# THE STUDY OF HPRT GENE EXPRESSION USING GENE TARGETING AND TRANSGENIC MICE

Simon Thompson

A Thesis Presented for the Degree of Ph.D

Department of Molecular Biology
University of Edinburgh
December 1989

**To My Mother**

## ABSTRACT

The mammalian HPRT gene is a housekeeping gene. It is expressed constitutively at a low level in all cell-types, with the exception of the brain where expression is elevated. It has been shown in human brains that HPRT expression is particularly high in the basal ganglia. Total HPRT deficiency in man is the causal defect in the Lesch-Nyhan syndrome, a severe and complex neurological disorder. Partial HPRT deficiency leads to gouty arthritis.

The expression of the mouse HPRT gene has been extensively studied in cultured cells and promoter elements essential for a basal level of transcription have been identified. The aim of the work presented in this thesis was to identify the nucleotide sequence elements responsible for the elevation of HPRT expression in brain. A line of mice transgenic for an HPRT minigene was produced. The HPRT minigene contains 635bp of HPRT 5' flanking sequence and the HPRT 5' and 3' untranslated regions. This minigene contains sufficient nucleotide sequence to be expressed at an elevated level in brain relative to other tissues. Gene targeting in ES cells was also used to manipulate the HPRT gene in the mouse germ line. An HPRT mutant gene corrected by gene targeting and also containing 635bp of 5' flanking sequence is expressed at elevated levels in brain. This experiment provided important information about gene targeting and its potential as a way of manipulating the mammalian genome to study the expression of endogenous genes.

## DECLARATION

The composition of this thesis and the work presented within it are my own, unless otherwise stated. The experiments presented were devised in collaboration with my supervisor, Dr. David W. Melton.

Simon Thompson December 1989

## ACKNOWLEDGEMENTS

I am most grateful to my supervisor, David Melton, for his excellent supervision and advice. I would like to thank Angus King, Carolanne McEwan, Angela Pow, Jim Selfridge and Nik Somia for technical help and discussion. I am grateful to Martin Hooper and Alan Clarke for a very profitable collaboration, to Paul Simons for making the transgenic mice experiment possible, and to the SERC for funding. I am grateful to many members of the department for advice and assistance, in particular, Noreen Murray and David Finnegan. Thanks to Graham Brown and Frank Johnston for photography, and to Hilary Lothian for typing this thesis.

I would especially like to thank Alex Gann and Jim Selfridge for their friendship, and my mother for her support, both moral and financial.

Most of all, thank you to Julia, for everything.

## LIST OF CONTENTS

## ABBREVIATIONS

| | |
|---|---|
| A | adenine |
| bp | base pair(s) |
| BSA | bovine serum albumin |
| C | cytosine |
| cDNA | DNA complementary to mRNA |
| CNS | central nervous system |
| dATP | deoxyadenosine triphosphate |
| dCTP | deoxycytidine triphosphate |
| dGTP | deoxyguanosine triphosphate |
| DNA | deoxyribonucleic acid |
| DTT | dithiothreitol |
| EDTA | ethylenediamine tetraacetic acid |
| ES cells | embryonic stem cells |
| G | guanine |
| GMP | guanine monophosphate |
| HAT | hypoxanthine, aminopterin, thymidine |
| HEPES | N'-2-hydroxyethyl piperazine-N'-2-ethanesulfonic acid |
| IMP | inosine monophosphate |
| kb | kilobase pairs |
| MOPS | 3-[N-Morpholino] propanesulfonic acid |
| mRNA | messenger RNA |
| OD | optical density |
| PBS | phosphate bufferred saline |
| RNA | ribonucleic acid |
| SDS | sodium dodecyl sulphate |
| SSC | standard saline citrate |
| ssDNA | single-stranded DNA |
| T | thymine |
| TEMED | N,N,N',N'-tetramethyl ethylene diamine |
| Tris | Tris-(hydroxymethyl)-methylamine |
| U | Uracil |

# 1. INTRODUCTION

## 1.1 HPRT enzyme function and tissue distribution

Hypoxanthine-guanine phosphoribosyltransferase (HPRT; IMP: pyrophosphate phosphoribosyltransferase; EC 2.4.2.8) is an enzyme of the purine salvage pathway. The role of HPRT in purine metabolism is illustrated in Figure 1.1. It catalyses the conversion of hypoxanthine and guanine to 5'-IMP and 5'-GMP, respectively (Kornberg et al., 1955; Korn et al., 1955). These products are precursors in DNA and RNA synthesis, and in the synthesis of cofactors for a range of enzymatic reactions. This activity is required in all mammalian cells and so the HPRT enzyme is said to be a 'housekeeping enzyme' and the gene a 'housekeeping gene'.

Mammals are also able to synthesise nucleotides de novo from simple precursors. After these nucleotides have been metabolized, the free purine bases can be reconverted directly to the respective nucleotides by phosphoribosyltransferase activity in the presence of 5'-phosphoribosyl - 1 - pyrophosphate (PRPP). HPRT is the key enzyme in the salvage of guanine and hypoxanthine. Adenine phosphoribosyltransferase (APRT) catalyses the salvage of adenine (Figure 1.1). In humans 90% of free purines are recycled in this way.

HPRT is a soluble, cytoplasmic enzyme which constitutes between 0.005 and 0.04% of total cell protein (Kelley and

Wyngaarden, 1983). The native enzyme is probably a tetramer. Each subunit is 217 amino acids long with a mass of 24,000 daltons (Wilson et al., 1982).

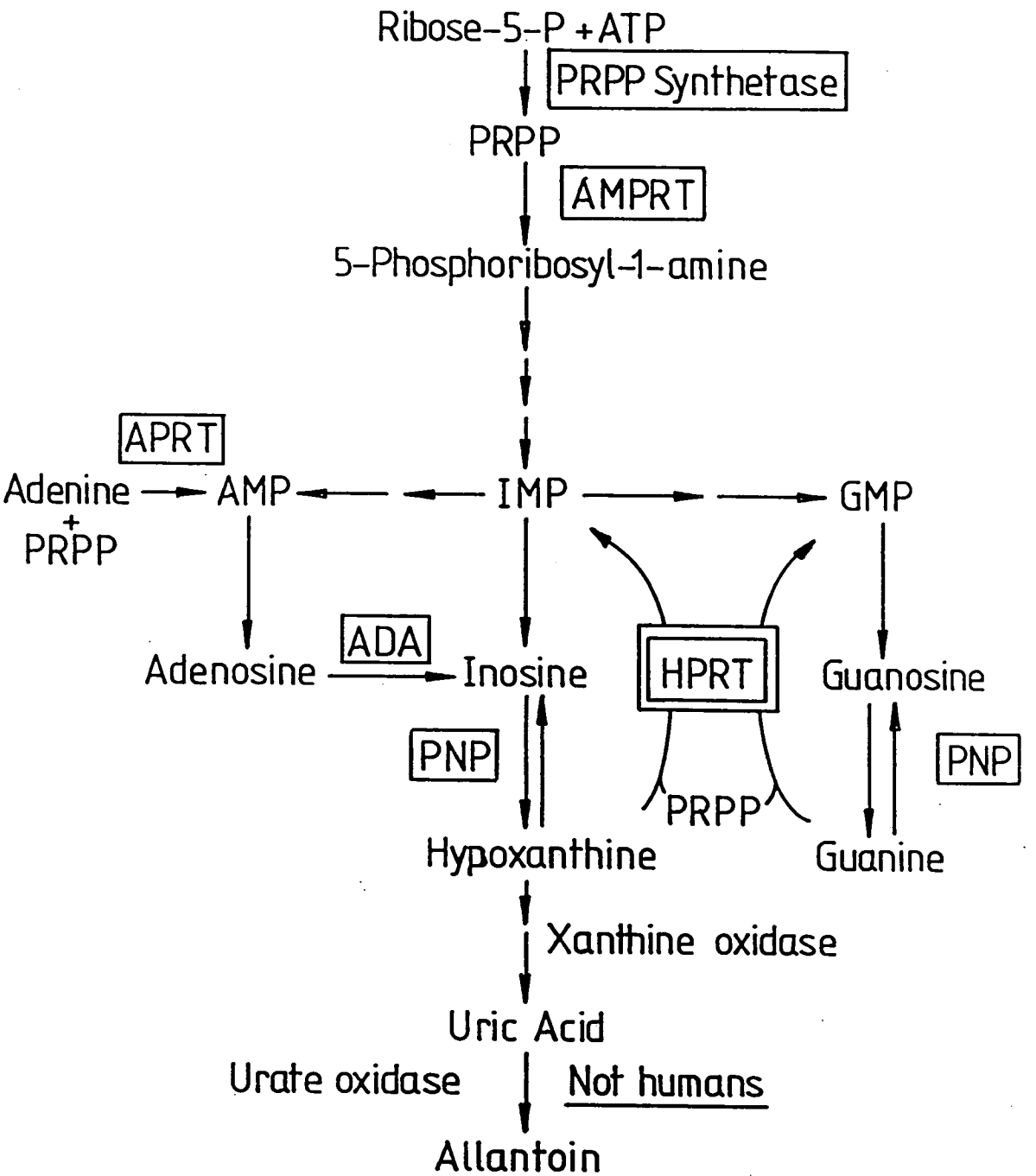
The tissue distribution of HPRT activity has been analysed in humans (Rosenbloom et al.,1967). While HPRT enzyme activity is detectable in all tissues, there is considerable variation in the level of activity between tissues. Enzyme assays on human autopsy material showed that many tissues, including liver, spleen, kidney and pancreas have low and approximately equivalent levels of HPRT activity. However, in brain tissue the level of activity is elevated. This elevation is particularly pronounced in the basal ganglia where the level of HPRT specific activity is approximately 20-fold higher than in liver tissue. The level of HPRT activity in each tissue is inversely proportional to the level of activity of amidophosphoribosyltransferase (AMPRT) enzyme, which catalyses the rate limiting step in the purine de novo synthesis pathway (Figure 1.1). Perhaps, therefore, the basal ganglia are particularly dependent upon purine salvage. This may be related to the neurological dysfunction associated with HPRT deficiency in humans, the Lesch-Nyhan syndrome.

Rats also demonstrate elevated HPRT enzyme activity in the central nervous system (CNS) (Gutensohn and Guroff, 1972). HPRT activity was assayed in the basal ganglia but was found to be no higher than in the brain as a whole. Gutensohn and Guroff (1972) also demonstrated that in rat brain there is a

**Figure 1.1.**

Purine Metabolism. PRPP,5'-phosphoribosyl-1-pyrophosphate; ATP, adenosine triphosphate; AMP, adenosine monophosphate; IMP, inosine monophosphate; GMP, guanosine monophosphate; AMPRT, amidophosphoribosyltransferase; APRT, adenine phosphoribosyltransferase; ADA, adenosine deaminase; PNP, purine nucleotide phosphorylase.

Figure 1.1


Ribose-5-P + ATP
PRPP Synthetase
PRPP
AMPRT
5-Phosphoribosyl-1-amine
APRT
Adenine
+
PRPP
AMP
IMP
GMP
Adenosine
ADA
Inosine
HPRT
Guanosine
PNP
PRPP
PNP
Hypoxanthine
Guanine
Xanthine oxidase
Uric Acid
Urate oxidase
Not humans
Allantoin

threefold increase in HPRT activity during the first 15 to 20 days of life. Similarly, enzyme assays on mouse tissues show that protein derived from brain has the highest specific activity of HPRT (Lo and Palmour, 1979). There are no reports of attempts to localise, more specifically, the elevated level of expression in mice.

There are no published studies of HPRT mRNA levels in tissues for any organism and consequently the mechanism which elevates HPRT expression in the brain has not been determined. Electrophoretic variants of the HPRT enzyme have been observed (Lo and Palmour, 1979; Gutensohn and Guroff, 1972). The basis of this variation has not been established and it has no apparent functional significance (Kelley and Wyngaarden, 1979). There is no correlation between the electrophoretic variants and levels of enzyme activity.

## 1.2 The Lesch-Nyhan syndrome

The Lesch-Nyhan syndrome, first described by Lesch and Nyhan in 1964, is the inherited neurological disorder associated with a complete deficiency of HPRT enzyme activity (Seegmiller et al., 1967). Because the HPRT gene is X-chromosome linked the syndrome almost exclusively affects males, with an estimated frequency of 1 in 10,000 (Kelley and Wyngaarden, 1983). The Lesch-Nyhan syndrome is reviewed extensively by Stout and Caskey (1989). A partial deficiency

of HPRT activity results in gouty arthritis.

Lesch-Nyhan patients generally appear normal at birth. However, by three to four months of age most patients exhibit delayed motor development. Between eight and twelve months of age more pronounced signs become evident. Initially, these may be slow involuntary and sometimes rhythmical movements of the limbs and hands, but later develop to include more sudden and violent uncontrollable actions. The most disturbing and curious feature of the Lesch-Nyhan syndrome is the compulsive self-injurious behaviour which presents in 85% of patients. The onset of self-mutilation can be as early as 1 year of age or as late as 16 years of age, and is characterised by biting of the fingers, lips and the inside of the cheeks. This may become so extreme that it is necessary to keep the elbows of a patient in extension with splints to protect the fingers. In many patients it has been necessary to extract teeth to prevent mutilation of the lips. If physical restraints are removed the patient often pleads for them to be replaced. Other neurological disorders also exhibit self-injurious behaviour but only in Lesch-Nyhan syndrome is this so severe that it leads to the loss of tissue.

Most Lesch-Nyhan patients appear to be mentally deficient. Although, poor performance in IQ tests is in part due to communication difficulties arising as a result of the lack of muscle control. One patient was found to have normal intelligence when testing was adapted to compensate for

these disabilities.

Autopsies on Lesch-Nyhan patients have revealed no consistent distinctive pathological changes in the CNS, and the connection between HPRT enzyme deficiency and the complexities of the Lesch-Nyhan behavioural phenotype is not known. The analysis of autopsy tissues is complicated by damage due to hyperuricemia, which is common in patients with partial and complete HPRT deficiency. Hyperuricemia, the accumulation of excess uric acid, results because HPRT deficiency leads to a build up of hypoxanthine which is converted to uric acid by xanthine oxidase (Figure 1.1).

Levels of the neurotransmitter dopamine in the basal ganglia of Lesch-Nyhan patients are consistently lower than 30% of the normal level (Rosenbloom et al., 1967). It has been suggested that a deficiency of purine nucleotides due to the absence of HPRT activity could impair the arborisation of nigrostriatal neurones, which in humans reaches a peak a few months after birth, at the same time as the behavioural disorders become evident (Baumeister and Frye, 1985). Thus it is interesting that HPRT activity in rat brains increases threefold during the first few weeks after birth (Gutensohn and Guroff, 1972). If as a result of HPRT deficiency a critical stage in neural development is disrupted, it may not be possible to alleviate the neurological symptoms by restoring HPRT activity at a later stage.

A partial HPRT deficiency is associated with hyperuricemia

and gouty arthritis. Hypoxanthine levels become elevated as a result of diminished feedback inhibition of the de novo purine synthesis pathway (Figure 1.1). Excess hypoxanthine is converted by xanthine oxidase to uric acid which precipitates in the joints. Self-injurious behaviour has not been observed in patients with a partial HPRT deficiency. However, evidence of spasticity, cerebellar ataxia, and mild mental retardation can be found in approximately 20% of patients. Approximately 50% of patients with partial deficiency have evidence of mental retardation without accompanying cerebellar symptoms (Stout and Caskey, 1989).

## 1.3 Therapy for HPRT deficiency

Allopurinol, which inhibits xanthine oxidase and thereby prevents the build up of uric acid (Figure 1.1), is an effective treatment for hyperuricemia in patients with either complete or partial HPRT deficiency. However, there is no therapy for the behavioural disturbances of Lesch-Nyhan syndrome. Many drug therapies have been tested without success and development is hindered by the lack of knowledge about the basic physiological defect. Failure to identify a successful therapy may reflect irreversible neural damage early in development.

The Lesch-Nyhan syndrome has been considered a candidate for gene replacement therapy. Retroviral vectors have been used to introduce the human HPRT gene into cultured fibroblasts

and lymphoblasts (Miller et al., 1983). Transient expression of human HPRT in mice has been obtained by transplanting bone marrow cells infected with a retrovirus expressing human HPRT into lethally irradiated mice (Miller et al., 1984). However, this may not be a profitable strategy because the restoration of HPRT activity in circulating blood by either blood transfusion (Edwards et al., 1984) or bone marrow transplantation (Nyhan et al., 1986) has no effect on neurological function in Lesch-Nyhan patients. With the aim of HPRT gene replacement directly into neuronal cells the neurotropic herpes simplex virus type 1 (HSV-1) has been used as a vector. An HSV-1 vector with the human HPRT cDNA under the transcriptional control of the viral thymidine kinase gene has been used to obtain transient HPRT expression in an HPRT deficient rat neuroma cell line (Palella et al., 1988). The vectors used in this study have the considerable disadvantage that they are cytopathic to the target cells. Before such a strategy can be pursued the mechanism which controls the elevated expression of HPRT in specific regions of the CNS needs to be determined, and more suitable vectors need to be developed.

## 1.4 Animal models for HPRT deficiency

Early attempts at producing an animal equivalent of Lesch-Nyhan syndrome involved the administration of caffeine, a methylxanthine, to rats. This resulted in self-mutilation similar to that seen in Lesch-Nyhan patients (Boyd et al.,

1965). Various dopamine agonists have been used to induce abnormal behaviour in monkeys (Goldstein et al., 1986). Surgical denervation of nigrostriatal dopamine neurons was performed when monkeys were 2 to 4 years old. Eight to twelve years later the administration of dopamine agonists, such as L-dopa, resulted in self-mutilative biting of the forelimb digits and spasticity of the hindlimbs.

Recently, mouse models for HPRT deficiency have been produced by Hooper et al. (1987) and Kuehn et al. (1987). Both groups used embryonic stem (ES) cells as a way of introducing an HPRT deficiency into the mouse germ line. Mouse ES cells, obtained from disrupted 3½ day embryos, can be cultured in vitro and still retain their pluripotency (Evans and Kaufman, 1981; Martin, 1981). ES cells taken from a cultured line can be reintroduced into host blastocysts which develop to produce chimeric adult mice. Cells derived from the cultured ES cells may contribute to the germ line of those chimeras (Bradley et al., 1984). Hooper et al. (1987) obtained an HPRT deficient ES cell line, named E14TG2a, by culturing a wild-type ES cell line, derived from strain 129 mice, in selective medium containing 6-thioguanine (6-TG). The characterisation of the genetic mutation in E14TG2a cells is described in Chapter 5 of this thesis. Germ line chimeras were obtained from blastocyst injections of E14TG2a cells. Kuehn et al. (1987) used multiple retroviral infections to generate ES cells with insertion mutations in the HPRT gene. They also obtained chimeras that transmitted the HPRT deficiency through the

germ line. Disappointingly, although these mice have an absolute HPRT deficiency, they have none of the symptoms associated with Lesch-Nyhan syndrome. In fact, they appear perfectly normal. However, the levels of striatal dopamine are reduced by approximately 20% in HPRT deficient mice relative to their normal littermates (Finger et al., 1988). In Lesch-Nyhan patients the levels of striatal dopamine are reduced by between 70% and 90% relative to normal levels.

Jinnah et al. (1989) have demonstrated that it is possible to induce behavioural abnormalities in HPRT deficient mice by the administration of intermediate doses (8mg/kg) of d-amphetamine. This treatment produced different behavioural patterns in HPRT deficient mice to wild-type mice. Normal mice exhibited predominantly locomotor hyperactivity, while HPRT deficient mice engaged in a brief phase of locomotor hyperactivity followed by continuous stereotypic behaviour characterised by gnawing, grooming, sniffing or licking, with little or no locomotor activity. At higher doses (16 to 32mg/kg) and lower doses (2 to 4mg/kg) of amphetamine wild-type and HPRT deficient animals behaved similarly. Why HPRT deficiency leads to increased amphetamine sensitivity is not known although it does suggest a dopaminergic dysfunction in the HPRT deficient mice. Possible mechanisms include enhanced release of dopamine, dopamine receptor sensitivity, reduced metabolism of dopamine or inhibitory systems in the striatum. The neurological abnormalities associated with Lesch-Nyhan syndrome are also thought to result from dysfunction of dopaminergic systems in the striatum

(Baumeister and Frye, 1985). This behavioural difference at least provides an observable test for the effectiveness of new therapies.

Why HPRT deficient mice have none of the symptoms associated with Lesch-Nyhan syndrome remains unknown, but presumably represents a basic metabolic difference between the species. This may be connected with the finding that while HPRT expression is elevated specifically in the basal ganglia of human brains (Rosenbloom et al., 1967), in rats HPRT expression is homogeneously high throughout the brain with no specific elevation in the basal ganglia (Gutensohn and Guroff, 1972). Perhaps this indicates increased de novo purine synthesis and less dependence on the salvage pathway in rodents. Humans also differ from mice in that they lack urate oxidase enzyme activity (Lee et al., 1988). In most mammals urate oxidase is present in the liver and catalyses the oxidation of uric acid to allantoin (Figure 1.1). Humans and certain primates lack this activity and this may influence the consequences of HPRT deficiency.

## 1.5 Structure of the mammalian HPRT gene

HPRT gene sequences have been cloned from three mammalian species. Mouse HPRT cDNA sequences were first cloned from NBR4, an HPRT revertant mouse neuroblastoma cell line which overproduces a mutant HPRT protein approximately 20-fold (Brennand et al., 1982; Konecki et al., 1982). The longest

mouse cDNA clone, pHPT5, was used to isolate cDNA clones from a Chinese hamster cell line (Konecki et al., 1982) and from human cell lines (Brennand et al., 1983). Jolly et al. (1983) independently isolated a human cDNA clone by screening a cDNA library with a human HPRT genomic fragment, obtained from an HPRT deficient mouse cell line into which the human gene had been introduced by DNA-mediated gene transfer.

