAATGGGAATAACCAGAAACTACTAAATTTTG
Gen contig2a
Gen contig2b
Gen contig 3+4          CCATTAAGGAGACTTTAGGCTTTTTCATCTGGTTTTCATTGGGGCAAAGCAAAGCATATCCTGA

                    8130       8140       8150       8160       8170       8180       8190
                     |          |          |          |          |          |          |
Contig# 2               GTGTTGGTTTGTTTCCGCCCTCCACGCCACCCCTGAGACACCGAGCATGCGCATTAACCCAGAG
Gen contig 1
MOBP69
MOBP81B
MOBP81A                 TTTGTTGTAGAAATAAAACCTGGGGATGGATCAAAA
MOBP99
rOPRP1
rOP1                    TTTGTTGTAGAAATAAAACCTGGGGATGGATC
Gen contig2a
Gen contig2b
Gen contig 3+4          GTGTTGGTTTGTTTCCGCCCTCCACGCCACCCCTGAGACACCGAGCATGCGCATTAACCCAGAG

                        8200       8210       8220       8230       8240       8250
                         |          |          |          |          |          |
Contig# 2               CGCGCTAACCCATGCGCAGCGTTGGACGTTAGCTCATTAATGTGATTAACCAGAAACCACTAAA
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4          CGCGCTAACCCATGCGCAGCGTTGGACGTTAGCTCATTAATGTGATTAACCAGAAACCACTAAA

                    8260       8270       8280       8290       8300       8310       8320
                     |          |          |          |          |          |          |
Contig# 2               TTTTGTTTGTTGTAGAAATAAAACCTGGGGATGGATCAAACTATATCTGGAGATTGCTTGAAGC
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Big contig2a
Gen contig2b
Gen contig 3+4          TTTTGTTTGTTGTAGAAATAAAACCTGGGGATGGATCAAACTATATCTGGAGATTGCTTGAAGC
```

```
                      8330       8340       8350       8360       8370       8380
                       |          |          |          |          |          |
Contig# 2        TGATGGGTTACTAAGCCAAATGTGCTTTGGAATTCCTCACTGAGTCGCGTTGTAATCACCACGT
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4   TGATGGGTTACTAAGCCAAATGTGCTTTGGAATTCCTCACTGAGTCGCGTTGTAATCACCACGT

                    8390       8400       8410       8420       8430       8440
                       |          |          |          |          |          |
Contig# 2        ACTGTTTTTCTACTTAACATCGACTCTCAGGAACATTCCCAAGTTAATACCGGAGACATTTCAT
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Big contig2a
Gen contig2b
Gen contig 3+4   ACTGTTTTTCTACTTAACATCGACTCTCAGGAACATTCCCAAGTTAATACCGGAGACATTTCAT

                 8450       8460       8470       8480       8490       8500       8510
                    |          |          |          |          |          |          |
Contig# 2        AAAACTCCCAAAGCTGTTCTTCATACATTTCTTGAATTAGCAATCCAAGCCCCCATGTTCTGAC
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4   AAAACTCCCAAAGCTGTTCTTCATACATTTCTTGAATTAGCAATCCAAGCCCCCATGTTCTGAC

                       8520       8530       8540       8550       8560
                          |          |          |          |          |
Contig# 2        ATTATTCCCGAATGTTTAATTAATCCTTACATTTTGGGGAAAAAACGGAAGGC
Gen contig 1
MOBP69
MOBP81B
MOBP81A
MOBP99
rOPRP1
rOP1
Gen contig2a
Gen contig2b
Gen contig 3+4   ATTATTCCCGAATGTTTAATTAATCCTTACATTTTGGGGAAAAAACGGAAGGC
```

Appendix X Sequence multialignment of contigs 1, 2 and 3, and published sequence data. Sequences were imported to the Genejockey software (Cambridge, Biosoft), from the database according to their EMBL accession numbers (defined in Table 2.3.1 and Section 2, 2.5.1) and aligned by eye within the Genejockey.