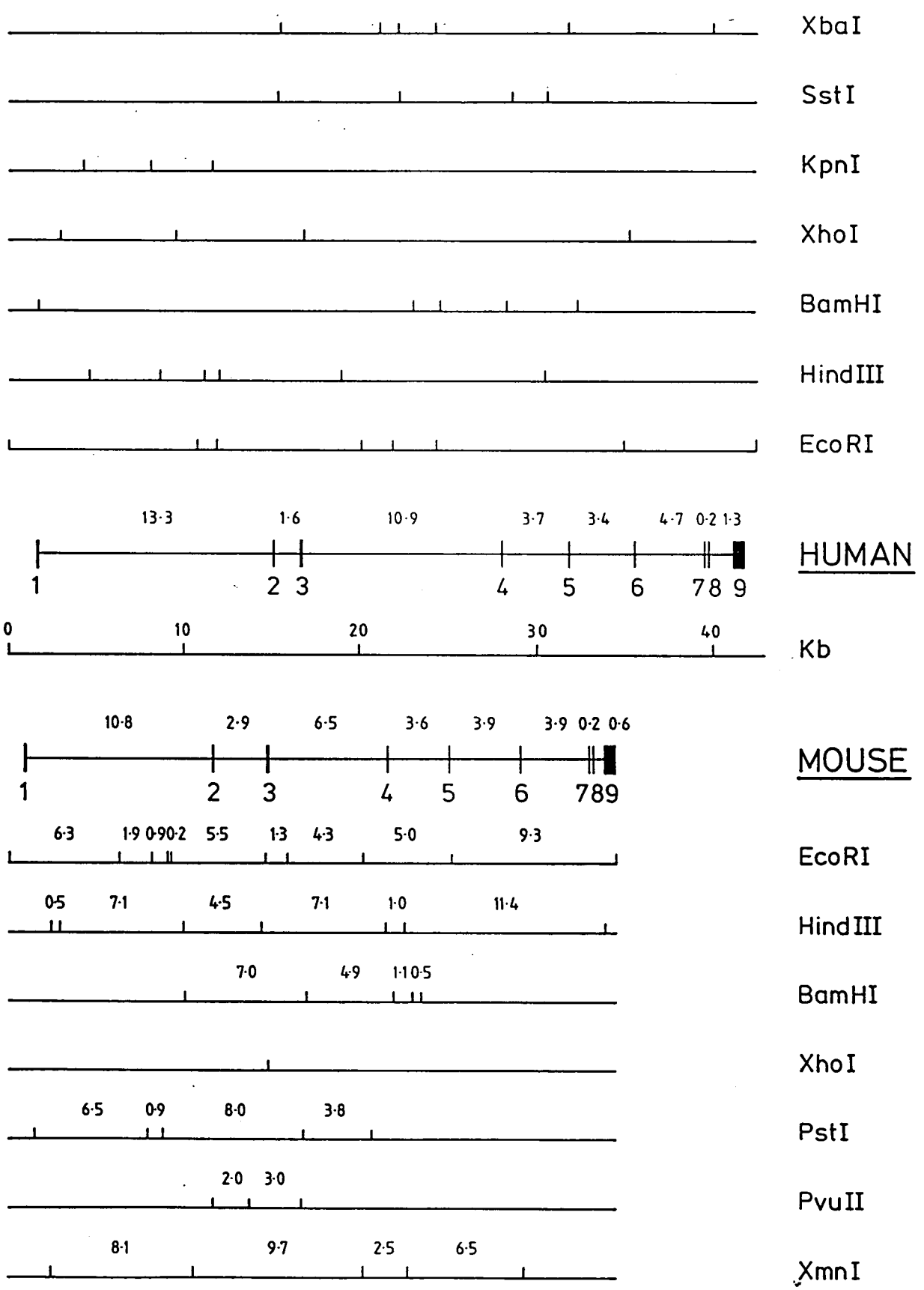
Comparison of the nucleotide sequences of the cDNA from mouse, human and Chinese hamster shows that they each code for a 217 amino acid protein. The initiating methionine is removed by post-translational processing (Wilson et al., 1982). Within the protein coding regions there is greater than 90% nucleotide sequence identity between the three species (Melton, 1987). The mouse and human proteins differ in the amino acid sequence at just seven positions. A high level of nucleotide sequence conservation between species is also found in the 3' untranslated region of the HPRT cDNA. This is discussed in greater detail in the context of gene expression.

The HPRT cDNA clones were used as hybridisation probes to identify HPRT genomic recombinants in libraries from both mouse (Melton et al., 1984) and human cells (Kim et al., 1986; Patel et al., 1986). The structure of the mouse and human HPRT genes, with selected restriction enzyme sites, is shown in Figure 1.2. Characterisation of genomic EcoRI fragments and use of parts of the mouse HPRT cDNA clone,

**Figure 1.2.**

The structures of the mouse and human HPRT genes, with selected restriction enzyme sites. The genes are drawn to scale and exons are depicted as vertical bars. The size of each intron and the sizes of restriction fragments in the mouse gene are indicated (in kb).

# Figure 1.2


XbaI
SstI
KpnI
XhoI
BamHI
HindIII
EcoRI
13·3 1·6 10·9 3·7 3·4 4·7 0·2 1·3
1 2 3 4 5 6 7 8 9
HUMAN
0 10 20 30 40
Kb
10·8 2·9 6·5 3·6 3·9 3·9 0·2 0·6
1 2 3 4 5 6 789
MOUSE
6·3 1·9 0·9 0·2 5·5 1·3 4·3 5·0 9·3
EcoRI
0·5 7·1 4·5 7·1 1·0 11·4
HindIII
7·0 4·9 1·1 0·5
BamHI
XhoI
6·5 0·9 8·0 3·8
PstI
2·0 3·0
PvuII
8·1 9·7 2·5 6·5
XmnI

pHPT5, as hybridisation probes in Southern analysis of wild-type genomic DNA showed that the gene extends over 33kb of the mouse X-chromosome. DNA-mediated gene transfer experiments and the comparison of the HPRT Southern hybridisation patterns for a wild-type cell line and NBR4, in which the gene is amplified, allowed the identification of fragments containing HPRT exons and demonstrated the presence of an HPRT pseudogene. The mouse HPRT gene is divided into nine exons. The first exon contains the 5'-untranslated sequence and 27bp of protein coding sequence. The size of the 5' untranslated region is 118bp or smaller depending on which transcription initiation site is used. The ninth exon contains 45bp of coding sequence and the entire 548bp of 3' untranslated sequence. The intron sizes are shown in Figure 1.2. Five of the introns are positioned between codons and three are positioned within codons.

The human gene is also divided into nine exons by introns which nucleotide sequence analysis showed are in identical positions to those of the mouse gene (Kim et al., 1986). The human gene spans 42kb of the X-chromosome, the greater length relative to the mouse gene, being due to larger introns. The sizes of the exons differ between the two species only in the lengths of the 5' and 3'untranslated regions. The structure of the Chinese hamster HPRT gene has not been completely determined, but where introns have been mapped they are in identical positions to human and mouse introns.

## 1.6 The HPRT promoter

S1 nuclease protection and primer extension analyses showed that the major transcription initiation site of the mouse HPRT gene was just 13bp upstream of the start of pHPT5, the longest cDNA clone (Melton et al., 1984). Nucleotide sequencing of the region upstream of this transcription initiation site (which is denoted nucleotide position +1) in the mouse gene failed to reveal the sequence motifs which at the time were believed to be normally present in eukaryotic gene promoters, having been identified in tissue-specific genes. For instance, the nearest match to the TATA box consensus, usually located 20 to 30bp upstream of the transcription start site (Corden et al., 1980), occurs greater than 700bp upstream of the gene. Furthermore, there is no sequence match for the CAAT box which is situated about 80bp upstream of many eukaryotic genes (Benoist et al., 1980). The 5' end of the gene was assayed for promoter activity. An HPRT minigene was constructed (Melton et al., 1984) which consisted of 845bp of 5' flanking sequence, the putative promoter region, fused onto the wild-type human HPRT cDNA sequence through a restriction site in the 5' untranslated region. In DNA-mediated gene transfer experiments this minigene, pHPT44, transformed HPRT deficient Chinese hamster, RJK88, cells to the HPRT+ phenotype. Thus, it was demonstrated that the 845bp immediately upstream of the mouse HPRT gene, while not necessarily containing all the elements regulating transcription, does have promoter activity in cultured

cells.

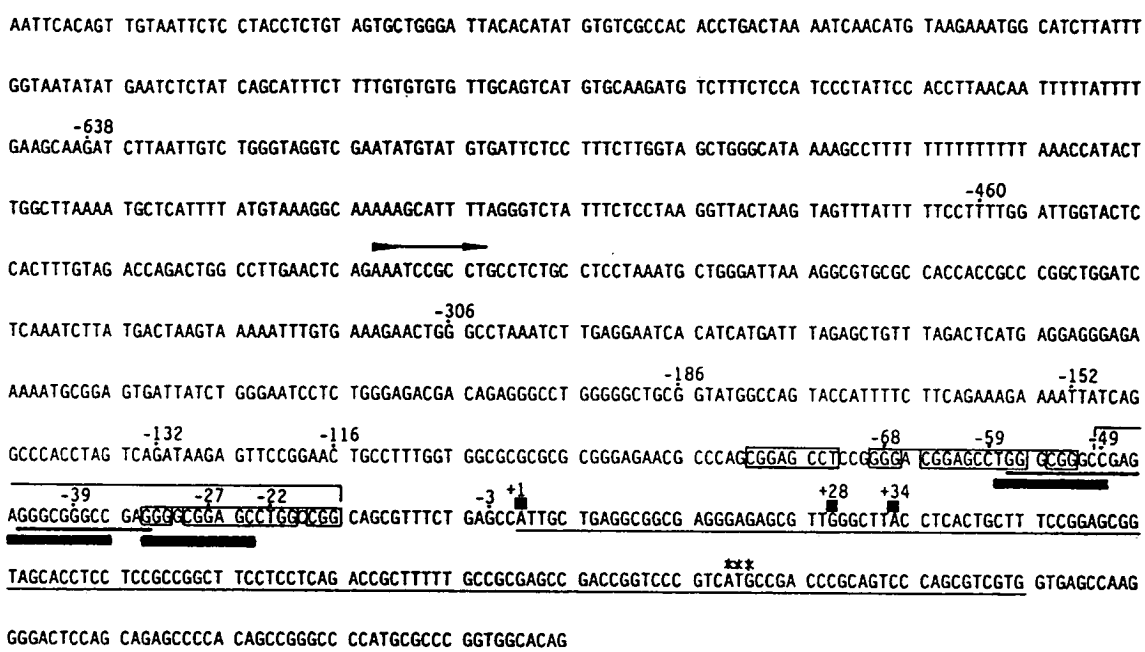
Further analysis of the mouse HPRT promoter region was reported by Melton et al. (1986). Figure 1.3 shows the nucleotide sequence around the 5' end of the HPRT gene. As stated previously there are no TATA or CAAT consensus sequence matches in the expected positions. However, the region in which the 5' end of the gene is situated is highly CG-rich. In particular, the 100bp immediately upstream of the start site is 80% CG-rich. The possible significance of this is discussed below in relation to X-inactivation and DNA methylation.

Computer-assisted analysis of the nucleotide sequence of the promoter region revealed a complex pattern of direct repeats (Figure 1.3). There are two 12bp perfect repeats in tandem, between 30 and 60bp upstream of the transcription start site. Between nucleotide positions -80 and -15 there are two imperfect 18bp repeats, the central region of which is repeated a third time. Nine out of ten bases from the centre of the 18bp repeat are found in the enhancer region of the DNA tumour virus SV40 (Moreau et al., 1981). In the same region of the mouse HPRT promoter there are three matches to the consensus sequence for the binding site of the transcription factor Sp1 (Kadonaga et al., 1986). Sp1 is discussed in detail in section 1.7. A match to the consensus sequence for the AP2 binding site is also found in this region of the HPRT promoter. AP2 binding sites together with AP3 binding sites constitute the functional core of the SV40 enhancer (Herr and

**Figure 1.3.**

Structure of the mouse HPRT promoter region. The sequence is numbered from the main transcription initiation site (+1). The 145bp first exon is underlined and within it the transcription initiation sites at +28 and +34 are indicated. Three asterisks indicate the translation initiating codon. The 18bp imperfect direct repeats and the third repeat of part of the central section are boxed. The 12bp direct repeats are underlined. Three matches to the consensus Sp1 binding site are heavily underlined. The sequence enclosed by the square bracket has been used as a synthetic promoter (Section 1.7). The end points of 5' deletions used to map functional elements (Section 1.6) are indicated by a dot over the position of the first base present. The arrow indicates the 5' boundary of the CG-rich region.

# Figure 1.3


AATTCACAGT TGTAATTCTC CTACCTCTGT AGTGCTGGGA TTACACATAT GTGTCGCCAC ACCTGACTAA AATCAACATG TAAGAAATGG CATCTTATTT
GGTAATATAT GAATCTCTAT CAGCATTTCT TTTGTGTGTG TTGCAGTCAT GTGCAAGATG TCTTTCTCCA TCCCTATTCC ACCTTAACAA TTTTTATTTT
-638
GAAGCAAGAT CTTAATTGTC TGGGTAGGTC GAATATGTAT GTGATTCTCC TTTCTTGGTA GCTGGGCATA AAAGCCTTTT TTTTTTTTTT AAACCATACT
-460
TGGCTTAAAA TGCTCATTTT ATGTAAAGGC AAAAAGCATT TTAGGGTCTA TTTCTCCTAA GGTTACTAAG TAGTTTATTT TTCCTTTTGG ATTGGTACTC
CACTTTGTAG ACCAGACTGG CCTTGAACTC AGAAATCCGC CTGCCTCTGC CTCCTAAATG CTGGGATTAA AGGCGTGCGC CACCACCGCC CGGCTGGATC
-306
TCAAATCTTA TGACTAAGTA AAAATTTGTG AAAGAACTGG GCCTAAATCT TGAGGAATCA CATCATGATT TAGAGCTGTT TAGACTCATG AGGAGGGAGA
-186 -152
AAAATGCGGA GTGATTATCT GGGAATCCTC TGGGAGACGA CAGAGGGCCT GGGGGCTGCG GTATGGCCAG TACCATTTTC TTCAGAAAGA AAATTATCAG
-132 -116 -68 -59 -49
GCCCACCTAG TCAGATAAGA GTTCCGGAAC TGCCTTTGGT GGCGCGCGCG CGGGAGAACG CCCAGCGGAG CCTCCGGGGA CGGAGCCTGG GCGGGCCGAG
-39 -27 -22 -3 +1 +28 +34
AGGGCGGGCC GAGGGCGGGA GCCTGGCGGG CAGCGTTTCT GAGCCATTGC TGAGGCGGCG AGGGAGAGCG TTGGGCTTAC CTCACTGCTT TCCGGAGCGG
TAGCACCTCC TCCGCCGGCT TCCTCCTCAG ACCGCTTTTT GCCGCGAGCC GACCGGTCCC GTCATGCCGA CCCGCAGTCC CAGCGTCGTG GTGAGCCAAG
GGGACTCCAG CAGAGCCCCA CAGCCGGGCC CCATGCGCCC GGTGGCACAG

Gluzman, 1985). Mutations which prevent AP2 binding reduce enhancer activity in HeLa cells by approximately 60%.

The nucleotide sequence of the promoter region of the human gene has also been determined (Kim et al., 1986; Patel et al., 1986). The overall homology between the mouse and human sequences for a region including the 5' untranslated region and extending 150bp upstream of the major transcription initiation site is only 51% (Kim et al., 1986). However, localised similarity becomes obvious if the two sequences are aligned around the common region of multiple Sp1 binding sites. Allowing for up to three mismatches, five putative binding sites are present in the mouse promoter and six in the human. The spacing between sites is similar in both species, with a tendency for adjacent sites to be approximately one helical turn apart. There are also sequences similar to the central region of the 18bp imperfect repeats in the mouse in equivalent positions in the human promoter. Both Patel et al. (1986) and Kim et al. (1986) have identified additional regions of homology between the two species, some of which also occur in the SV40 enhancer. In addition, one region of the human promoter has 17 out of 22 residues in common with the SV40 enhancer but has no equivalent in the mouse promoter.

To facilitate functional analysis of the mouse HPRT promoter, an HPRT minigene, pDWM1, was constructed in which the 33kb chromosomal gene is represented in 3kb of sequence (Melton et al., 1986). The pDWM1 minigene contains 845bp of

5' flanking sequence, the entire 5' and 3' untranslated regions and the HPRT protein coding sequence interrupted by the last two introns. This minigene can transform HPRT-deficient Chinese hamster cells to the HPRT+ phenotype (as assayed by the ability to grow in HAT-containing medium) at a frequency of 2 to 5 x $10^{-4}$ per 5μg of plasmid DNA.

Deletional analysis of the 845bp of 5' flanking sequence present in pDWM1 was used to identify functionally important regions. Exonuclease Bal31 was used to generate a progressive series of deletions of the promoter region extending towards the gene from the 5' end. The end points of the 5' deletions are indicated in Figure 1.3, numbered relative to the major transcription initiation site +1. Each deletion construct was used to transform RJK88 cells (Chinese hamster fibroblasts with a deletion of the HPRT gene) to the HPRT+ phenotype. HAT-resistant clones were assayed for their ability to incorporate [$^{3}$H]-hypoxanthine. Because less than 10% of normal HPRT activity is sufficient to support growth in HAT-containing medium (Melton, 1981), a wide range of expression levels can be detected using this assay. Melton et al. (1986) showed that it is possible to delete as far as nucleotide -39 and still retain greater than 30% of the incorporation for the starting minigene pDWM1. However, a minigene in which the promoter was deleted as far as -27 gave an incorporation level of just 2.5% of the original value. The sequence between -39 and -27 contains one of the putative Sp1 binding sites. Deletions further into the promoter region towards the initiation site gave

similar low levels of incorporation. A reproducibly higher level of incorporation than for pDWM1 was obtained for deletion mutants with end points between -306 and -152. This may indicate the presence of a negative control element but has not been analysed further.

3' deletions of the promoter region fused to the neomycin phosphotransferase (neo) transcription unit, from pSV2 neo, lacking its own promoter were assayed for the ability to transform RJK88 cells to G418 resistance. Deletion to -38 gave a transformation frequency of 40% of a construct containing HPRT sequences from -638 to +113 fused to neo. While deleting as far as -44 reduces the transformation frequency by 95%. Thus, it is possible to delete from the 3' side of the promoter as far as -38, which loses the proximal 18bp repeat and Sp1 site, and still retain transcriptional activity. While, from the 5' direction it is possible to delete as far as -39 with a similar lack of effect. The conclusion has to be that the mouse HPRT promoter is functionally duplicated. Functional redundancy has also been identified in the HMGCoA reductase promoter (Osbourne et al., 1987).

The wild-type endogenous HPRT gene has three main transcription initiation sites at nucleotides +1, +28 and +34, as determined by S1 nuclease protection and primer extension analysis. The transcription initiation sites used by the deletion mutants were analysed. For all the 5' deletion mutants only the three wild-type initiation sites

were used, although HPRT mRNA was barely detectable for the mutant deleted to -27. For the 3' deletion mutants initiation occurred at each of the three wild-type sites until they were deleted. Then initiation was directed to sites in the equivalent position with respect to the promoter. If the promoter region was deleted beyond -38, transcription initiation occurred from many sites extending up to 350bp upstream of +1. Nucleotide -350 is close to the boundary of the CG-rich region.

The experiments described above demonstrated that the DNA sequences essential for transcription of the HPRT gene in cultured Chinese hamster fibroblasts are all contained within the 50bp immediately upstream of the major transcription initiation site. Less detailed analysis of the human HPRT promoter (Patel et al., 1986) demonstrated that 100bp of 5' flanking sequence is sufficient for transcription of a human minigene.

## 1.7 The organisation of cis-acting DNA elements that regulate transcription initiation

Eukaryotic genes can be defined as either housekeeping genes or specialised genes, dependent upon their pattern of expression. Housekeeping genes, of which HPRT is an example, are constitutively expressed in all the cells of an organism. The expression of specialised genes is restricted to certain cell types or defined stages in development. In general

terms the organisation of the DNA elements involved in the control of transcription is different in specialised genes to housekeeping genes. However, there are no hard rules and some genes show characteristics of both housekeeping genes and specialised genes. Rather than representing two distinct classes, housekeeping genes and specialised genes, as defined above, are probably the extremes of a spectrum.

The transcription of specialised genes is characteristically regulated by two nominally distinct cis-acting DNA elements, the promoter and the enhancer (reviewed in Maniatis et al., 1987). The promoter region is typically situated within approximately 100bp immediately upstream of the transcription initiation site. Promoters contain a combination of binding sites for general transcription factors and factors which determine promoter specificity. The promoters of most specialised genes contain a binding site for the general transcription factor TFIID. This binding site, the TATA box, is usually situated between 20 and 30bp upstream of the transcription start site. The binding of TFIID to the TATA box is the first step in transcription initiation (Davison et al., 1983; Fire et al., 1984; Van Dyke et al., 1988; Buratowski et al., 1989) and focuses the assembly of RNA polymerase II with other general transcription factors to form the active transcription complex. The yeast gene encoding TFIID has been cloned (Hahn et al., 1989; Eisenmann et al., 1989; Horikoshi et al., 1989). TATA box-containing promoters usually have a single transcription initiation site. If the TATA box is deleted

transcription is then initiated from heterogeneous sites (Grosschedl and Birnstiel, 1980; Grosveld et al., 1982; Benoist and Chambon, 1981). Hence TFIID and associated factors determine the site of transcription initiation.

Additional promoter elements, 30 to 110bp upstream of the start site regulate the frequency of transcription initiation. A consensus sequence commonly found in this region is the CAAT box. A group of proteins has been identified which bind to the CAAT box (Santoro et al., 1988). The protein factors that bind these upstream elements interact with the TFIID protein:DNA complex to regulate the frequency of transcription initiation (reviewed in Maniatis et al., 1987). The best characterised of the general transcription factors is Sp1 (reviewed in Kadonaga et al., 1986). Sp1 binding is essential for the expression of both the early and late genes of the SV40 virus (Gidoni et al., 1984). The consensus sequence for the Sp1 binding site is GGGCGG and is functional in either orientation. Matches to this consensus are found in the promoter regions of many genes, including HPRT. The human Sp1 gene has been cloned (Kadonaga et al., 1987).

In addition to these general transcription factor binding sites, specialised promoters contain sites for factors which determine the specialised properties of the promoter. An example of such a binding site is the heat shock element (HSE). Three matches to the HSE consensus sequence are found in the Xenopus hsp70 promoter and are responsible for

activation of that gene in response to hyperthermia. HSE consensus matches are found in a range of species. HSEs can act over considerable distances in an orientation independent manner and have optimal activity in combination with both a CAAT box and a TATA box. The same factor-binding sites from the Drosophila hsp70 gene can function either as regulatory promoter elements in proximity to the TATA box or, when duplicated, as inducible enhancer elements, active over distances of several kilobase pairs (Bienz and Pelham, 1986).

The human metallothionein promoter contains at least five distinct elements that regulate transcription and its cell specificity. The general transcription factor binding sites include a TFIID binding site, an Sp1 binding site and two AP1 binding sites (Lee et al., 1987). AP1 binding sites are also present in the SV40 enhancer and the promoters of a number of other genes (reviewed by Jones et al., 1988). Two other types of element exist in the metallothionein promoter. The metal regulatory elements which confer responsiveness to induction by heavy metals and the glucocorticoid responsive elements which confer responsiveness to induction by steroid hormones (Karin et al., 1984). Hence, promoters consist of a combination of general elements that are required for the transcription of many genes and specialised elements which determine the expression pattern specific to that gene.

The transcription of many specialised genes is also

regulated by enhancers. The enhancer effect was first identified for sequences upstream of the SV40 early promoter (Banerji et al., 1981; Moreau et al., 1981; Fromm and Berg, 1983). Enhancers can stimulate transcription from distances of greater than 10kb either upstream or downstream of the gene, or from within the gene itself. Enhancers like upstream promoter element function independent of orientation. Enhancers have now been identified for many genes although the SV40 enhancer remains the best characterised (reviewed in Jones et al., 1988). Like promoters, enhancers are composed of sequence elements which bind one or more protein factors. The short sequence elements, combinations of which constitute the enhancer, have been termed enhansons (Ondek et al., 1988; reviewed, Dynan, 1989). These protein binding sites often occur in pairs separated by a precise distance indicating a requirement for stereoscopic alignment to allow protein-protein interaction. Binding site pairs have been called modules (Dynan, 1989). Mutation of a single module in the SV40 enhancer which leads to diminished transcription stimulation, can be suppressed, or reverted, by duplication of remaining unmutated modules (Herr and Clarke, 1986). Enhancers may contain binding sites for many transcription factors with different tissue-specificities. The specificity of transcription stimulation is determined by the particular combination in a given enhancer.

The distinction between enhancers and promoters is nominal, that is, relating only to their position

relative to the gene. Structurally they are essentially the same and many transcription factors, including AP1 (Lee et al., 1987), bind to both promoters and enhancers (reviewed in Jones et al., 1988). Furthermore, elements taken from promoters can act as enhancers with heterogeneous promoters (Bienz and Pelham, 1986). A notable exception to these generalisations is the TATA box which is peculiar to promoters. This is not surprising, in light of its function, as discussed previously.

Mammalian HPRT is a housekeeping gene. It is expressed in all cell-types. A comparison of the HPRT promoter with the promoters of other housekeeping genes reveals a tendency towards certain characteristics which are different to specialised genes. All housekeeping gene promoters so far studied are in highly CG-rich regions of the genome. The HPRT promoter sits in a CG-rich island approximately 2kb long. More locally, the 100bp immediately upstream of the major transcription start site of the mouse HPRT gene are 80% CG-rich. The mouse adenine phosphoribosyl transferase (APRT) gene has a 200bp region which is 66% CG-rich (Dush et al., 1985), the mouse and hamster dihydrofolate reductase (DHFR) genes have a 140bp region which is 65% CG-rich (Farnham et al., 1985; Mitchell et al., 1986), and the hamster 3-hydroxy-3-methylglutaryl coenzyme A (HMGCoA) reductase gene promoter has a 100bp region which is also 65% CG-rich (Reynolds et al., 1984). Other housekeeping genes with CG-rich promoter sequences include human adenosine deaminase (Valerio et al., 1985), human phosphoglycerate

kinase (Singer-Sam et al., 1984) and human triose phosphate isomerase (Brown et al., 1985). All of the housekeeping genes listed above contain multiple copies of the consensus sequence for the Spl binding site (reviewed Kadonaga, 1986).

Most housekeeping genes lack matches to the TATA box consensus sequence in their promoter regions. Of the housekeeping genes discussed above, only triose phosphate isomerase appears to contain a TATA box (Brown et al., 1985). Presumably, a different protein performs the apparently pivotal role of TFIID in promoters that lack a TFIID binding site. Interestingly, housekeeping gene promoters initiate transcription from heterogeneous sites. As described previously, TFIID positions the site of transcription initiation in specialised genes.

Bidirectionality is a property that has been demonstrated for some housekeeping promoters. That is, the ability to promote the transcription of both DNA strands divergently from the promoter region. A 34bp sequence from the mouse HPRT promoter, from -49 to -16 (Figure 1.3), which corresponds to the region identified as being important by deletional analysis, has been synthesised chemically and analysed for its ability to initiate transcription. This synthetic promoter, which contains two putative Spl binding sites and the putative AP2 binding site, was able to direct transcription from the normal sites when placed upstream of a promoterless HPRT minigene, or a neo gene. Interestingly,

the synthetic promoter was able to promote transcription when placed in either orientation relative to the reporter gene. It was also capable of simultaneously promoting the transcription of both an HPRT minigene and a divergent neo gene when placed between the two. S1 nuclease protection and Northern analyses of wild-type mouse mRNA, using probes from the HPRT promoter region, indicate the presence of additional transcripts, some of which are divergent. Further analysis of these transcripts has been hindered by the presence of highly repeated sequences in the same region.

Divergent transcripts have also been identified originating adjacent to the DHFR gene (Farnham et al., 1985). The mouse DHFR gene has two promoter regions, approximately 300bp apart, both of which contain Sp1 binding sites and lack TATA boxes. Either promoter is able to initiate bidirectional transcription. The DHFR divergent transcripts have sequence identity with bacterial genes involved in DNA mismatch repair (Linton et al., 1989). Many transcription factor binding sites, including the Sp1 binding site, are functionally bidirectional or orientation independent, despite being asymmetrical. This combined with the absence of the focusing influence of a TATA box probably explains the bidirectional properties of these promoters.

In summary,while specialised genes tend to have control elements divided into promoters and enhancers which can be many kilobases from the gene, housekeeping genes tend to have

all control elements relatively close to the gene. The less regulated pattern of expression of housekeeping genes seems to be reflected in less strictly organised control regions which lack specialised elements. Having said that, housekeeping genes are not unregulated. The HMGCoA reductase gene (Reynolds et al., 1984) is subject to negative feedback regulation in response to cholesterol levels and DHFR (Farnham et al., 1985) expression increases during DNA synthesis. HPRT enzyme activity is elevated in brain tissue. While the mechanism of expression elevation is undetermined, data presented in this thesis demonstrate that the increased enzyme activity is paralleled by increased steady-state levels of HPRT mRNA.

## 1.8 DNA methylation and gene expression

As described earlier, the HPRT promoter region is highly CG-rich. The 100bp immediately upstream of the transcription start site are 80% CG-rich which compares to the average of 40% for bulk DNA. This concentrated CG-richness, lies within a more extensive CG-rich island spanning approximately 2kb. It has been noticed by Bird (1986) that large CG-rich islands tend to be associated with genes, predominantly but not exclusively with housekeeping genes as opposed to specialised genes. Within these CG-rich islands the density of CpG dinucleotides may be more than ten times higher than in bulk DNA. In mammals the majority of CpG dinucleotides are methylated at the cytosine residue. The low frequency of

CpG dinucleotides in bulk DNA may be due to the mutation of methylcytosines. CpG-rich islands tend to be hypomethylated and can consequently be identified using methyl CpG sensitive restriction enzymes, such as HpaII. Consequently, they are also known as HTF (HpaII tiny fragment) islands.

Lavia et al., (1987) randomly isolated two HTF islands from mouse chromosomal DNA. Both islands were found to be the origin of divergent transcripts that were detectable in a range of mouse tissues. Bidirectional initiation may therefore be a property of many housekeeping gene promoters. The apparent correlation between transcriptional activity and hypomethylation of HTF islands leads to the question of whether methylation is directly involved in the control of gene expression.

In female mammalian cells only one copy of the HPRT gene is expressed due to X-chromosome inactivation (Lyon, 1972), which occurs early in embryogenesis. The first evidence for the involvement of methylation in X-inactivation, albeit indirect, came from DNA transfection experiments. DNA from an active X-chromosome was much more efficient at transforming HPRT deficient cells to HPRT expressing cells than DNA from the inactive X-chromosome (Liskay and Evans, 1980; Chapman et al., 1982; Lester et al., 1982; Venolia and Gartler, 1983). In addition, the incorporation into DNA of the cytidine analogue, 5-azacytidine, which cannot be methylated results in a high frequency of reactivation of genes on the inactive X-chromosome (Mohandas et al., 1981;

Lester et al., 1982; Graves, 1982). It has subsequently been demonstrated using methylation sensitive enzymes which have CpG in their recognition sequence, that several CpGs around the HPRT promoter are methylated on the inactive but not on the active X-chromosome (Yen et al., 1984; Lock et al., 1986; Wolf et al., 1984). Other X-linked genes for which a similar correlation has been observed are the housekeeping genes: glucose-6-phosphate dehydrogenase (Toniolo et al., 1984; Wolf et al., 1984a) and phosphoglycerate kinase (Keith et al., 1986). It is important to note, that X-linked genes without CG-islands are still inactivated on the inactive chromosome. The precise interrelationship between methylation and inactivation is undetermined. In the mouse embryo inactivation of the HPRT gene on the inactive X-chromosome occurs several days before methylation is detectable in the CG-rich island (Lock et al., 1987). In addition, in the early extra-embryonic tissues, inactivation does not appear to be strongly associated with methylation (Kratzer et al., 1983).

A protein has been identified which specifically binds to methylated CpG-containing sequences (Meehan et al., 1989). Methyl-CpG binding protein (MeCP) is abundant in a range of tissues, but interestingly, is only present at very low levels in embryonal carcinoma cells, which are particularly active methylators of introduced DNA.

Methylation is also associated with gene inactivation as a

mechanism of parental imprinting (reviewed by Solter, 1987). Swain et al., (1987) demonstrated that the expression of a c-myc transgene, in one line of transgenic mice, was dependent upon whether it was inherited from the paternal or maternal parent. When inherited from the mother this autosomal transgene was methylated and not expressed. However, the transgene was expressed and hypomethylated when inherited from the father. The effect of methylating cytosine residues on DNA-protein interaction is discussed by Dynan (1989a).

## 1.9 Post-transcriptional regulation of gene expression

Gene promoters are not the only influences on gene expression and the steady-state levels of mRNA in a cell are not solely dependent upon the rate of transcription of the corresponding gene. Expression may also be regulated by RNA processing, translation or post-translational processing. The steady-state levels of mRNA are also a function of the rate of mRNA turnover (reviewed Ross, 1988). This is a possible mechanism for regulating the elevated expression of HPRT in brain tissue. There is currently no evidence for regulation at the level of transcription initiation in the HPRT gene. It has been demonstrated for some housekeeping genes that mRNA accumulation is controlled at the level of transcript degradation (Carneiro and Schibler, 1984). Ten cDNA clones were selected from a library on the basis of the abundance of their complementary mRNA. That is, five were selected for which the corresponding mRNAs are moderately

abundant (220 to 640 copies per L-cell) and a further five were chosen for which the corresponding mRNAs are rare (5 to 76 copies per L-cell). Northern analysis showed each of the ten mRNAs is present in a range of tissues. Therefore, all ten probably represent transcripts from housekeeping genes. Run-off transcription experiments performed on isolated L-cell nuclei indicated that there was no correlation between the relative abundance of a transcript and the rate of transcription of each of the ten genes. That is, the transcription rate of the genes with moderately abundant mRNA was no higher than of genes with rare mRNA, suggesting that regulation is post-transcriptional. This contrasts with earlier findings that transcriptional regulation was predominantly responsible for the control of expression of tissue-specific genes (Derman et al., 1981). The relative stability of each of the mRNAs was compared in a kinetic labelling experiment. L-cells were incubated with [$^{32}$P] orthophosphate for three hours and 24 hours. Purified polyadenylated mRNA was analysed by dot blot hybridisations. Quantitation of hybridisation signals showed that the majority of the moderately abundant mRNAs accumulate to a higher extent than the rare mRNAs, between the labelling times of 3 hours and 24 hours. Carneiro and Schibler (1984) concluded that the steady-state levels of the mRNAs studied are determined by differential stabilities, rather than rates of transcription.

Most mRNA molecules are protected against exonucleolytic attack by poly (A) sequences (Bergmann and Brawerman, 1977)

or, in the case of some histone mRNAs which lack a poly (A) tail, by stem-loop secondary structures (Georgiev and Birnstiel, 1985). More specific elements must be involved if controlled mRNA turnover is a mechanism for regulating gene expression. In this respect, it is interesting that the 3' untranslated region of the HPRT mRNA, which is 545bp long in the mouse and approximately 50bp longer in the human, shows 70% nucleotide identity between the mouse and human sequences. Within the 3' untranslated region there are large blocks of sequence with absolute identity between mouse, human and Chinese hamster. One block of absolute sequence identity which is 40 nucleotides long is extremely AU-rich (92.5%). Shaw and Kamen (1986) have identified a conserved AU-rich sequence from the 3' untranslated region of the lymphokine granulocyte-monocyte colony stimulating factor (GM-CSF) mRNA, which mediates selective degradation. These GM-CSF sequences are 93% conserved between mouse and human. Furthermore, similar sequences are present in the 3' untranslated regions of numerous other lymphokine, cytokine and proto-oncogene mRNAs. The common feature of these transcripts is that they all code for transiently expressed growth-regulating proteins. The 51 nucleotide long GM-CSF AT-rich DNA sequence was synthesised chemically and introduced into the 3' untranslated region of the rabbit β-globin gene, which produces relatively stable mRNA (half-life greater than 17 hours). In cultured cell transfection experiments the β -globin transcript containing the AU-rich sequence accumulates to just 3% of the level of a control β-globin mRNA. The control gene also had a 51bp insert but

this insert contained 14 G and C residues interspersed throughout the sequence. Shaw and Kamen (1986) demonstrated that the effect of introducing the AU-rich sequence is to decrease the β -globin mRNA half-life. The longest sequence motif in common between GM-CSF and the range of lymphokines and proto-oncogenes analysed is AUUUA. Although, the most common feature is a single adenine followed by three or more uridine residues. A cytoplasmic protein which binds specifically to mRNAs containing the AUUUA sequence has now been identified (Malter, 1989). Using either criterion, the HPRT 3' untranslated region has the same sequence elements. Controlled rapid degradation of mRNA is an obvious advantage for genes that code for proteins involved in the regulation of growth. However, it is less obvious why the HPRT gene should share this characteristic. The possible role of the 3' untranslated region in the control of HPRT expression is of considerable interest.

## 1.10 Project aim

As described above HPRT enzyme activity is found at relatively low levels in all mammalian tissues with the exception of the CNS, in which activity is elevated (Rosenbloom et al., 1967; Gatensohn and Guroff, 1972). In humans, at least, the elevation is concentrated in the basal ganglia. This tissue distribution is particularly interesting because of the consequences of HPRT deficiency on CNS function. At the outset, the aim of this project was

to gain further information about the control of HPRT gene expression. The detailed study of the mouse HPRT promoter has been described. To investigate the control of the brain specific elevation in expression we have extended our experiments to whole animal systems. Two approaches are described in this thesis. Firstly, the use of transgenic mice produced by DNA microinjection into fertilised eggs. Secondly, we have exploited gene targeting in embryonic stem cells to manipulate the HPRT gene in the mouse germ line.

Knowledge of the mechanism by which HPRT expression is elevated in brain, in addition to being of fundamental biological interest, may help elucidate the relationship between HPRT expression and Lesch-Nyhan syndrome. It would also be a pre-requisite for a successful gene therapy strategy.

# 2. MATERIALS AND METHODS

## 2.1 Materials

### 2.1.1 Suppliers of Laboratory Reagents

Standard laboratory reagents: BDH, Fisons, Sigma, May and Baker, Koch-Light and Bio-Rad.

Restriction enzymes and DNA modifying enzymes: Bethesda Research Laboratories (BRL), Boehringer Mannheim and Pharmacia.

Deoxyribonucleotides: Sigma.

Agarose (Ultra pure and low melting-point): BRL.

Bacterial Media Reagents: Difco Laboratories.

Media Reagents for mammalian cell culture: Flow laboratories and Sera-lab.

### 2.1.2 E.coli Growth Media

Solutions were autoclaved and stored at room temperature.

Luria Broth (LB): 10g Bacto-Tryptone, 5g yeast extract, 10g NaCl, made to 1 litre and adjusted to pH7.2 with NaOH.

Luria-agar (LB-agar): LB plus 15g agar per litre.

When required antibiotics were added to media immediately before use.

LB - ampicillin: LB plus 0.05mg/ml ampicillin.

Ampicillin stock: 10mg/ml in $H_2O$, stored at 4°C.

LB-tetracycline: LB plus 0.025mg/ml tetracycline.

Tetracycline stock: 20mg/ml in ethanol/water (50%, v/v), stored at -20°C.

### 2.1.3 E.coli Strains

RR1: $F^-$, hsdS, ara, proA, lacY, galk, rpsL ($Sm^r$), xyl, mtl, supE, $\lambda^-$. Bolivar et al., (1977).

JM83: ara, Δlac-pro, strA, thi, ϕ80dlacZΔM15. Vieira and Messing (1982).

### 2.1.4 ES cell culture media

Embryonic stem cells were grown in Buffalo rat liver cell (BRL)-conditioned medium, without a feeder layer. The

medium used was Glasgow modified Eagle's medium (Flow laboratories) supplemented with 1 x nonessential amino acids and 1mM sodium pyruvate, and 5% (v/v) of both fetal calf serum and newborn calf serum (Sera-Lab). ES cells were grown in the absence of antibiotics in 37°C incubators with 5% $CO_2$ in air.

BRL-conditioned medium: BRL cells were grown to confluence in 75$cm^2$ tissue culture flasks. The medium was discarded and 15ml of fresh medium was added. After 2 days this was collected and replaced with a further 15ml of fresh medium. In total 45ml of conditioned medium was collected from each flask. The conditioned medium was filtered (pore size, 0.2 µm) and stored at -20°C. ES cells were grown on gelatin-coated tissue culture flasks in 60% BRL-conditioned medium, 40% fresh medium, with 0.1mM β -mercaptoethanol.

HAT-containing medium: 60% BRL-conditioned medium supplemented with $10^{-4}$M hypoxanthine, $10^{-6}$M aminopterin and $10^{-5}$M thymidine.

### 2.1.5 ES cell lines

E14: A wild-type male ES cell line isolated from strain 129/Ola blastocysts (Handyside et al., 1989).

E14TG2a: An HPRT deficient mutant ES cell line derived from E14 by selection in 6-thioguanine (Hooper et al., 1987).

### 2.1.6 Mouse strains

Mouse strains, obtained from Olac 1976 Ltd., were a gift from M.L. Hooper, Department of Pathology, University of Edinburgh.

129/Ola: an inbred strain which is homozygous for the recessive alleles chinchilla ($\underline{c}^{ch}$), pink-eyed dilution ($\underline{p}$) and white-bellied agouti ($\underline{A}^{W}$).

HPRT-deficient mice: mice of the outbred strain MF1, homozygous or hemizygous for the $\underline{hprt}^{b-m3}$ mutant allele.

### 2.1.7 Plasmids and Constructs

pDWM1LS+127/+135, pDWM101 and pDWM100 were constructed by Carolanne McEwan.

pDWM1LS+127/+135: this is a linker scanning mutant derived from the HPRT minigene pDWM1Δ1 (Melton et al., 1986). It was constructed by annealing a 3' Bal 31 deletion of pDWM1Δ1, with end point +127, to a 5' Bal 31 deletion of pDWM1Δ1, with end point +135, through a 10bp BamHI linker. The resultant minigene contains 635bp of 5' flanking sequence, the entire 5' untranslated region, the HPRT coding sequence interrupted by the seventh and eighth introns, and the HPRT 3'

untranslated region. The sequence differs from wild-type in the region of the BamHI linker, as illustrated in Figure 3.2.

pDWM101: the structure of pDWM101 is illustrated in Figure 5.1. pDWM101 consists of: a 4.7kb BamHI-EcoRI restriction fragment containing exon 2 and derived from the mouse HPRT genomic clone, λHPT 1 (Melton et al., 1984); a 3.0kb BglII - BamHI fragment containing the HPRT promoter and exon 1, derived from λHPT32; and a 1.35kb EcoRI fragment, containing exon 3 derived from λHPT1; cloned in pUC8 (Vieira and Messing, 1982). The resulting 11.7kb plasmid has the same organisation as the equivalent part of the HPRT gene itself except that the first intron is reduced from 10.8kb to 4.1kb in the vector.

pDWM100: this vector was constructed as a positive control for transformation of ES cells. It was derived from pDWM101. The 1.25kb XhoI-EcoRI fragment in pDWM101 was replaced by the 1.8kb XhoI-EcoRI fragment from pDWM1 (Melton et al., 1986) which contains the remainder of the HPRT coding sequence and the HPRT 3' untranslated region. The resultant minigene, pDWM100, has HPRT activity and can restore activity to HPRT deficient ES cells without a requirement for homologous recombination.

pHPT5: this construct consists of the full-length (1300bp) mouse HPRT cDNA cloned by GC tailing into the PstI restriction site of pBR322 (Konecki et al., 1982).

pHPT4: this construct also consists of the full-length mouse HPRT cDNA cloned into the PstI restriction site of pBR322. It differs from pHPT5 in that it lacks polyA sequences (Konecki et al., 1982).

plasmid 91 (mouse α-actin cDNA): 1350bp of the mouse α-actin cDNA inserted into the PstI site of pBR322. This sequence will hybridise to two tissue mRNA bands in Northern analyses. A band at approximately 2100 nucleotides represents cytoplasmic (β and γ) actin mRNA. A band at approximately 1650 nucleotides represents muscle (α) actin mRNA (Minty et al., 1981). This plasmid was a gift from Kathy Pritchard-Jones, MRC Human Genetics Unit, Western General Hospital, Edinburgh.

## 2.2 Methods

### 2.2.1 E.coli culture

E.coli cells carrying plasmids were cultured in L-broth plus antibiotic where appropriate, at 37°C with constant shaking.

### 2.2.2 ES cell methods

General culturing: ES cells were grown in 37°C incubators

with 5% $CO_2$ in air. Tissue culture flasks were gelatin-coated before use. ES cells were grown in 60% BRL-conditioned medium, 40% fresh medium, containing 0.1mM β-mercaptoethanol. The medium was changed each day and cultures were trypsinised (0.025% (w/v) trypsin, 1mM$Na_2$ EDTA, 1% (v/v) chick serum in PBS) and split every 3 to 4 days.

Electroporation: vector DNA was introduced into ES cells by electroporation (Bio-Rad Gene Pulser). Approximately 5 x $10^6$ cells were resuspended in 0.8ml of HEPES-phosphate buffered saline (pH7.05) (Graham and van der Eb, 1973) with 20 μg of linearised vector DNA. Cells were given a single pulse (800V, path length 0.4cm, 3μF) then allowed to stand at room temperature for 10 minutes before plating in non-selective medium. HAT selection (Littlefield, 1964) was imposed after 24 hours.

### 2.2.3 Protein Methods

Isolation of soluble proteins from cultured cells and mouse tissues

Adapted from Wahl et al., (1975).

Lysis buffer: 10mM Tris.HCl pH7.5, 30mMKCl, 10mM $MgCl_2$, 1mM Dithiothreitol (DTT), 0.5% Triton X-100. This buffer was kept on ice.

The following procedure was carried out in a 4°C ('cold') room. Medium was aspirated from cell culture plates. Cells were washed twice in PBS, and plates were left to drain. 0.5ml of lysis buffer was washed over the surface of the cell culture and then left for 20 minutes with regular agitation. The lysate was transferred to a microcentrifuge tube and cell debris was removed by spinning in a microcentrifuge (17,000 x g) for 10 minutes at 4°C. The supernatant was transferred to a new microcentrifuge tube and stored at -20°C.

Mouse tail biopsies: approximately 1cm of tail was mechanically disrupted in 0.125ml lysis buffer and kept on ice for 1 hour. Tissue debris was pelleted by spinning in a microcentrifuge for 10 minutes at 4°C. The supernatant was stored at -20°C.

Protein gel electrophoresis

Protein lysate samples were electrophoresed through nondenaturing discontinuous polyacrylamide gels.