```
                                10        20        30        40        50        60
                                 |         |         |         |         |         |
OC II 15inv 5'UTR  AAGCTCCACCAGGGCCCGGTATCCACAGGAACCTTTCACAACAGCCATTACTCGCAGGGCA
MOBP69 5'                              TATCCACAGGAACCTTTCACAGCAGCCAATACCTGCAGGGCA
MOBP81 B 5'                                  AGGGAACCTTTCACAGCAGCCAATACCTGCAGGGCA
MOBP81A 5'                                       CTTTCACAGCAGCCAATACCTGCAGGGCA
MOBP99 5'                                                GCCAATACCTGCAGGGCA
rOP1 5'
rOPRP1 5'

                               70        80        90       100       110       120
                                |         |         |         |         |         |
OC II 15inv 5'UTR  ACAAAGAATCAAATGAGAGCAAGACAAGCCGAGAACAGAGTCCTGGTTGCCAGACCGGCAC
MOBP69 5'          ACAAAGAATCAAATGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
MOBP81 B 5'        ACAAAGAATCAAATGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
MOBP81A 5'         ACAAAGAATCAAATGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
MOBP99 5'          ACAAAGAATCAAATGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
rOP1 5'                      TGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/
rOPRP1 5'                    TGAGAGCGAGACAAGCTGGGAATGAAGTCCTGGTTGCCAGA/TGGGA/

                          130
                           |
OC II 15inv 5'UTR  GGATGAAAACCCAG|***
MOBP69 5'          GCTTGAAAACACAG|***
MOBP81 B 5'        GCTTGAAAACACAG|***
MOBP81A 5'         GCTTGAAAACACAG|***
MOBP99 5'          GCTTGAAAACACAG|***
rOP1 5'            GCTTGAAAACACAG|***
rOPRP1 5'          GCTTGAAAACACAG|***
```

Appendix XI Sequence multi-alignment of 5' UTR sequences from published (Yamamoto *et al.*, 1994; Holz *et al.*, 1996) rat splice variants and sequence from a 5' cDNA fragment generated in an oligo-capping assay to identify transcriptional start sites (Section 2, 2.5). 5' donor splice sites are indicated by |***. N.B. Nucleotide numbers, listed above sequence, are purely arbitrary and bear no relevance to the genomic organisation of the gene. /, introduced into 5'UTR sequence of rat splice variants to maintain the alignment of rat and mouse sequences. These do not represent unidentified nucleotides.

Publications

McCallion, A. S., Guenet, J. L., Montague, P., Griffiths, I. R., Savioz, A., and Davies, R. W. (1996). The Mouse Gene (*Mobp*) Encoding Myelin-Associated Oligodendrocytic Basic-Protein Maps to Distal Chromosome-9. *Mammalian Genome* **7**, 847-849.

McCallion, A. S., Guenet, J. L., Stewart, G. J., and Davies, R. W. (1996). Genetic mapping of 14 ESTs expressed in the mouse ventral midbrain. Submitted *Mammalian Genome.*

Montague, P., Dickinson, P. J., McCallion, A. S., Stewart, G. J., Savioz, A., Davies, R. W., Kennedy, P. G. E., and Griffiths, I. R. (1997). Developmental expression of the murine *Mobp* gene. *J. Neurosci. Res.* **49**, 133-143.

Montague, P., Kirkham, D., McCallion, A. S., Davies, R. W., Kennedy, P. G. E., Klugmann, M., Groome, N. P., and Griffiths, I. R. (1997). Expression of the *Mobp* gene in the *shiverer* mouse. *Submitted Developmental Neurosci.*

Montague, P., Barrie, J.E., Thomson, C.E., McCallion, A. S., Davies, R. W., Kennedy, P. G. E., and Griffiths, I. R. (1997). Cytoskeletal and nuclear localisation of MOBP polypeptides. In press *European J. Neurosci.*

Duran Alonso, M.B., Sheils, P., McCallion, A.S., Bennet, N.K., Payne, A.P., Szpirer, J.A., Szpirer, C., Brodie, M.J., Davies, R.W. and Sutcliffe, R. G. The cystatin S gene maps to rat chromosome 3, to which *D1Mgh18* is re-assigned from chromosome 1. In press, *Mammalian Genome 8.*