Separating gel solution: 7.5% acrylamide, 0.25% N, N'-methylene-bisacrylamide (bisacrylamide), 5 μg/ml riboflavin (Vit.$B_{12}$), 100mM Tris.HCl.pH8.1. 40ml of this solution was polymerised by adding 0.3ml 10% ammonium persulphate and 0.02ml TEMED.

Stacking gel solution: 2.5% acrylamide, 0.625% bisacrylamide, 5 μg/ml riboflavin, 60mM Tris.HCl pH6.7, 20% sucrose. 16ml of this solution was polymerised by adding 0.2ml 10% ammonium persulphate and 0.012ml TEMED.

Well-forming gel: 5% acrylamide, 0.16% bisacrylamide, 125mM Tris.HCl pH7.0. 20ml of this solution was polymerised by adding 0.15ml 10% ammonium persulphate and 0.010ml TEMED.

The separating gel (16cm x 16cm x 0.15cm) was cast between glass plates and overlaid with butanol. After polymerisation the butanol was removed, and the stacking gel solution was poured to a depth of 3cm. After polymerisation the well-forming gel solution was added and the comb was inserted. The gel was assembled into a vertical gel kit.

Sample buffer: 166mM Tris.HCl pH6.7, 25% sucrose, 0.0013% bromophenol blue (BPB).

20 μl of a protein lysate was added to 30 μl of sample buffer. The mixture was mixed, spun in a microcentrifuge for 5 minutes at 4°C and loaded onto a vertical gel.

10 x Electrophoresis buffer: 0.1M Tris (hydroxymethyl)-methylamine, 0.75M glycine, 0.15% β-mercaptoethanol.

Electrophoresis was carried out in 1 x buffer at 4.5V/cm for 16 hours, or until the BPB dye front reached the end of the gel, at 4°C.

HPRT in situ enzyme assay

A modification of the method of Epstein (1970).

Enzyme assay solution: 2mM Na 5-phosphoribosyl-1-pyrophosphate (PRPP), 50mM Tris.HCl pH7.4, 5mM Mg $Cl_2$, 25 µCi/ml [$^3$H]-hypoxanthine ( 5 Ci/mmol, Amersham).

After electrophoresis gel plates were separated and the separating gel was overlaid with 0.6ml of the assay solution. The product of the reaction, [$^3$H] - 5' - IMP was separated from the substrates using DEAE Ion exchange chromatography paper (DE81, Whatman). This paper was soaked in water and laid over the gel and assay solution. The reaction was incubated in a humid incubator at 37°C for 2 hours. The Ion exchange chromatography retains the charged [$^3$H] - 5' - IMP ions, but uncharged [$^3$H] -hypoxanthine was washed from the paper by three rinses in 50% (v/v) methanol/water. The dried chromatography paper was exposed to autoradiographic film for 3 days at -70°C in the presence of intensifying screens.

### 2.2.4 DNA and RNA methods

Large scale preparation of plasmid DNA

L-broth (5ml) supplemented with the appropriate selecting

antibiotic was inoculated with a single colony of the plasmid-bearing strain. The culture was grown to stationary phase and then diluted into 30ml of L-broth plus antibiotic. This culture was grown for 3 to 4 hours and then diluted to 500ml of L-broth plus antibiotic, and grown for a further 3 hours. Cells were harvested by centrifugation at 4,000 x g for 10 minutes and then resuspended in 2.5ml of sucrose mix (50mM Tris.HCl pH8.1, 40mM EDTA, 25% (w/v) sucrose). Lysozyme (1ml of 10mg/ml) and 0.5M EDTA (1ml) were added, the suspension was mixed gently and left on ice for 10 minutes. Then 13ml of Triton mix (50mM Tris.HCl pH 8.1, 63mM EDTA, 0.1% (v/v) Triton X-100) was added and mixed by swirling before returning to the ice for a further 10 minutes. The lysate was cleared by centrifugation at 23,000 x g for 30 minutes at 4°C. The supernatant was then mixed thoroughly but gently with 15ml Phenol/STE (redistilled phenol equilibrated with 10mM NaCl, 10mM Tris. HCl pH8.0, 1mM EDTA) and centrifuged at 4,000 x g for 10 minutes. DNA was precipitated from the aqueous layer by adding 0.1 x volume of 3M sodium acetate and 2.5 x volume absolute ethanol, and leaving for at least 1 hour at -70°C. The precipitate was pelleted by centrifugation at 14,000 x g for 20 minutes. The supernatant was discarded and the pellet was resuspended in 12ml TE (10mM Tris.HCl pH8.0, 1mM EDTA). 19g of CsCl was added to the DNA suspension and the volume was made up to 26ml with TE. 100 µl of ethidium bromide (10mg/ml) was added and the density was adjusted to between 1.55 and 1.59 g/ml. Supercoiled plasmid DNA was separated from chromosomal and nicked plasmid DNA by isopycnic ultracentrifugation at

38,000 rpm in a Sorvall Ti50 rotor (140,000 x g) for 48 hours at 20°C. DNA bands were visualised under long-wave U.V. illumination and the lower (supercoiled plasmid) band removed. Ethidium bromide was removed by repeated extraction with an equal volume of butanol, and subsequent dialysis against TE, at 4°C. DNA was recovered by precipitation with sodium acetate and ethanol.

Alkaline lysis plasmid miniprep method

Adapted from the method of Birnboim and Doly (1979). 5ml of L-broth plus antibiotic was inoculated with a single colony of the plasmid-carrying strain, and grown to stationary phase (overnight). 1.5ml of this culture was transferred to a microcentrifuge tube and spun in a microcentrifuge for 15 seconds. The cell pellet was resuspended in 0.1ml of 50mM glucose, 10mM EDTA, 25mM Tris. HCl pH8.0 solution and incubated at room temperature for 15 minutes. 0.2ml of 1% SDS, 0.2M NaOH solution was then added, mixed by inversion, and kept on ice for 5 minutes. Chromosomal DNA was removed by precipitation with 0.15ml of 3M sodium acetate pH5.0 on ice for 30 minutes. The precipitate was pelleted by spinning in a microcentrifuge (17,000 x g) for 15 minutes. The supernatant was transferred to a fresh microcentrifuge tube and spun for a further 10 minutes as above. To the supernatant was added 2.5 x volume of absolute ethanol (-20°C). The mixture was vortexed and kept at -70°C for 30

minutes followed by spinning in a microcentrifuge for 6 minutes. The pellet was washed twice with 70% (v/v) ethanol/water.

Preparation of high molecular weight DNA from cultured mammalian cells

The method of Pellicer et al., (1978). Between $10^7$ and $10^8$ cells were scraped from culture dishes in 20ml of PBS at 4°C. Cells were pelleted by centrifugation in a bench top centrifuge at 1300 rpm. The cell pellet was resuspended in 10ml PBS at 4°C and pelleted again as above. This pellet was resuspended gently in 10ml hypotonic solution (10mM Tris.HCl pH8.0, 10mM NaCl, 3mM $MgCl_2$) and then spun at 1300 rpm for 5 minutes. The pellet obtained was resuspended in 10ml of hypotonic solution plus Triton (0.2% Triton X-100) which lyses cells but not nuclei, and kept on ice for 10 minutes. The suspension was centrifuged at 1300 rpm for 5 minutes in a bench top centrifuge. The pellet was resuspended in 7.6ml of 10mM Tris. HCl pH8.0, 400mM NaCl, 10mM EDTA. To this was added 0.4ml of 10% SDS and 0.1ml proteinase k (30mg/ml) and the mixture was incubated overnight at 37°C. 0.2ml of RNaseA (2mg/ml) was then added and incubated at 37°C for 1 hour. More proteinase k (0.1ml of 30mg/ml stock) was added and incubated for a further hour at 37°C. The sample was then extracted with an equal volume of PCA (25 parts redistilled phenol:24 parts chloroform:1 part iso-amylalcohol) by gently mixing for 15 minutes and spinning in a bench top centrifuge

at 4,000 rpm for 15 minutes. DNA was precipitaed from the aqueous layer by adding 0.25 x volume 5M NaCl and 2 x volume absolute ethanol (-20°C). The DNA precipitate was removed by spooling around the sealed end of a Pasteur pipette, and washed in 70% (v/v) ethanol/water. The DNA was then dried, resuspended in $H_2O$ and stored at 4°C.

Preparation of high molecular weight DNA from mouse tails

Mice were anaesthetised by interperitoneal injection with 2.5% avertin. Avertin (1,2,3 - tribromoethanol) was dissolved at a concentration of 1g/ml in 3° amylalcohol. This 100% stock was diluted to 2.5% in $H_2O$. All solutions were stored at 4°C. Mice were injected with between 0.015 and 0.017ml per gram of body weight.

1 to 2cm of tail was cut from anaesthetised mice, using a red hot scalpel which simultaneously cauterized the cut end of the tail.

DNA was prepared by a method adapted from that of Palmiter et al., (1982).

Tail buffer: 0.3M sodium acetate (not titrated), 10mM Tris. HCl pH8.0, 1mM EDTA, 1% SDS.

1 to 2cm of mouse tail was placed in a 20ml sterilin plastic universal containing 1.5ml of tail buffer plus 15 µl of

proteinase k solution (final concentration 0.2mg/ml) and incubated overnight at 37°C with gentle shaking. Subsequently, each sample was split into two 0.75ml aliquots in 1.5ml microcentrifuge tubes. One aliquot was stored at -20°C, the other was used to prepare DNA. The suspension was mixed thoroughly but gently with an equal volume of PCA (25 parts redistilled phenol:24 parts chloroform:1 part isoamylalcohol). Following a 5 minute spin in a microcentrifuge the aqueous (top) layer was transferred to a fresh tube and the PCA extraction was repeated. The aqueous layer was then extracted with an equal volume of CA (24 parts chloroform:1part isoamylalcohol). 1/30 x volume of 3M sodium acetate pH5.0 and 1 x volume of isopropanol was added to the aqueous layer to precipitate DNA, which was then pelleted by spinning in a microcentrifuge for 30 seconds. The pellet was washed twice with 70% ethanol, dried, resuspended in 0.2ml $H_2O$ and stored at -20°C.

Preparation of RNA from cultured mammalian cells or mouse tissues

A modification of the method described in Melton et al., (1981).

Guanidine hydrochloride (GnHCl) solution: 6M guanidine hydrochloride, 10mM DTT, 25mM EDTA pH7.0.

Between $10^7$ and $10^8$ cultured cells were washed and pelleted

in PBS, before homogenisation in 10ml of GnHCl solution. Mouse tissue was homogenised in between 20 and 30ml of GnHCl solution. Removal and homogenisation of mouse tissue was performed as quickly as possible after death. In a fresh 50ml tube 1/30 x volume of 3M potassium acetate pH5.0 and 1/2 x volume of absolute ethanol was added to each sample which was then left at -20°C for 4 hours. The precipitate was pelleted by spinning at 18,000 x g for 20 minutes in a refrigerated centrifuge (0°C). The pellet was resuspended in GnHCl solution, precipitated and pelleted again as above. This pellet was resuspended in 0.1M Tris. HCl pH8.9, 0.1M NaCl, 1mM EDTA, 1% SDS solution (3 to 5ml for cultured cells and 5 to 10ml for tissues) by vigorous vortexing. An equal volume of PCA was added and vortexed thoroughly for 15 minutes before spinning in a refrigerated bench top centrifuge for 15 minutes at 3000 rpm and 10°C. To the aqueous layer was added 1/15 x volume of 3M sodium acetate pH5.0 and 2 x volume of absolute ethanol. This was left to precipitate for 4 hours at -20°C and then spun for 20 minutes at 18,000 x g in a refrigerated centrifuge (0°C). The pellet was dissolved in the minimal possible volume of $H_2O$ (approximately 1.5ml for cultured cells and between 0.5 and 1.0ml for mouse tissues) and split into 0.5ml aliquots in microcentrifuge tubes. To each aliquot 1/10 x volume of 3M sodium acetate pH5.0 and 2 x volume of absolute ethanol was added. Samples were left at -70°C for 15 minutes and then spun in a microcentrifuge for 5 minutes. The RNA pellet was washed twice in 70% ethanol, dried and resuspended in the minimal volume of $H_2O$. RNA solutions were stored at -70°C.

The concentration of RNA and DNA suspensions was estimated by spectrophotometric analysis and comparison of samples in ethidium bromide agarose gels, visualised by U.V. illumination.

Restriction endonuclease digestion

Enzymes were purchased from Bethesda Research Laboratories (BRL). DNA was digested for between 1 and 3 hours at 37°C, with the exception of TaqI digestions which were carried out at 65°C.

BRL restriction enzyme buffers were used at 1 x concentration. RNase A was added to a final concentration of 0.1mg/ml. Reactions were terminated by heat inactivation or phenol extraction.

10 x React 2 Buffer: 50mM Tris. HCl pH8.0, 10mM $MgCl_2$, 50mM NaCl. Used for PstI, Hind III, TaqI and XhoI.

10 x React 3 Buffer: 50mM Tris. HCl pH8.0, 10mM $MgCl_2$, 100mM NaCl. Used for BamHI and EcoRI.

10 x React 6 Buffer: 50mM Tris. HCl pH7.4, 6mM $MgCl_2$, 50mM KCl, 50mM NaCl. Used for PvuII.

Agarose gel electrophoresis

Electrophoresis through horizontal submerged agarose gels was used for both analysis and purification of DNA. Ultrapure agarose (BRL) was used for analytical gels and low melting-point agarose (BRL) was used for preparative gels.

Agarose gels: 0.8 to 1.5% (w/v) agarose, 1 x TBE,0.5 µg/ml ethidium bromide.

10 x TBE: 0.89M Tris. base, 0.89M boric acid, 0.02M EDTA.

Gels were run in 1 x TBE electrophoresis buffer.

Samples were mixed with 0.2 x volume of sample buffer (100mM EDTA, 0.1% BPB, 20% Ficoll) and heated at 65°C for 5 minutes before loading. DNA was visualised by illumination with long-wave U.V. light.

DNA fragment purification from low melting-point agarose gels

The DNA band was cut from the low melting-point agarose gel with the minimal volume of agarose, and added to an equal volume of 1 x TBE, 0.2M NaCl solution. The agarose was melted by heating in a 65°C water bath. Two phenol extractions with redistilled phenol equilibrated with 1 x TBE, 0.1M NaCl were performed. The aqueous layer was extracted with 4 x volume of butanol repeatedly, until the volume of the aqueous layer

was reduced to approximately 0.4ml. DNA was precipitated by adding 0.1 x volume of 3M sodium acetate and 2.5 x volume of absolute ethanol (-20°C). DNA was pelleted by spinning in a microcentrifuge (17,000 x g) for 6 minutes. The pellet was resuspended in 0.1ml of 0.3M sodium acetate and 0.25ml of absolute ethanol was added to precipitate DNA for a second time. The DNA pellet was washed twice with 70% ethanol, dried and resuspended in $H_2O$. DNA solutions were stored at -20°C.

Radiolabelling DNA by random priming

This method (Feinberg and Vogelstein, 1983) was used to make probes for both Southern and Northern hybridisations. Approximately 2µg of restriction fragment DNA, purified from low melting-point agarose, was used in each reaction. DNA was denatured before use by boiling for 5 minutes and then incubating for 10 minutes at 37°C. The reaction mixture was made by adding in the following order: 10µl OLB buffer, 2µl bovine serum albumin (BSA, 10mg/ml), 32 µl denatured DNA solution, 5 µl (50µCi) α - [$^{32}$P] - dCTP (3000 Ci/mmol, Amersham), 1 µl (5u) E.coli DNA polymerase I klenow fragment (BRL). The reaction was left for 4 hours at room temperature and stopped by phenol extraction.

OLB buffer: 50 µl solution A (1.25M Tris.HCl pH8.0, 0.125M $MgCl_2$, 25mM β-mercaptoethanol, 0.5mM each of TTP, dGTP and dATP), 125 µl solution B (2M HEPES (4-(2-hydroxyethyl)-1-

piperazine-ethanesulphonic acid) adjusted to pH6.6 with 5M NaOH) and 75 µl solution C (random hexanucleotides $OD_{260nm}$ 90, in TE).

The radiolabelled probe was purified from unincorporated deoxyribonucleotides by passing through a column of sephadex G-50-80 (Sigma).

Southern transfer hybridisations

Adapted from the method of Southern (1975).

20 x SSC: 3M NaCl, 0.3M sodium citrate, adjusted to pH7.0.

Genomic DNA (5µg) was digested by the appropriate restriction enzyme(s) (50 u) in 100 µl reaction volumes, for 5 hours. Samples were electrophoresed through 0.8% agarose gels. After electrophoresis the gel was soaked for 30 minutes with gentle agitation in 0.5M NaOH, 1.5M NaCl. DNA was then transferred to Gene Screen Plus (New England Nuclear Products) for 18 hours using 0.5M NaOH, 1.5M NaCl as the transfer buffer. After transfer the charged nylon membrane was neutralised by soaking in 0.5M Tris, 3.0M NaCl pH7.4 for 15 minutes and dried.

The membrane was pre-hybridised in 30 to 50ml of 6 x SSC, 1% SDS, 10% dextran sulphate (Sigma) with 100µg/ml of sonicated herring sperm DNA for 2 hours at 65°C in a sealed

hybridisation bag (BRL). After 2 hours the denatured radiolabelled DNA probe (approximately $10^6$ cpm per ml of hybridisation buffer) was added with sonicated herring sperm DNA to total concentration of 250 µg/ml. Hybridisation was for 18 hours at 65°C. Following hybridisation the membrane was washed, twice for 5 minutes in 2 x SSC at room temperature, twice for 30 minutes in 1% SDS, 2 x SSC at 65°C and finally twice for 30 minutes in 0.1 x SSC at room temperature. The damp filter was sealed into a polythene bag and exposed to preflashed autoradiographic film at -70°C in light proof casettes with intensifying screens (Dupont).

Northern transfer hybridisations

Adapted from the method of Lehrach et al., (1977).

10 x MOPS: 0.2M MOPS, 50mM sodium acetate, 10mM EDTA, adjusted to pH7.0 and autoclaved.

Formamide sample buffer (FSB): 2.5 x MOPS, 50% deionised formamide, 10% formaldehyde.

30 µg of total RNA in a volume of 20 µl was added to an equal volume of FSB, and heated at 65°C for 5 minutes before snap cooling on ice. 0.25 x volume of standard electrophoresis buffer (100mM EDTA, 0.1% BPB, 20% Ficoll) was added. Samples were electrophoresed through formaldehyde agarose gels (1.4% agarose, 1 x MOPS, 0.66M formaldehyde, 0.5 µg/ml ethidium

bromide) for 3 to 4 hours at 4V/cm. After electrophoresis gels were washed twice in 10 x SSC for 20 minutes. RNA was transferred to Gene Screen Plus (New England Nuclear Products) with 10 x SSC transfer buffer for 18 hours. After transfer the nylon membrane was rinsed in 2 x SSC, dried and baked for 2 hours at 80°C to reverse the formaldehyde reaction.

Subsequent hybridisation and washing was carried out as for Southern membranes except that the hybridisation temperature and the temperature of the high stringency wash was 60°C (rather than 65°C). Autoradiography was carried out at -70°C with preflashed film and intensifying screens.

Single-stranded DNA probes for S1 nuclease protection analyses

From the method of Berk and Sharp, (1977).

Before radiolabelling probe DNA was phosphatased (i.e. 5' terminal phosphates were removed).

10 x Phosphatase buffer: 650mM Tris.HCl pH7.5, 95mM $MgCl_2$, 50mM DTT.

Reactions were carried out in 100 µl with approximately 1µg restriction fragment DNA, 1 x phosphatase buffer and 125u of bacterial alkaline phosphatase enzyme (BRL). Reactions were

incubated for 1 hour at 65°C and stopped by thorough phenol extraction, butanol extraction and precipitation with sodium acetate and ethanol.

DNA recovered from one phosphatase reaction was end-labelled in a kinase reaction.

10 x oligonucleotide kinase buffer: 0.5M Tris. HCl pH7.5, 0.1M $McCl_2$, 50mM DTT, 1mM EDTA.

Phosphatased DNA was added to 1 x oligonucleotide kinase buffer, 0.1mM spermidine, 1mM $NaPO_4$ pH7.0, 50 µCi $\gamma$-[$^{32}$P] - ATP (5000 Ci/mmol, Amersham) and 10 u of T4 polynucleotide kinase (BRL) in a total volume of 50µl. The reaction was incubated at 37°C for 1 hour. The reaction was then phenol extracted, butanol extracted and precipitated with sodium acetate and ethanol. The labelled DNA pellet was resuspended in 10 µl of strand separating buffer (SSB; 30% dimethyl sulphoxide, 0.03% BPB, 0.05% xylene cyanol, 1mM EDTA, 10mM NaOH). The sample was incubated at 90°C for 3 minutes and loaded onto a vertical strand-separating acrylamide gel (8% acrylamide/bis-acrylamide 39:1, 1 x TBE). The gel was run for 18 hours at 10V/cm. The position of radiolabelled DNA was determined by exposing the gel to autoradiography film for 20 minutes. Both strands of the DNA fragment are labelled. The required band on the gel was determined previously by assaying each strand. Acrylamide containing the required (antisense) strand was cut from the gel and DNA was eluted from it in 3ml of elution buffer (10mM

Tris.HCl pH8.0, 50mM NaCL, 5mM EDTA) for 18 hours. The elution buffer was vortexed and filtered through glass wool in a 10ml syringe. The DNA suspension was phenol extracted, and butanol extracted to reduce the volume to 1ml.

S1 nuclease analysis

30 µg of total RNA from cultured cells or tissues was precipitated with the single-stranded DNA probe in a total volume of 0.4ml. All the single-stranded probe obtained in a kinase reaction was divided equally between the RNAs under investigation. The pellet was resuspended in 35µl of 2 x SET (0.3M NaCL, 40mM Tris.HCl pH8.0, 2mM EDTA). Samples were incubated at 90°C for 15 minutes and at 65°C for 4 hours. Then 0.3ml of ice cold S1 nuclease buffer (280mM NaCl, 50mM sodium acetate pH5.0, 4.5mM $ZnSO_4$) containing 50 u of S1 nuclease enzyme (BRL) was added to each sample. The reaction mixture was incubated at 37°C for 15 minutes. Reactions were stopped by adding 60 µl of 100mM EDTA, followed by phenol extraction and DNA precipitation with sodium acetate and ethanol. Pellets were resuspended in 6 µl of Maxam-Gilbert sequencing buffer (80% formamide, 0.03% BPB, 0.05% xylene cyanol, 100mM NaOH, 1mM EDTA). Each sample was incubated at 90°C for 10 minutes and loaded onto an 8.0% acrylamide urea sequencing gel. After electrophoresis at 30 W the gel was removed from the glass plates, sealed in Saran wrap and exposed to autoradiography film at -70°C with intensifying screens.

## Radiolabelled standards for S1 analyses

pBR322 DNA cut with BamHI and MspI was radiolabelled by filling in 5' overhangs.

10 x Nick translation buffer: 500mM Tris.HCl pH7.5, 100mM $MgCl_2$, 500µg/ml BSA.

DNA was incubated with 1 x Nick translation buffer, 0.2mM dATP, 0.2mM dGTP, 0.2mM TTP, 5 u E.coli DNA polymerase I klenow fragment (BRL) and 10µCi $\alpha$-[$^{32}$P] - dCTP (3000Ci/mmol, Amersham) for 1 hour at room temperature. Cold dCTP was added to a final concentration of 0.2mM and the reaction was incubated for a further 30 minutes. The reaction mixture was diluted with 50µl of Maxam-Gilbert sample buffer and stored at -20°C.

## Densitometric analysis

Densitometric analysis of autoradiographs was performed using a Shimadzu dual-wavelength thin-layer chromato scanner, model CS-930, according to the accompanying instruction manual. Linearity of densitometric analysis was confirmed using autoradiographs exposed to the same source for varying lengths of time.

## 3. HPRT TRANSGENIC MICE

A description of the transcriptional control of the mammalian X-linked HPRT gene, based on experiments involving the transformation of cultured cells with derivatives of the HPRT minigene, pDWM1 (Melton et al., 1986), has been presented in the Introduction of this thesis. To try to validate these findings *in vivo* and to determine the mechanism by which HPRT expression is elevated in brain tissue, mice transgenic for an HPRT minigene were produced. Transgenic mice expressing human HPRT have been produced previously (Stout et al., 1985). The transgene used consisted of the human HPRT cDNA under the transcriptional control of the mouse metallothionein-1 (MT-1) promoter and fused to the 3' untranslated region from the human growth hormone (hGH)gene. Three transgenic lines were obtained and analysed for human HPRT expression. The expression pattern in each transgenic line was different, presumably influenced by the integration site of the transgene. However, all mice examined had elevated human HPRT expression in the CNS. It is difficult to interpret this result because an analogous construct in which the HPRT cDNA was replaced by the human growth hormone gene also gave elevated expression in the CNS of transgenic mice (Swanson et al., 1985). This suggests the expression pattern is due to the combination of the MT-1 promoter and the hGH 3' untranslated region rather than the HPRT cDNA.

## The transgene

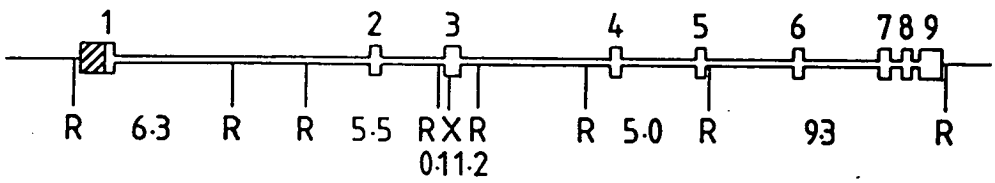
The aim of our experiment was primarily to determine whether the HPRT minigene contains sufficient DNA sequence to be expressed with the same tissue-specific variations as the endogenous gene. In particular, we wanted to know whether the minigene was expressed at high levels in brain. The recipient mice for the transgene have endogenous HPRT activity. Hence, it was necessary to use a minigene the expression of which would be distinguishable from that of the endogenous gene. To satisfy this requirement the vector pDWM1LS+127/+135 was used. The structure of pDWM1LS+127/+135 is illustrated schematically in Figure 3.1. pDWM1LS+127/+135 is derived from the basic HPRT minigene pDWM1 (Melton et al., 1986). It was constructed as a linker scanning mutant, and is composed of a 3' deletion of pDWM1Δ1 with deletion end point at +127 (number relative to the major transcription initiation site, +1), ligated through a 10bp BamHI linker to a 5' deletion of pDWM1Δ1 with deletion end point +135. pDWM1LS+127/+135 contains 635bp of mouse genomic sequence upstream of the major transcription initiation site. It contains the entire 5' and 3' untranslated regions of the mouse cDNA and the mouse protein coding region interrupted by the seventh and eighth introns only. The differences between the wild-type DNA sequence and the sequence of pDWM1LS+127/+135 at the BamHI linker are shown in Figure 3.2. These changes make expression from the transgene distinguishable from endogenous HPRT gene expression. The DNA sequence changes occur at the 5' end of

**Figure 3.1.**

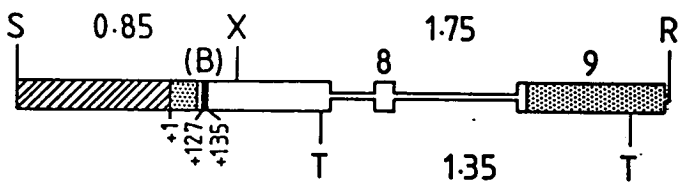
Structure of pDWM1LS+127/+135. The organisation of the X-linked gene is shown. Exons 1 to 9 (not to scale) are represented by open boxes and numbered. EcoRI, R, and XhoI, X, restriction sites are indicated and the size of each restriction fragment is shown in kb. The cross-hatched box represents the 635bp of 5' flanking sequence which is also present in pDWM1LS+127/+135. The structure of pDWM1LS+127/+135 is shown with selected restriction sites: S, SalI; X, XhoI; R, EcoRI; T, TaqI. Restriction fragment sizes are indicated in kb. The cross-hatched box represents the 635bp of 5' flanking sequence. The stippled boxes represent the 5' and 3' untranslated regions. The open boxes represent protein coding regions. The open lines represent introns 7 and 8. The position of the BamHI linker (B) is indicated by the solid vertical bar, between nucleotides +127 and +135 of the wild-type sequence. +1 indicates the major transcription initiation site.

Figure 3.1

HPRT Gene




pDWM1LS+127/+135



the protein coding region. Hence, protein derived from the transgene will have a different amino acid sequence for the amino-terminal residues. The mouse endogenous gene codes for a protein which has the amino-terminal amino acid sequence: methionine, proline, threonine, arginine, serine, proline, serine. The initiating methionine is post-translationally cleaved from the HPRT protein (Wilson et al., 1982). Whereas, the transgene codes for a protein which has the amino-terminal sequence: methionine, proline, threonine, proline, aspartic acid, proline, alanine, serine. That is, the three amino acids arginine, serine and proline which are basic, polar and non-polar respectively, are substituted for proline, aspartic acid, proline and alanine which are non-polar, acidic, non-polar and non-polar respectively. Consequently, the pDWM1LS+127/+135 encoded protein has a different isoelectric point and a higher mobility on a non-denaturing polyacrylamide gel than the wild-type protein at pH8.1.

pDWM1LS+127/+135 can transform HPRT deficient Chinese hamster,RJK88, cells to HPRT+ at a frequency of $10^{-4}$ per 5 g of plasmid DNA. The greater electrophoretic mobility of the transgene encoded protein is shown in Figure 3.3. Soluble proteins were purified from cultured cells and electrophoresed through non-denaturing 7.5% polyacrylamide gels. HPRT protein was then visualised by an in situ enzyme assay using [$^{3}$H]-hypoxanthine. The first lane contains protein derived from a wild-type mouse cell-line, RJK691. The second lane contains protein isolated from HPRT

**Figure 3.2.**

Sequence alterations in pDWM1LS+127/+135. The wild-type nucleotide sequence of the first seven translated codons is shown. Nucleotides +127 and +135 relative to the major transcription initiation site are indicated. The sequence of the seven N-terminal amino acids is shown. The initiating methionine is post-translationally cleaved. The nucleotide sequence of the corresponding region in pDWM1LS+127/+135 is shown, with the amino acid sequence directly below. The 10bp BamHI linker is underlined. Amino acids: Ala, alanine; Arg, arginine; Asp, aspartic acid; Met, methionine; Pro, proline; Ser, serine; Thr, threonine.

Figure 3.2

Wild-type

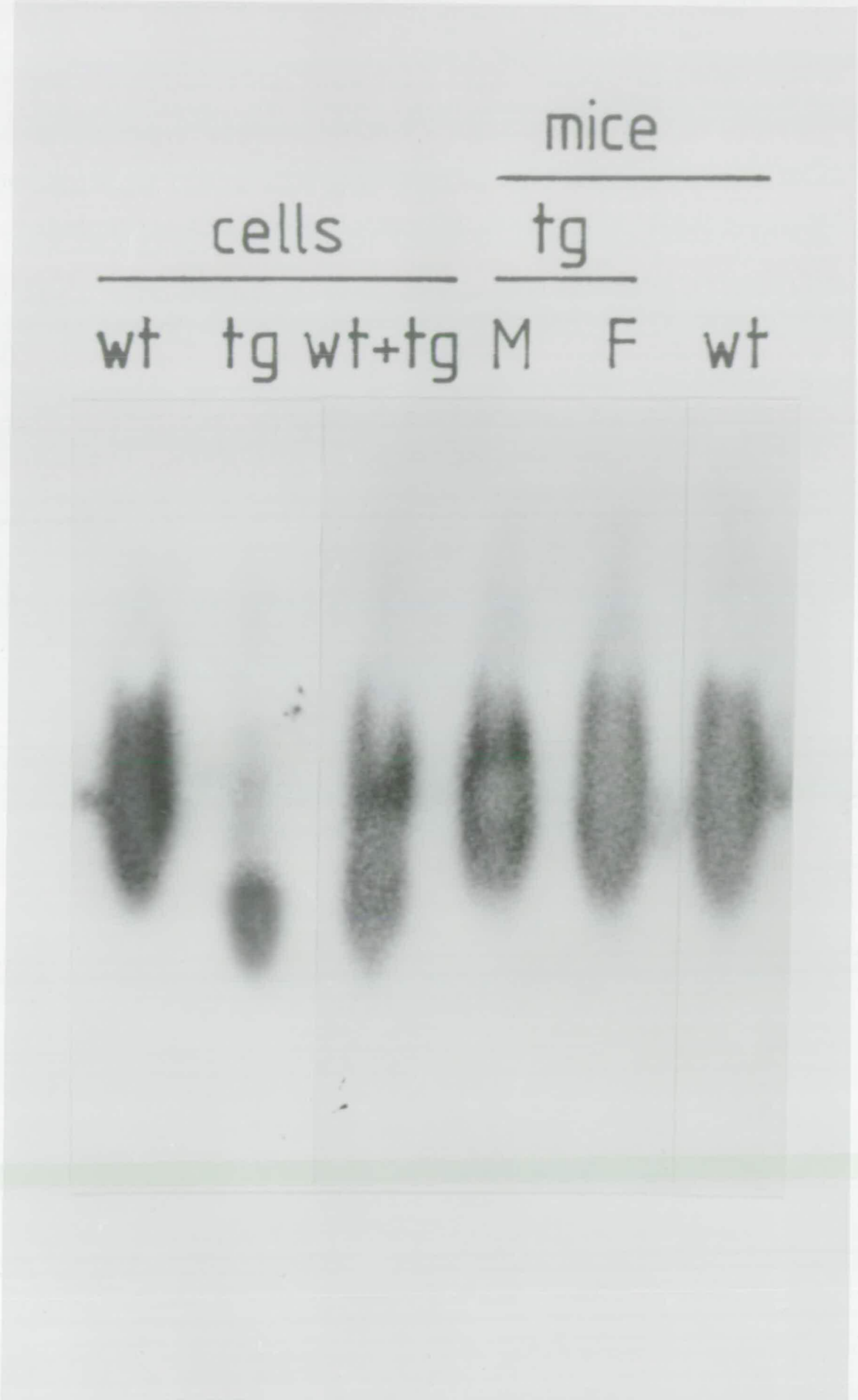
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DNA: | ATG | CCG | ACC$^{+127}$ | CGC | AGT | CCC$^{+135}$ | AGC | . . . |
| protein: | [Met.] | Pro. | Thr. | Arg. | Ser. | Pro. | Ser. | . . . |

pDWM1LS+127/+135

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| DNA: | ATG | CCG | ACC$^{+127}$ | CCG | GAT | CCG | GCC$^{+135}$ | AGC | . . . |
| protein: | [Met.] | Pro. | Thr. | Pro. | Asp. | Pro. | Ala. | Ser. | . . . |

**Figure 3.3.**

Non-denaturing electrophoresis of proteins. Proteins were electrophoresed through non-denaturing polyacrylamide gels and HPRT enzyme was visualised using an in situ enzyme assay with [$^{3}$H]-hypoxanthine. Protein samples from cells and mice are shown. Cells: wt, wild-type mouse cell line (RJK691); tg, HPRT deficient Chinese hamster cells (RJK88) transformed with pDWM1LS+127/+135; wt + tg, a 50% (v/v) mixture of protein from the wild-type mouse cell line and HPRT deficient Chinese hamster cells transformed with pDWM1LS+127/+135. Mice: protein samples from a male (M) and female (F) transgenic (tg) mouse are shown alongside protein from a wild-type mouse. Protein samples were obtained from tail biopsies.


mice
cells
tg
wt
tg
wt+tg
M
F
wt

deficient Chinese hamster, RJK88, cells transformed with pDWM1LS+127/+135. Non-transformed RJK88 cells have no detectable endogenous HPRT activity (data not shown). The third lane contains a 50% (v/v) mixture of RJK691 protein and pDWM1LS+127/+135-transformed RJK88 protein. Thus enzyme activity derived from pDWM1LS+127/+135 is distinguishable from enzyme activity derived from the endogenous gene in cultured cells by its different electrophoretic mobility.

It should also be possible to measure expression from pDWM1LS+127/+135 by analysis of mRNA. Because of the differences in DNA sequence between pDWM1LS+127/+135 and the endogenous gene, mRNA derived from the two genes can be distinguished by S1 nuclease protection analysis using a single-stranded DNA probe that spans the 5' end of the mRNA. The basis of this analysis is explained schematically in Figure 3.4. The results of such an analysis are shown in Figure 3.5. The DNA probe used was the 424bp BamHI-XhoI fragment from the -146 5' deletion mutant (Melton et al., 1986). The single-stranded DNA probe was end-labelled at the XhoI site (5' end) 271bp downstream of the major transcription initiation site. Hence mRNA from the wild-type mouse cell line RJK691 produces a protected fragment of 271bp (+1). Two additional protected fragments of approximately 30 bases shorter represent downstream initiation sites at +28 and +34 (Melton et al., 1986). The 424bp fragment represents undigested probe which is preserved probably due to the ability of this fragment to protect itself. Other feinter bands are apparently

**Figure 3.4.**

Schematic representation of the use of S1 analysis to distinguish between mRNA derived from the HPRT transgene (b) and mRNA derived from the X-linked endogenous HPRT gene (a). Endogenous HPRT mRNA, transcribed from nucleotide +1, produces a protected fragment of 271bp. Endogenous HPRT transcribed from nucleotides +28 and +34 produces a protected fragment 28 and 34bp shorter, respectively. The radiolabelled end of the single-stranded (ss) DNA probe is indicated by an asterisk. Transgene derived mRNA should produce a protected fragment of 136bp irrespective of the site of transcription initiation, due to the presence of the mismatched sequence at the BamHI linker (B).

Figure 3.4

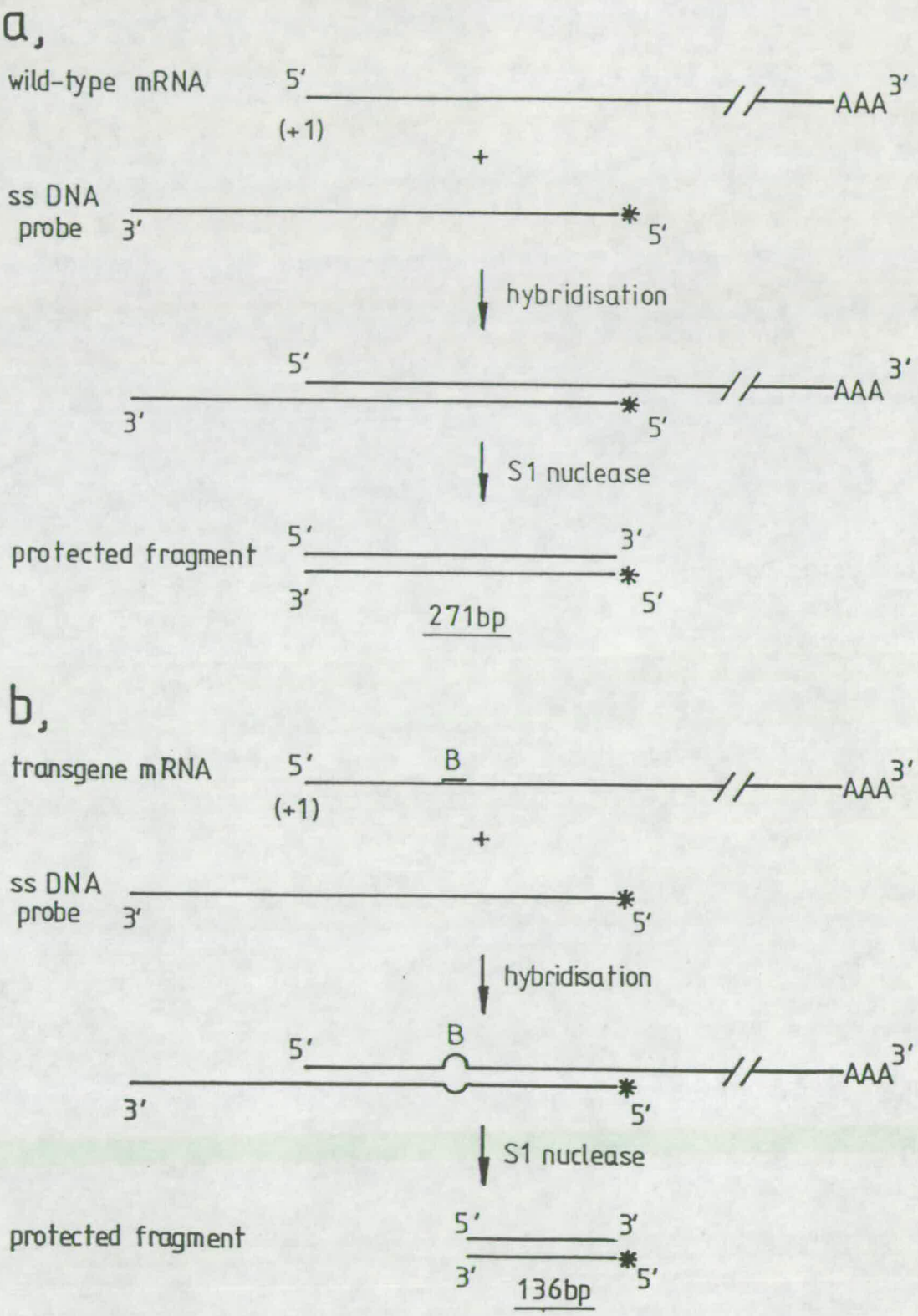
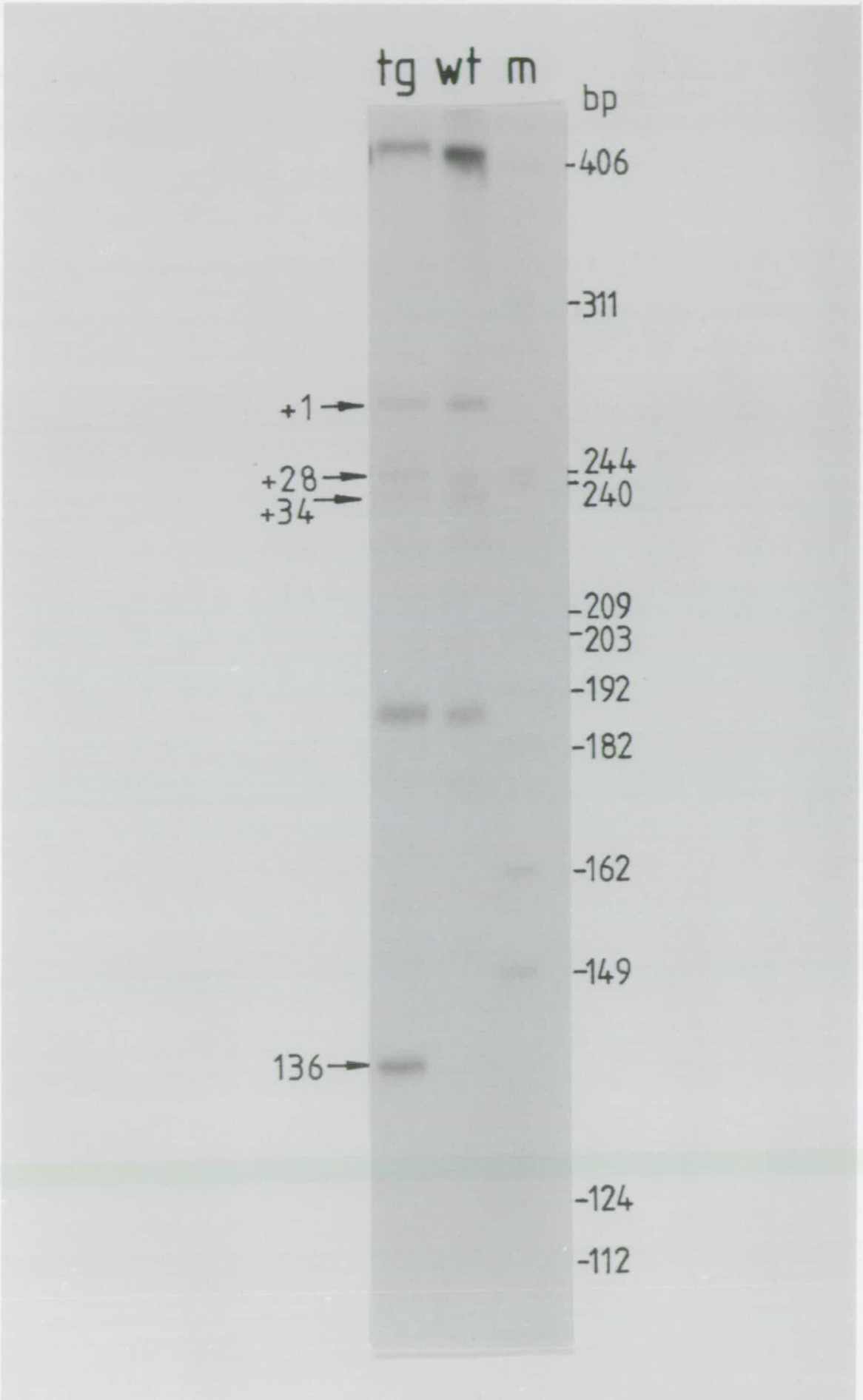

a,
wild-type mRNA
5'
AAA 3'
(+1)
+
ss DNA probe
3'
5'
hybridisation
5'
AAA 3'
3'
5'
S1 nuclease
protected fragment
5'
3'
3'
5'
271bp
b,
transgene mRNA
5'
B
AAA 3'
(+1)
+
ss DNA probe
3'
5'
hybridisation
5'
B
AAA 3'
3'
5'
S1 nuclease
protected fragment
5'
3'
3'
5'
136bp

**FIGURE 3.5.**

S1 analysis of transgene-derived mRNA from cultured cells. Cellular RNA was hybridised to a 424 base single stranded probe, which spanned the main transcription initiation site and was end-labelled at an XhoI site, 271 bases into the gene. Hybridisation mixtures were digested with S1 nuclease. Lanes: tg, RNA derived from HPRT deficient Chinese hamster (RJK88) cells transformed with pDWM1LS+127/+135; wt, RNA from wild-type mouse (RJK691) cells; m, molecular weight standards. The size of each molecular weight standard is indicated to the right of the panel. To the left of the panel, protected fragments representing the initiation sites +1, +28 and +34, and the 136 base protected fragment specific to transgene RNA, are indicated. Undigested probe (424 bases) is evident in both RNA tracks.


tg wt m
bp
406
311
+1
244
+28
240
+34
209
203
192
182
162
149
136
124
112

artefacts, as they do not consistently appear in S1 nuclease analyses and have no equivalent in primer extension analyses. That is, no products are seen in primer extension analyses which could correspond to initiation at the sites suggested by these bands. RNA derived from RJK88 cells that are transformed with pDWM1LS+127/+135 produces a different S1 protection pattern. In addition to the bands representing the transcription initiation sites at +1, +28 and +34 there is a band of approximately 136 bases long. This is the distance between the XhoI-site end of the DNA probe (+271) and the BamHI linker (+127 to +135) in pDWM1LS+127/+135 (see Figure 3.4). The protected fragment of 136 bases, therefore, represents S1 nuclease digestion at the region of mismatch between the pDWM1LS + 127/+135 derived mRNA and the single-stranded DNA probe, which has the wild-type sequence. Hence, the +136bp band is specific for pDWM1LS + 127/+135 derived mRNA. Unfortunately, not all of the transgene derived mRNA-DNA probe heteroduplexes are digested to completion at the mismatch region around the BamHI linker. Consequently, the transgene derived mRNA contributes to the protected signals which represent the transcription initiation sites. Therefore, it is not possible to quantify the transgene derived mRNA relative to the endogenous HPRT mRNA using this assay.

Production of HPRT transgenic mice

I am grateful to J. Paul Simons, of the Agricultural and Food

Research Council, Institute of Animal Physiology and Genetics Research, Edinburgh Research Station, for carrying out the microinjection experiment.

pDWM1LS+127/+135 DNA was digested with SalI and EcoRI and the 2.6kb fragment containing HPRT sequences was isolated from plasmid DNA by electrophoresis and purification from low melting-point agarose. A number of studies have shown that prokaryotic vector sequences can be inhibitory to transgene expression (Jaenisch, 1988; and references therein). Pronucleus stage eggs were obtained from superovulated C57BL/6 x CBA $F_1$ females, after mating with $F_1$ males, and injected with approximately 200 copies of the transgene per egg. Injection was into either pronucleus. Injected eggs (200) were cultured overnight and subsequently cleaved embryos (130) were transferred into the oviducts of pseudopregnant MF1 females, obtained by mating with vasectomised MF1 males. GO mice were screened for the presence of the transgene by Southern analysis of tail biopsy genomic DNA. Out of 27 GO mice weaned only one, female mouse 22, appeared to have the pDWM1LS+127/+135 gene integrated intact. Figure 3.6 shows a comparison of the Southern blot patterns of a wild-type mouse and female mouse 22. Genomic DNA was digested with EcoRI and XhoI, and hybridised with a full-length HPRT cDNA probe, pHPT5 (Konecki et al., 1982). The EcoRI-XhoI double digest generates a 1.75kb fragment (Figure 3.7) in the pDWM1LS+127/+135 sequence which is specific to the transgene. The sizes of transgene specific bands are indicated down the left hand side of Figure 3.6.

**Figure 3.6.**

Southern analysis of transgenic mice. Genomic DNA (5μg), obtained by tail biopsy, was digested with EcoRI and XhoI, electrophoresed, transferred and hybridised to a full-length HPRT cDNA probe, pHPT5. DNA: tg, a mouse transgenic for pDWM1LS+127/+135; wt, a wild-type mouse. The sizes (in kb) of endogenous gene fragments are indicated to the right of the panel. A fragment containing HPRT pseudogene sequences is indicated (PG). The sizes of transgenic mouse specific fragments are indicated to the left of the panel.

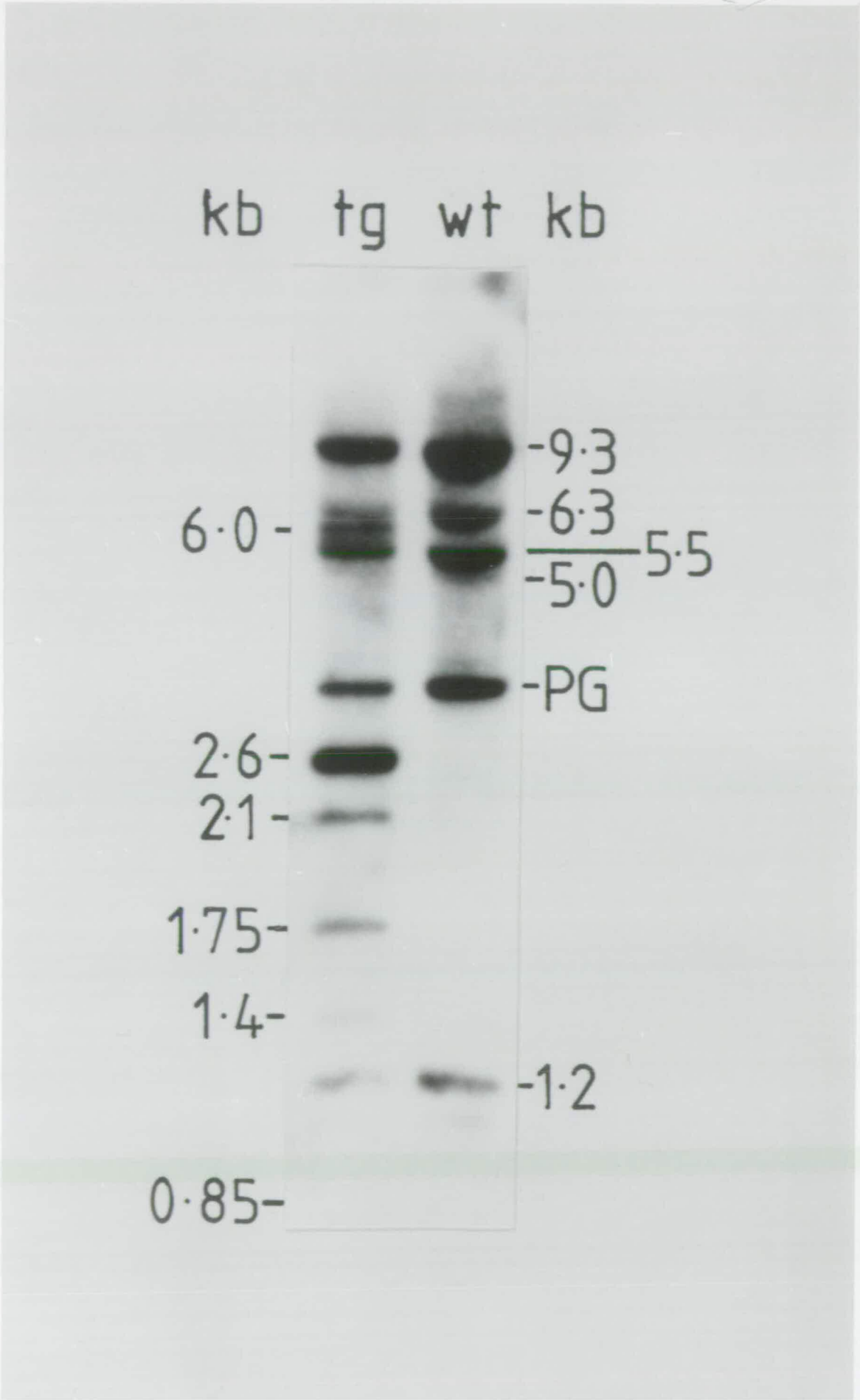
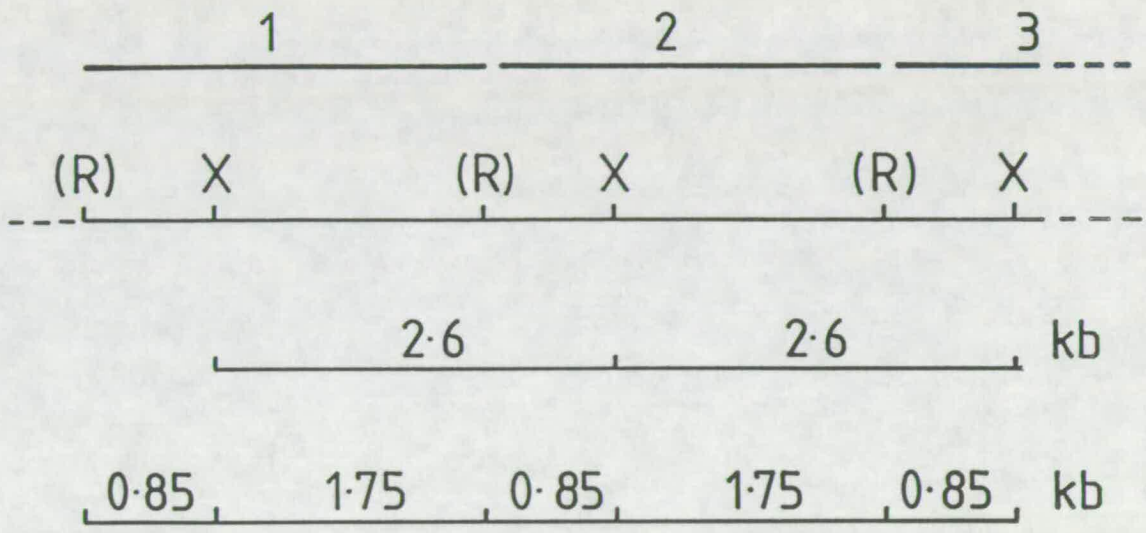

kb
tg
wt
kb
-9·3
-6·3
6·0-
5·5
-5·0
-PG
2·6-
2·1-
1·75-
1·4-
-1·2
0·85-

**Figure 3.7.**

Schematic representation of the tandem array of pDWM1LS+127/+135 transgenes. The top line shows the position of three transgenes in a head-to-tail tandem array. The second line shows the restriction sites for EcoRI, R, and XhoI, X. EcoRI sites are in parentheses because they are present only at some of the transgene:transgene junctions. The two lower lines illustrate the origins of transgene specific fragments observed in Southern analyses, see Figure 3.6.

Figure 3.7



The pattern observed is consistent with several copies of the transgene having integrated in tandem (Figure 3.7). At some transgene-transgene junctions the EcoRI site at the 3' end of the transgene is preserved, at other junctions it is lost. Hence a band of 850bp, which represents the sequences between the internal XhoI site and the 5' end of transgene (Figure 3.7), is generated from some copies. Where the EcoRI site is not maintained an XhoI-XhoI fragment of 2.6kb (the length of the transgene) is produced. Southern blot analysis of XhoI digested genomic DNA from mouse 22 produces a hybridising band of 2.6kb (data not shown). Because 2.6kb is the size of the transgene, all the copies of pDWM1LS+127/+135 must be integrated head to tail in tandem. Head to head and tail to tail junctions would generate bands of 1.7kb and 3.5kb respectively. Bands of these sizes were not seen.

As well as mouse 22, two additional G0 mice were obtained that contained exogenous HPRT sequences. One of these mice had died within a few hours of birth. Southern analysis (not shown) of EcoRI-XhoI digested genomic DNA showed that it contained the 1.75kb fragment from the 3' end of the transgene (Figure 3.1), although the 5' end of the transgene appeared to be absent. The transgene had undergone structural rearrangment. The second mouse containing exogenous HPRT sequences had a single additional band that hybridised to HPRT cDNA. However, the size of this band was not reconcilable with the structure of the transgene.

The transgenic female mouse 22 was tested for transmission of

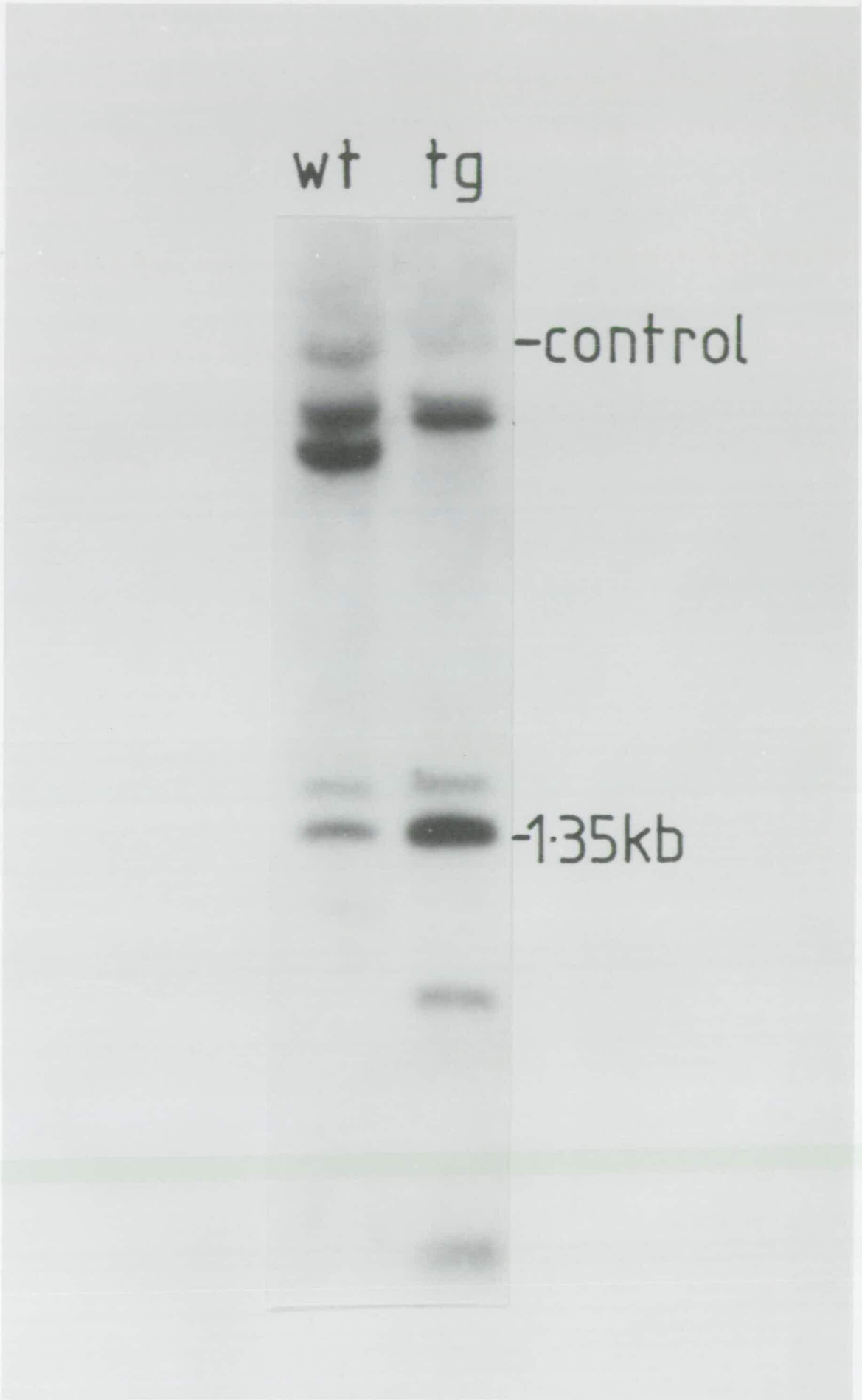
the pDWM1LS+127/+135 sequences by mating to C57BL/6 x CBA $F_1$ males. Offspring from mouse 22, G1 mice, were screened by Southern blot analysis of EcoRI-XhoI digested genomic DNA obtained from tail biopsies. Of 23 G1 mice, one male, 22.5, and two females, 22.4 and 22.17, had the identical Southern blot pattern to mouse 22 (Figure 3.6), demonstrating inheritance of the HPRT transgene. Inheritance of exactly the same Southern blot pattern is consistent with there being a single transgene integration site in this line. Considerably less than 50% (3 out of 23) of G1 mice inherited the transgene indicating that mouse 22 was probably mosaic. Male transgenic mouse 22.5 was used to produce further generations to provide material for expression studies. Three matings between 22.5 and wild-type females produced G2 offspring which were screened for the presence of pDWM1LS+127/+135 DNA. Twelve mice, eight females and four males, out of twenty three were transgenic. Transmission of the transgene from male parent to male progeny demonstrated that the integration site was not X-chromosome linked.

Transgene Copy Number

The number of copies of the pDWM1LS+127/+135 transgene that had integrated in mouse 22 was estimated by a comparison of the Southern blot patterns for TaqI digested genomic DNA from a wild-type and a transgenic animal. TaqI digestion of pDWM1LS+127/+135 DNA generates an internal fragment of 1.35kb (Figure 3.1). TaqI digestion of the X-linked

**Figure 3.8.**

Southern analysis to determine transgene copy number. Genomic DNA (5μg), obtained by tail biopsy, was digested with TaqI, electrophoresed, transferred, and hybridised to a full-length HPRT cDNA probe, pHPT5. DNA: wt, a wild-type mouse; tg, a mouse transgenic for pDWM1LS+127/+135. The hybridising band and the control band used to determine the transgene copy number by densitometric analysis are indicated.


wt
tg
-control
-1·35kb

endogenous gene generates the same fragment. The 1.35kb fragment is present at one copy per genome in TaqI digested genomic DNA from a wild-type male mouse. In TaqI digested DNA from a hemizygous transgenic male mouse the number of 1.35kb fragments per genome will be equivalent to the transgene copy number plus one, the endogenous gene fragment. The number of copies of the 1.35kb fragment in transgenic DNA can be estimated from the intensity of that band relative to the same band in wild-type DNA on a Southern blot. To correct for loading variations between tracks the intensity of the 1.35kb band is normalised relative to an endogenous gene specific band. The results of this analysis are shown in Figure 3.8. The 1.35kb band and the internal control are indicated. Intensities were measured by densitometric analysis. In wild-type DNA the intensity of the 1.35kb band relative to the control band is 1.8. In the transgenic DNA the normalised intensity of the 1.35kb band is 18.7. Hence there are 10 (18.7 ÷ 1.8 = 10.4) copies of the 1.35kb fragment per genome in transgenic mice. Allowing one copy for the endogenous gene, there are approximately 9 copies of the pDWM1LS+127/+135 transgene.

Transgene expression

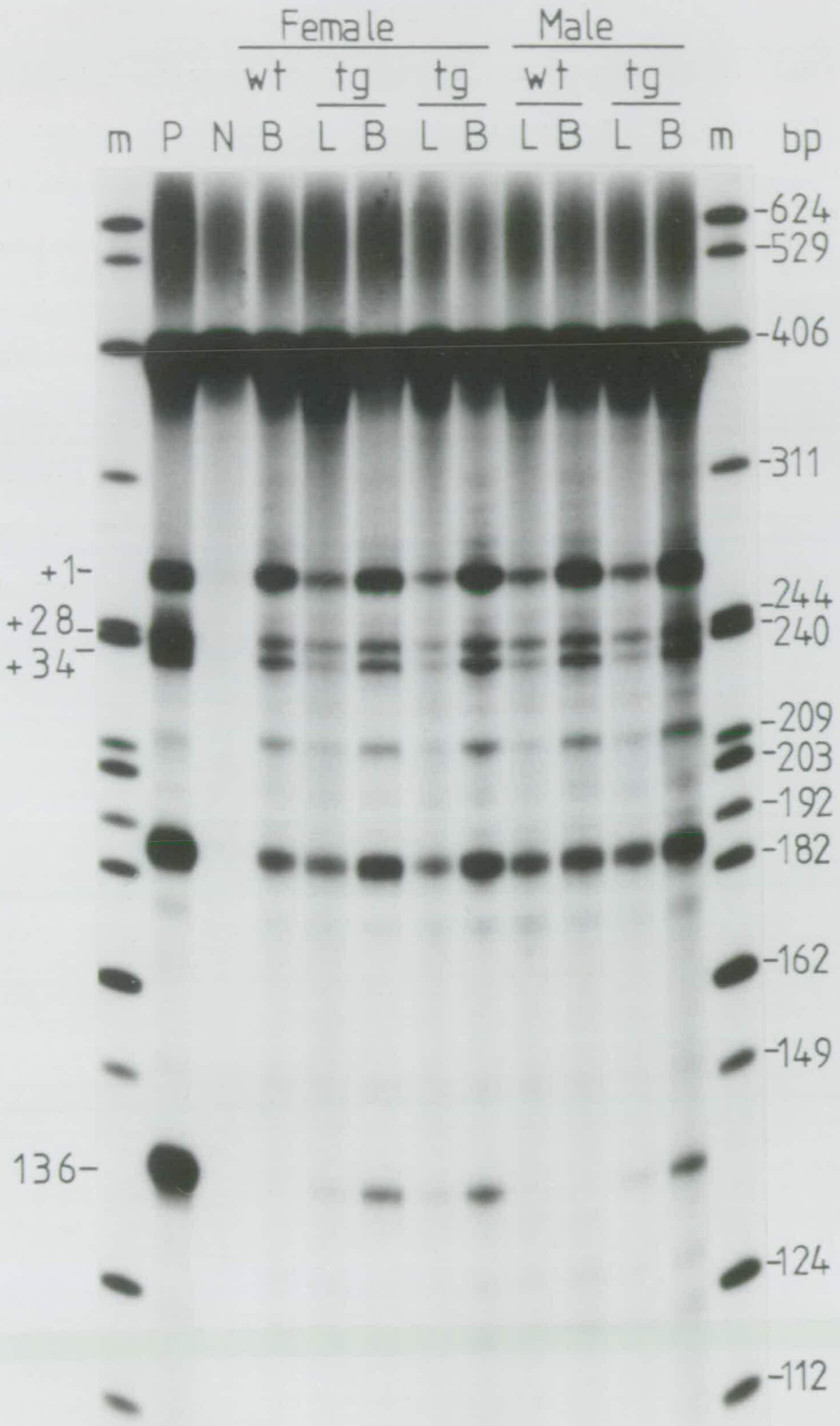
Soluble protein was prepared from tail biopsies of transgenic mice and HPRT enzyme activity and electrophoretic mobility was analysed on non-denaturing polyacrylamide gels. Figure 3.3 shows the results of such an analysis. The first

lane shows the position of HPRT enzyme derived from a wild-type mouse cell line (RJK691). The second lane shows the position of HPRT enzyme derived from the pDWM1LS+127/+135 transgene when expressed in HPRT deficient cultured cells (RJK88). The third lane contains both the wild-type and the transgene derived protein. Two lanes containing protein derived from tail biopsy of transgenic mice (both male and female) are shown adjacent to protein from a wild-type sibling. None of the transgenic mice tested showed any evidence of transgene derived HPRT enzyme activity. Although samples were not matched for protein concentration there is little variation between animals in the endogenous HPRT signal which provides an internal control. By using mixed samples, the sensitivity of the assay was tested. Activity as low as 20% of the wild-type level is readily detectable.

Having failed to detect any transgene derived enzyme activity in pDWM1LS+127/+135 transgenic mice, analysis of HPRT mRNA was carried out. Total RNA was prepared from the tissues of transgenic and wild-type mice and analysed by S1 nuclease protection assays. The results of an S1 nuclease protection analysis of RNA from the livers and brains of wild-type and transgenic mice is shown in Figure 3.9. S1 nuclease reactions were performed on 30 µg of total RNA and half a reaction was loaded in each lane. RNA derived from an HPRT deficient cell line (RJK88) transformed with pDWM1LS+127/+135 is included as a positive control. The major transcription initiation site (+1) is represented by

**Figure 3.9.**

S1 analysis of HPRT mRNA from transgenic mice. Cellular RNA was hybridised to a 424 base single stranded probe, which spanned the main transcription initiation site and was end-labelled at an XhoI site, 271 bases into the gene. Hybridisation mixtures were digested with S1 nuclease. RNA derived from the liver (L) or brain (B) of 5 mice is shown. Both male and female mice are shown that are either transgenic (tg) or wild-type (wt). The positive control (P) lane contains RNA from HPRT deficient Chinese hamster (RJK88) cells transformed with pDWM1LS+127/+135. The negative control (N) lane contains RNA from HPRT deficient Chinese hamster (RJK88) cells. The size of each molecular weight marker (m) is indicated (in bp). The fragments representing the transcription initiation sites (+1, +28, +34) and the 136bp fragment specific to transgenic mRNA, are indicated to the left of the panel.


Female
Male
wt
tg
tg
wt
tg
m P N B L B L B L B L B m bp
-624
-529
-406
-311
+1-
+28-
+34-
-244
-240
-209
-203
-192
-182
-162
-149
136-
-124
-112

the band at 271 bases. The doublet approximately 30 bases shorter represents the two initiation sites (+28 and +34). The band at approximately 136 bases is specific to pDWM1LS+127/+135 mRNA. As explained previously the 424bp band in each track represents undigested probe. The second lane contains RNA from the HPRT deficient cell line RJK88, as a negative control. As expected no HPRT transcripts are evident in this sample. The remaining tracks present a comparison of RNA from liver and brain tissue of three transgenic mice (male and female) with their wild-type (non-transgenic) counterparts. The transgenic mice are hybrids and so to avoid strain specific variations siblings are used as non-transgenic controls.

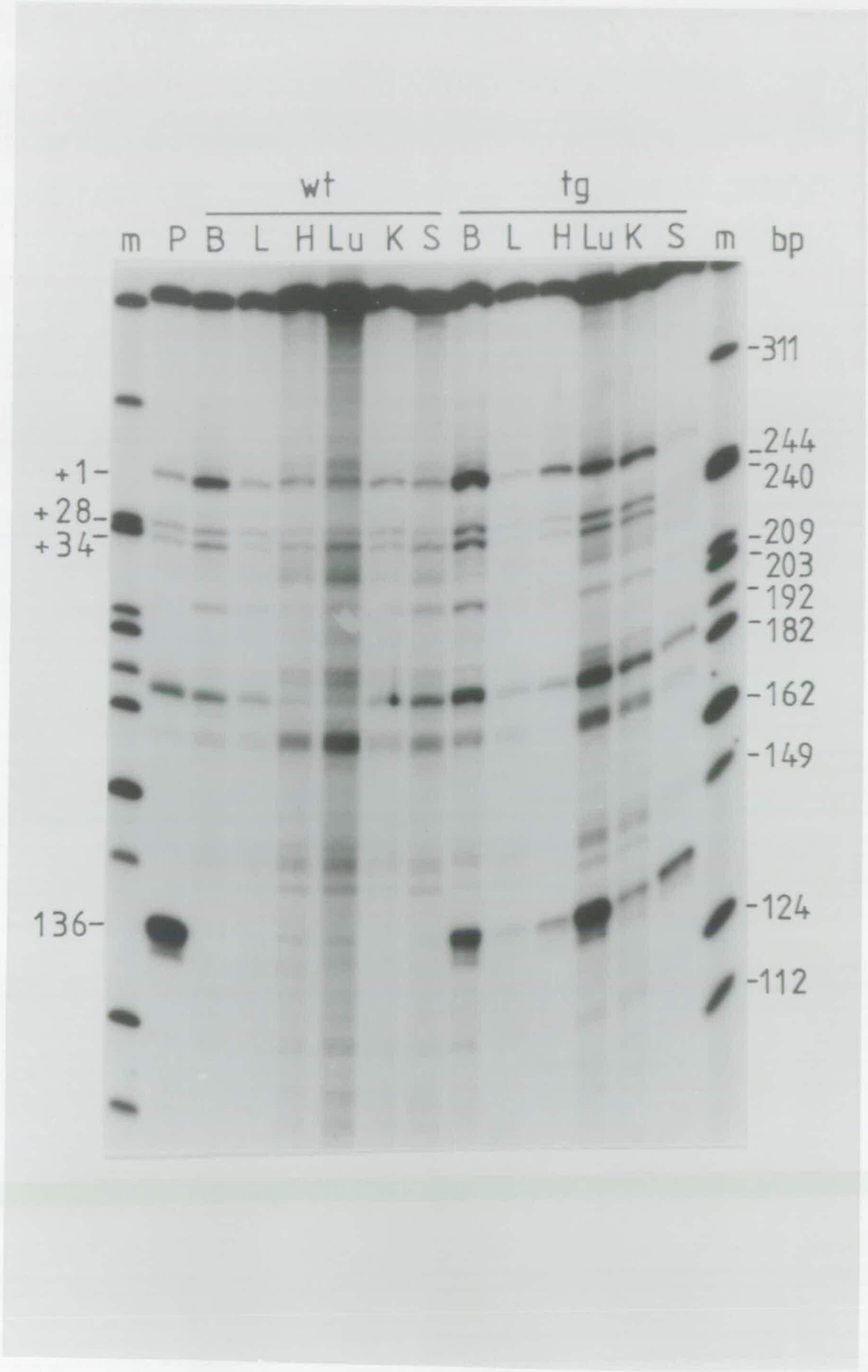
The elevation in HPRT expression in brain relative to liver in normal mice can be seen by comparing the intensity of the +1 band for these two tissues in the wild-type male (Figure 3.9). A wild-type female RNA from brain is also presented. The presence of the 136bp band in each of the transgenic mouse RNA samples demonstrates that the transgene is being expressed. No band is evident in the same position in wild-type tracks. The intensity of the 136bp band is greater in brain RNA than in liver RNA for each of the transgenic mice, indicating that the transgene is expressed at a higher level in brain than liver. The relative elevation in brain is approximately equivalent to that for the endogenous gene.

For

The conclusions that can be drawn from this HPRT transgenic mouse experiment are limited, because only one transgenic line was obtained. Hence the expression of the transgene was studied at just one chromosomal integration site. While the elevated expression of the transgene in brain relative to other tissues may by due to sequences within the transgene, the possibility that the expression pattern observed is due to position effects cannot be excluded. The fact that transcription is initiated from the transgene at sites equivalent to those in the endogenous gene indicates that the transgene contains sequences which are able to direct the transcription initiation from the sites that are used in vivo. The pDWM1LS+127/+135 transgene contains sequences that allow it to be expressed in a range of tissues when inserted into the mouse genome.

**Figure 3.10.**

S1 analysis of HPRT mRNA from a range of wild-type and transgenic tissues. Cellular RNA was hybridised to a 424 base single stranded probe, which spanned the main transcription initiation site and was end-labelled at an XhoI site, 271 bases into the gene. Hybridisation mixtures were digested with S1 nuclease. RNA from a wild-type female (wt) and a transgenic female (tg) is presented. Tissues: B, brain; L, liver; H, heart; Lu, lung; K, kidney; S, spleen. The positive control (P) is RNA from HPRT deficient Chinese hamster (RJK88) cells transformed with pDWM1LS+127/+135. The size (in bp) of each molecular weight marker (m) is indicated to the right of the panel. Fragments representing the transcription initiation sites (+1, +28, +34) and the 136bp fragment specific to transgenic mRNA are indicated to the left of the panel.


wt
tg
m P B L H Lu K S B L H Lu K S m bp
+1
+28
+34
136
311
244
240
209
203
192
182
162
149
124
112

A single HPRT transgenic mouse line was obtained by DNA microinjection into the pronuclei of fertilised eggs. The transgene used pDWM1LS+127/+135 is a derivative of pDWM1 (Melton et al., 1986) which is marked so that its expression is distinguishable from endogenous HPRT expression. It was not possible to detect transgene derived HPRT enzyme activity in mice. However, transgene derived mRNA was found in all tissues analysed of both male and female transgenic mice. The absolute level of HPRT expression from the transgene in this line is lower than from the endogenous gene. However, the tissue-specific variation that is seen in the expression of the endogenous gene, in particular elevated expression in brain, was also seen for the transgene in brain, liver, heart and kidney. Anomalously high pDWM1LS+127/+135 expression was seen in lung and spleen.

could be low - Lewis et Ford -

Having the HPRT transgene present in mice with endogenous HPRT activity has the advantage that it should be possible to directly compare the expression of the transgene with expression of the endogenous X-linked gene, which provides a perfect internal control. However, in practise S1 nuclease protection failed to distinguish absolutely between endogenous gene and transgene expression. It was found that the transgene derived mRNA also contributed to the signal of endogenous mRNA in such analyses. For this reason it was not possible to quantify transgene expression precisely. At this time mice deficient for HPRT expression became available (Hooper et al., 1987). Matings were carried out to produce mice that carried the pDWM1LS+127/+135 transgene but

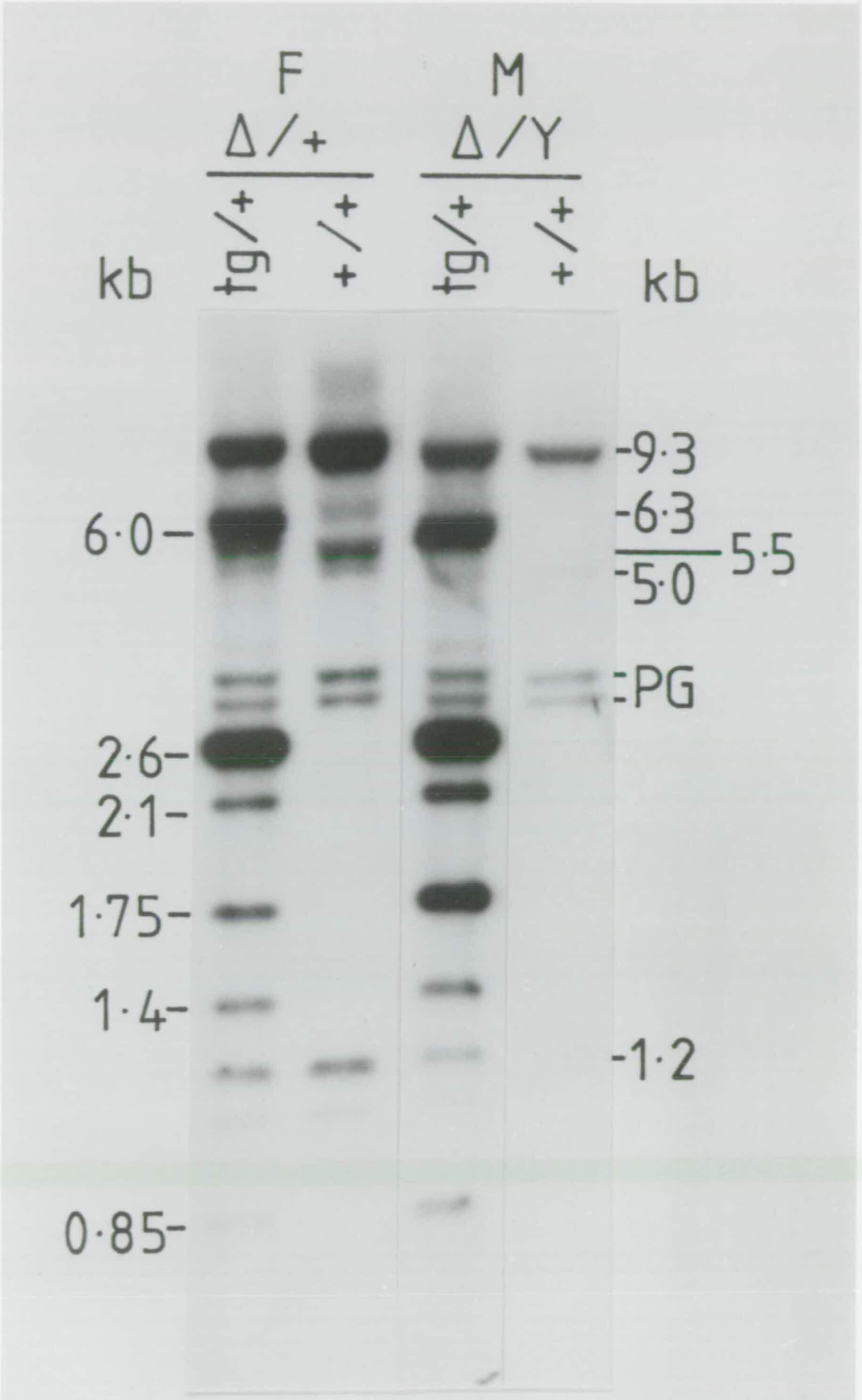
that were also deficient for endogenous gene activity. In these mice transgene expression could be assayed without complications arising from the endogenous gene activity.

# 4. TRANSGENE EXPRESSION ON A DEFICIENT BACKGROUND

To facilitate the analysis of expression from the HPRT transgene pDWM1LS+127/+135, matings were carried out to generate mice which were transgenic but which were deficient for endogenous HPRT activity. HPRT deficient mice (Hooper et al., 1987) were kindly provided by Martin Hooper, Department of Pathology, University of Edinburgh. The characterisation of the molecular genetic basis for the HPRT deficiency is presented in Chapter 5. The mutant HPRT allele can be traced in mice by Southern blot analysis (see Chapter 5). The wild-type locus is denoted $\underline{Hprt}^{b}$ and the mutant locus is denoted $\underline{hprt}^{b-m3}$. Male mice hemizygous for the transgene integration site ($\underline{tg}/+$) were mated to females homozygous for the null HPRT allele ($\underline{hprt}^{b-m3}/\underline{hprt}^{b-m3}$). There are two possible genotypes for each sex in the progeny from this mating. All male progeny are deficient for the endogenous gene ($\underline{hprt}^{b-m3}/Y$) but can be either hemizygous for the transgene or not transgenic. All female progeny are heterozygous deficient at the X-linked locus ($\underline{hprt}^{b-m3}/\underline{Hprt}^{b}$) but can also be either hemizygous for the transgene or not transgenic. These genotypes can be distinguished by Southern blot analysis of tail biopsy genomic DNA (Figure 4.1). Genomic DNA was digested with EcoRI and XhoI and probed with the full-length HPRT cDNA. For each sex it is necessary only to distinguish between those that carry transgene sequences and those that do not. Transgene specific bands are indicated to the left of the panel in Figure 4.1. Thus, determining the genotype of

**Figure 4.1.**

Screening mice for the transgene by Southern analysis of tail DNA. Genomic DNA (5 µg), obtained by tail biopsy, was digested with EcoRI and XhoI, electrophoresed, transferred and hybridised with the full-length HPRT cDNA probe. The sizes (in kb) of hybridising bands from the endogenous HPRT gene are indicated to the right of the panel. PG indicates the pseudogene containing band (Isamat et al., 1988). The sizes of bands specific to transgene sequences are indicated to the left of the panel. Both female (F) mice are heterozygous for the X-linked HPRT null mutation ($\underline{Hprt}^{b}/\underline{hprt}^{b-m3}$) and either hemizygous for the HPRT transgene (tg/+) or nontransgenic (+/+). Male (M) mice are hemizygous for the HPRT null mutation ($\underline{hprt}^{b-m3}$/Y) and either hemizygous for the HPRT transgene (tg/+) or nontransgenic (+/+).

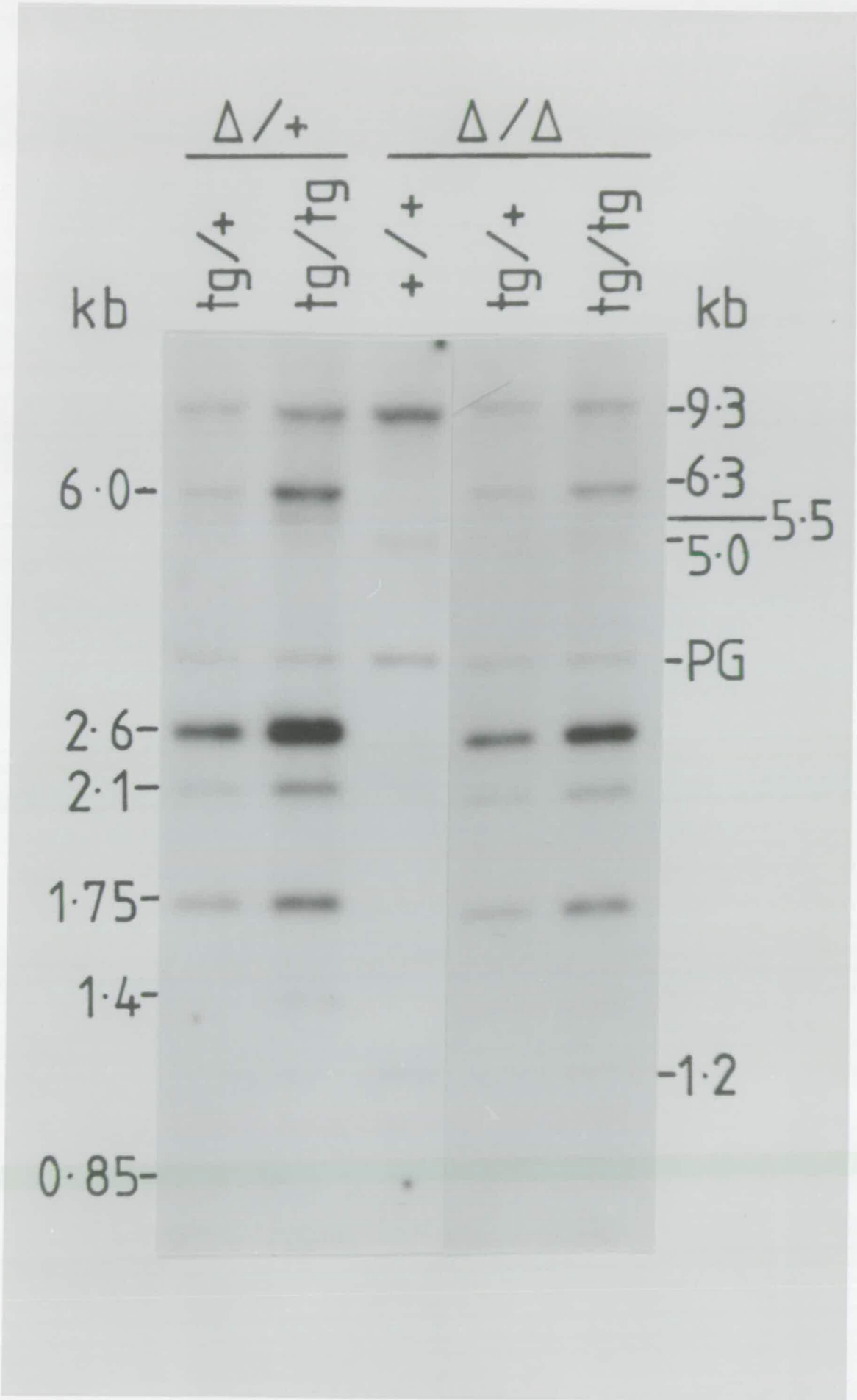

F
Δ/+
M
Δ/Y
tg/+
+/+
tg/+
+/+
kb
kb
-9·3
-6·3
-5·0
5·5
PG
-1·2
6·0
2·6
2·1
1·75
1·4
0·85

progeny is straightforward. Deficiency of the X-linked endogenous HPRT gene in male progeny is indicated by the absence of the 5.5kb and the 6.3kb bands. (This pattern is explained in Chapter 5). In females heterozygosity for this deficiency is masked by bands representing the wild-type allele.

To generate female mice that were transgenic and homozygous for the mutant X-linked HPRT allele a further cross was carried out. Male mice hemizygous for the HPRT transgene and mutant at the X-linked gene ($\underline{hprt}^{b-m3}$/Y) were mated to female mice hemizygous for the transgene and heterozygous for the X-linked mutant allele ($\underline{hprt}^{b-m3}/\underline{Hprt}^{b}$). Progeny were screened by Southern blot analysis of tail biopsy genomic DNA (Figure 4.2). DNA was digested with EcoRI and XhoI and probed with the full-length HPRT cDNA, pHPT5 (Konecki et al., 1982). Female offspring can be either homozygous for the mutant X-linked HPRT allele or heterozygous for this allele. In addition they can be either non-transgenic, hemizygous for the transgene or homozygous for the transgene. Mice that were homozygous for the mutant X-linked gene were distinguished from heterozygotes by the absence of the 5.5kb and 6.3kb bands. Non-transgenic mice obviously have none of the transgene-specific bands. Mice homozygous for the transgene have to be distinguished for hemizygous mice on the basis of the intensity of the transgene specific bands (Figure 4.2). The pseudogene containing band (PG) can be used as an internal control for loading. The 2.6kb transgene specific band in the second lane of Figure 4.2 appears to be

**Figure 4.2.**

Screening transgenic mice by Southern analysis of tail DNA. Genomic DNA (5μg), obtained by tail biopsy, was digested with EcoRI and XhoI, electrophoresed, transferred and hybridised with the full-length HPRT cDNA probe, pHPT5. The sizes (in kb) of hybridising bands from the endogenous HPRT gene are indicated to the right of the panel. The sizes of bands specific to transgene sequences are indicated to the left of the panel. The female mice shown are either heterozygous (Δ /+) or homozygous (Δ / Δ) for the X-linked HPRT null mutation (<u>hprt</u>$^{b-m3}$), and either nontransgenic (+/+), hemizygous for the transgene (tg/+), or homozygous for the transgene (tg/tg). PG indicates the pseudogene containing band.


Δ/+
Δ/Δ
tg/+
tg/tg
+/+
tg/+
tg/tg
kb
kb
-9·3
-6·3
5·5
-5·0
-PG
6·0-
2·6-
2·1-
1·75-
1·4-
-1·2
0·85-

twice as strong as the same band in the first track. Hence the second track contains DNA from a mouse homozygous for the transgene, while the first track contains DNA from a mouse hemizygous for the transgene. Relative intensities were confirmed by densitometric analysis (see Materials and Methods). Similarly, the mice represented in the third and fourth tracks of Figure 4.2 are hemizygous and homozygous for the transgene, respectively.

Having obtained transgenic mice that were deficient for X-linked HPRT activity, tissues for both males and females were analysed for HPRT expression. In these mice the only potentially active HPRT gene was the pDWM1LS+127/+135 transgene. Hence, it was not necessary to distinguish between different HPRT mRNAs, and HPRT transgene expression can be measured by Northern blot analysis. Total RNA (30 g) prepared from tissues, was electrophoresed on formaldehyde agarose gels, transferred and probed with full-length HPRT cDNA, pHPT4 (Konecki et al., 1982). Subsequently, to provide a control for loading, filters were stripped of the HPRT probe and rehybridised with an actin cDNA (Minty et al., 1981). Actin mRNA levels are not equivalent in all tissues and therefore actin does not provide a control for comparisons between different tissues of the same mouse. However, because actin mRNA is at a given level in a particular tissue it does provide a control for comparisons between the same tissue in different animals.

Northern analysis of RNA from a range of tissues from a

female mouse which is hemizygous for the transgene and homozygous deficient at the X-linked gene ($\underline{hprt}^{b-m3}/\underline{hprt}^{b-m3}$) is presented in Figure 4.3. For comparison, RNA from the same tissues of a normal female mouse are included. The actin cDNA reprobe of the same filter is presented below the HPRT cDNA hybridisation. The actin cDNA probe hybridises to both cytoplasmic actin mRNA (2100nts) and muscle actin mRNA (1650nts). The higher molecular weight cytoplasmic actin mRNA is used as the control for variations in loading. In the left hand panel of Figure 4.3 the pattern of wild-type HPRT expression for liver, brain, heart, lung and kidney can be seen. These relative levels of HPRT mRNA were seen in all wild-type animals studied by Northern analysis, and fit the same pattern that was seen in S1 analyses (Chapter 3). This demonstrates that the elevated HPRT enzyme activity seen in mouse brains, correlates with elevated steady-state levels of HPRT mRNA. Band intensities were measured by scanning densitometry of autoradiographs (Table 4.1). By eye it is evident that HPRT mRNA levels are highest in brain and lowest in liver. Also, it is clear that heart and kidney have approximately equivalent levels of HPRT mRNA, with a slightly lower level in lung. Assuming that the loads are equal in each track, a comparison of the absolute value for the intensity of each HPRT signal obtained by densitometry (Table 4.1) supports the above conclusion. The value for HPRT mRNA in brain is the highest ($18.8 \times 10^4$) and is sixfold higher than the value for HPRT mRNA in liver ($3.2 \times 10^4$) which is the lowest of the tissues analysed. Heart and kidney have

**Figure 4.3.**

Northern analysis of tissue RNA from wild-type and transgenic female mice. Total RNA (30µg) was prepared from mouse tissues, electrophoresed on formaldehyde-agarose gels, transferred, and hybridised with the HPRT cDNA probe, pHPT4. The filter was then rehybridised with an actin cDNA probe. Tissues from a wild-type female mouse and a female mouse hemizygous for the transgene and homozygous for the X-linked HPRT null mutation are presented. Tissues: L, liver; B, brain; H, heart; Lu, lung; K, kidney. The actin cDNA probe can hybridise to both cytoplasmic actin mRNA and the lower molecular weight muscle actin mRNA. The size of the HPRT mRNA is approximately 1500 nucleotides.

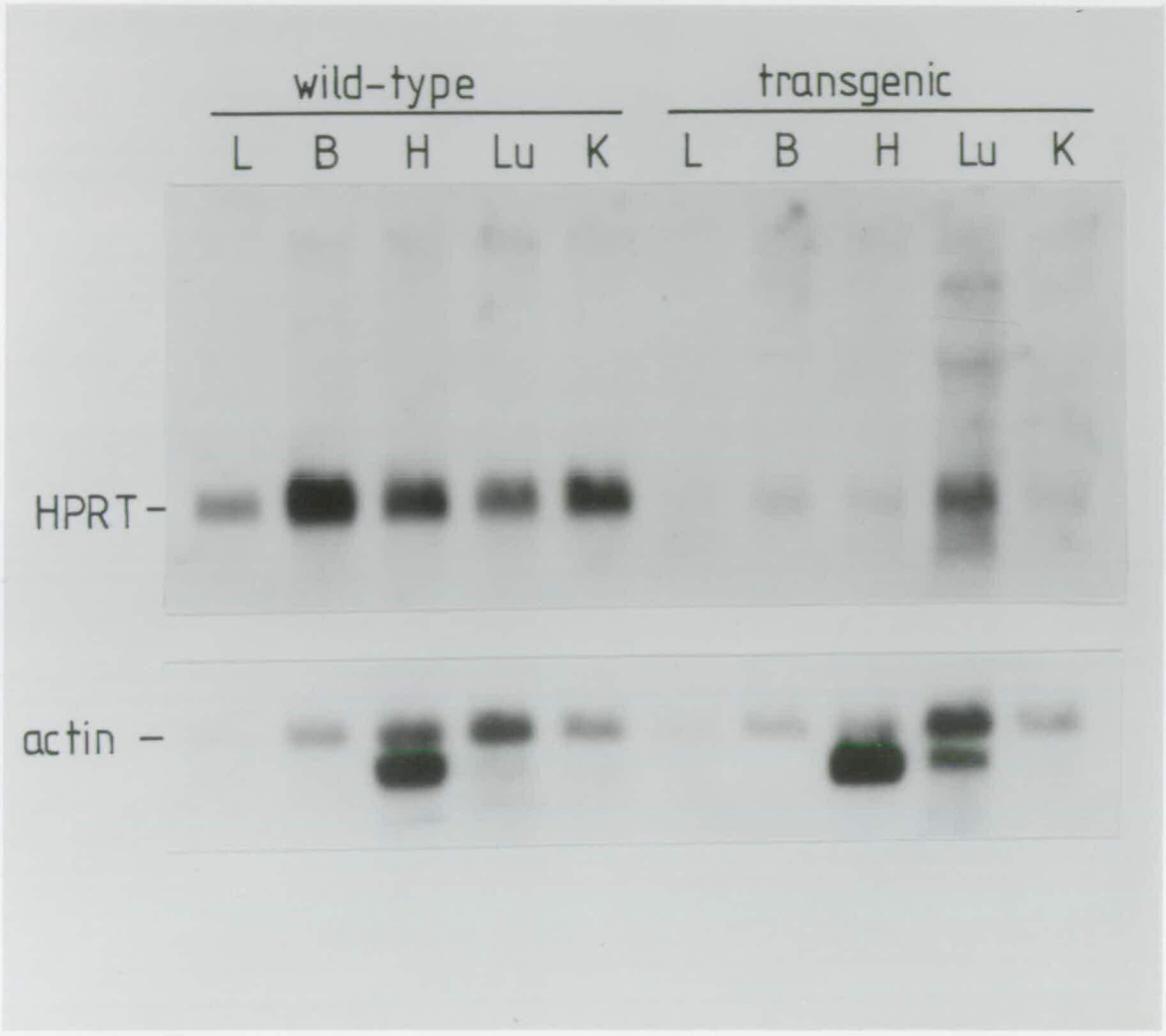

wild-type
transgenic
L B H Lu K L B H Lu K
HPRT-
actin -

**Table 4.1 Densitometric analysis of HPRT expression in wild-type and transgenic female tissues**

| | | H (HPRT x $10^{-4}$) | A (actin x $10^{-4}$) | H/A | % wild-type |
|---|---|---|---|---|---|
| Wild-type: | Liver | 3.2 | 0.5 | 6.4 | - |
| | Brain | 18.8 | 1.9 | 9.7 | - |
| | Heart | 11.0 | 4.3 | 2.5 | - |
| | Lung | 6.0 | 5.9 | 1.0 | - |
| | Kidney | 10.9 | 2.4 | 4.6 | - |
| Transgenic: | Liver | 0.5 | 1.3 | 0.39 | 6 |
| | Brain | 0.5 | 1.4 | 0.38 | 4 |
| | Heart | 0.5 | 3.5 | 0.14 | 3 |
| | Lung | 7.9 | 6.5 | 1.2 | 120 |
| | Kidney | 0.4 | 1.7 | 0.23 | 3 |

H and A values were obtained from the scanning densitometer and represent the area, in arbitrary units, under the peak corresponding to the HPRT and actin mRNA bands, respectively. H/A equals the value H divided by the value A, i.e. it is the standardised HPRT value. % wild-type is the mRNA level in transgenic tissue relative to wild-type expressed as a percentage, and equals (H/A transgenic ÷ H/A wild-type) x 100.

intermediate HPRT mRNA levels of $11.0 \times 10^4$ and $10.9 \times 10^4$ respectively. Lung has a lower level of $6.0 \times 10^4$.

Analysing the same tissues from a female mouse hemizygous for the transgene, it is obvious that the transgene is expressed at a considerably lower level than the X-linked gene in wild-type mice. There is one major exception to this, the level of HPRT expression in the lung of the transgenic female is approximately equivalent to the wild-type level in lung. Variations in loading were compensated for by standardisation relative to the actin signal. This allows comparison of the level of expression in a given tissue in the transgenic mouse with the same tissue in the normal mouse (Table 4.1). A standardised value of HPRT mRNA levels, called H/A, was obtained for each tissue by dividing the densitometry value obtained for the HPRT signal by that obtained for the cytoplasmic actin mRNA signal (see H/A, Table 4.1). For wild-type liver H/A equals 6.4, whereas in transgenic liver H/A equals 0.39. From this it can be calculated that in the liver of the transgenic female, the level of HPRT mRNA is approximately 6% ($100 \times (0.39 \div 6.4)$) of the wild-type level. In brain H/A values for wild-type and transgenic are 9.7 and 0.38, respectively. Thus, in brain pDWM1LS+127/+135 expression is 4% of wild-type. Calculated similarly, transgene expression in heart is 2% of wild-type, and in kidney 3% of wild-type. The level of HPRT mRNA in transgenic lung is 120% of the wild-type level. With the exception of lung, when examined by eye the transgenic tissues appear to have the same pattern of relative levels of

HPRT mRNA as wild-type tissues. That is, liver has the lowest level, brain has the highest level, and heart and kidney have an intermediate level. The absolute levels of HPRT mRNA in transgenic tissues are too low for densitometric analysis to verify this pattern.

RNA derived from a hemizygous transgenic male was analysed similarly. The pattern and absolute level of expression of the pDWM1LS+127/+135 transgene in male tissues was the same as in the female tissues described above, including the anomalously high level of expression in lung (data not presented).

The analyses so far presented of transgene expression have been limited to mice which were hemizygous for the transgene. Cells in normal mice only ever contain one active HPRT allele. Because the gene is X-linked, males contain a single copy and females are functionally hemizygous due to random inactivation of the genes on one X-chromosome in all cells. The pDWM1LS+127/+135 transgene in this line is integrated on an autosome, and should not, therefore, be subject to X-inactivation. Both males and females which are homozygous for the transgene should contain two active alleles. To test this prediction the level of HPRT expression in mice hemizygous for the transgene was compared to the level of expression in mice homozygous for the transgene by Northern blot analysis (Figure 4.4). The mice analysed in Figure 4.4 are all female and were identified by Southern analysis of tail-biopsy DNA as described previously. Densitometric

**Figure 4.4.**

Northern analysis of HPRT expression in hemizygous and homozygous transgenic mice. Total RNA (30μg) was prepared from mouse tissues, electrophoresed on formaldehyde-agarose gels, transferred, and hybridised with the HPRT cDNA probe, and subsequently with the actin cDNA probe. Liver (L) and brain (B) RNA are shown for four mice. Mice: wt indicates a wild-type female mouse; Δ/Δ indicates homozygosity for the X-linked HPRT null mutation (<u>hprt</u>$^{b-m3}$); tg/+ indicates hemizygosity for the transgene; tg/tg indicates homozygosity for the transgene.


Δ/Δ
wt
Δ/Δ
tg/+
tg/tg
L B L B L B L B
HPRT-
actin-

**Table 4.2 Densitometric analysis of HPRT expression in hemizygous and homozygous transgenic females**

| | | H (HPRT x $10^{-4}$) | A (actin x $10^{-4}$) | H/A | % wild-type |
|---|---|---|---|---|---|
| Wild-type: | Liver | 1.1 | 1.8 | 0.6 | - |
| | Brain | 5.6 | 3.3 | 1.7 | - |
| HPRT Deficient: | Liver | - | 2.0 | - | - |
| | Brain | - | 9.8 | - | - |
| Hemizygous Transgenic: | Liver | - | 2.2 | - | - |
| | Brain | 1.9 | 10.0 | 0.2 | 12 |
| Homozygous Transgenic: | Liver | 0.5 | 2.2 | 0.23 | 33 |
| | Brain | 3.3 | 7.8 | 0.4 | 24 |

H and A values were obtained from the scanning densitometer and represent the area, in arbitrary units, under the peak corresponding to HPRT and actin mRNA bands, respectively. H/A equals the value H divided by the value A, i.e. it is the standardised HPRT value. % wild-type is the mRNA level in transgenic tissue relative to the wild-type level expressed as a percentage, and equals (H/A transgenic ÷ H/A wild-type) x 100.

analysis of the autoradiographs shown in Figure 4.4. is summarised in Table 4.2. In the wild-type female the approximately 6-fold elevation in HPRT expression in brain relative to liver is evident. No HPRT mRNA is detectable in either the liver or brain of the HPRT deficient ($\underline{hprt}^{b-m3}/\underline{hprt}^{b-m3}$) non-transgenic female. HPRT mRNA was only detectable in liver from the female hemizygous for the transgene with long autoradiography exposure times, and is not visible in Figure 4.4. HPRT mRNA is detectable in the brain of this animal in Figure 4.4, at a level of approximately 12% of wild-type. In the female mouse homozygous for the transgene the level of HPRT mRNA in liver and brain is 34% and 24% of wild-type, respectively. Hence, the level of HPRT mRNA in mice homozygous for the transgene is approximately double the level in mice hemizygous for the transgene. As predicted, there is no evidence of inactivation of one allele of the transgene in homozygous mice.

In conclusion, the autosomal HPRT transgene is expressed in most tissues analysed at approximately 10% of the level of the X-linked HPRT gene. Anomalously high levels of transgene-derived HPRT mRNA are detectable in lung and spleen, but otherwise the relationship between the levels of transgene-derived mRNA in different tissues has the same pattern as for endogenous HPRT mRNA. Most interestingly, pDWM1LS+127/+135 expression is elevated in the brain of transgenic mice. A detailed discussion of these results is presented in Chapter 7.

1985) the mutant allele present in E14TG2a was designated as *hprt*$^{b-m3}$.

In collaboration with Oliver Smithies, University of North Carolina, Chapel Hill, USA, we were able to demonstrate that homologous recombination can be used to target specific modifications to the genome of mammalian cells (Doetschman et al., 1987). HPRT expression in E14TG2a was restored by correction of the *hprt*$^{b-m3}$ allele by gene targeting. The targeting vector used contained mouse HPRT exons 1 and 2 along with the human HPRT promoter. When the experiments described below were initiated a few reports had been published demonstrating the feasibility of gene targeting in ES cells (Thomas and Capecchi, 1987; Doetschman et al., 1987). However, at that time no one had succeeded in introducing a targeted modification into the mouse germ line using ES cells.

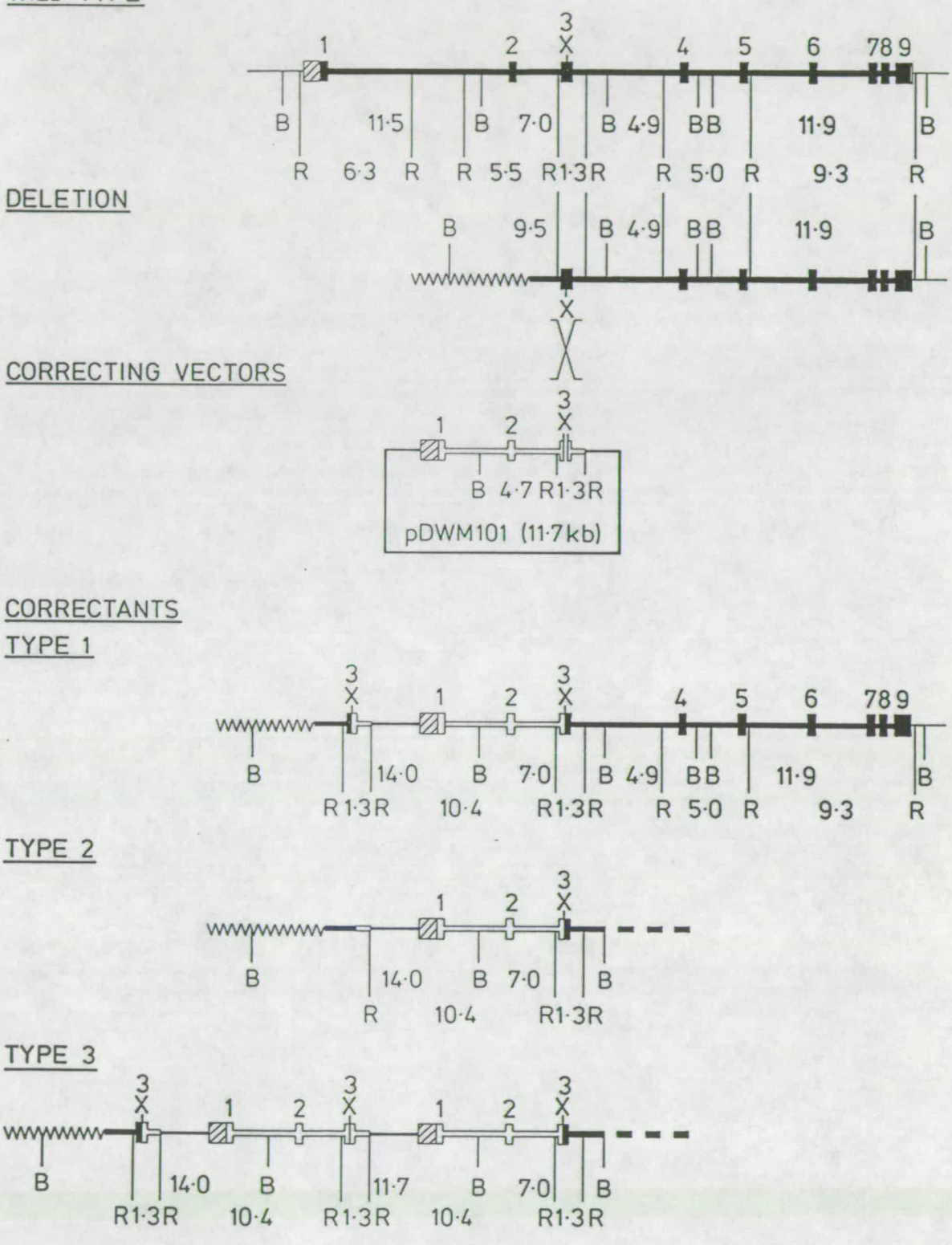
## Targeted correction of the E14TG2a deletion

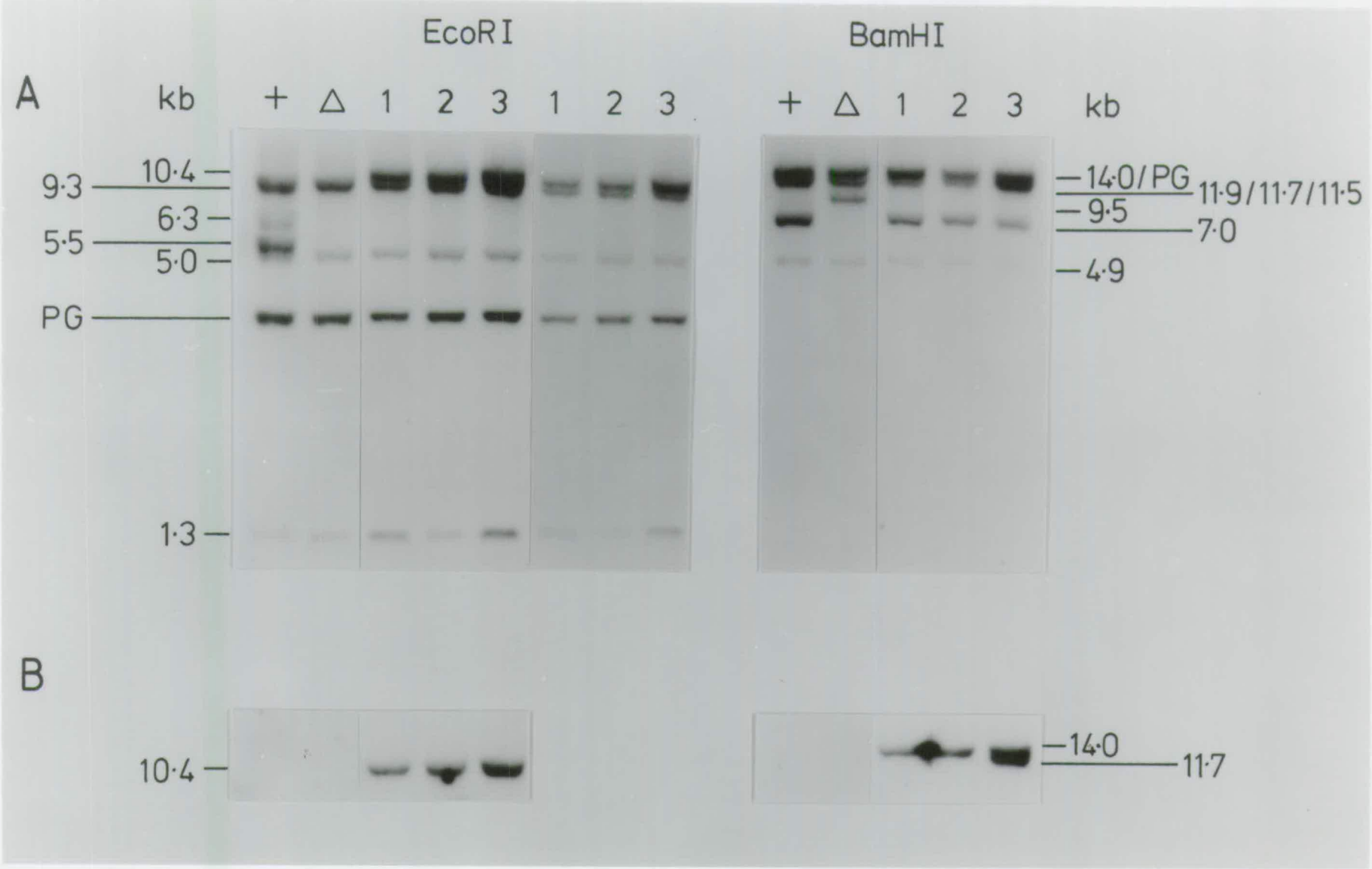
The vector pDWM101, constructed by Carolanne McEwan, was designed to correct the mutant HPRT allele in E14TG2a ES cells by gene targeting. Details of pDWM101 construction are presented in the Materials and Methods section. The vector and the gene targeting experiment are shown schematically in Figure 5.1. The structure of the mutant *hprt*$^{b-m3}$ allele is shown below the wild-type allele. As described above the E14TG2a mutation is a deletion of the first two HPRT exons

and at least 12kb of 5' flanking sequence including the HPRT promoter region. pDWM101 contains mouse exons 1, 2 and 3 plus the mouse HPRT promoter. Exons 1 and 2 are necessary to replace the deleted coding regions in E14TG2a. The exon 3 region provides homology with the target locus. The extent of homology is between 2.3kb and 4.2kb, depending upon the precise downstream end point of the deletion. In pDWM101 the HPRT promoter is provided in 635bp of 5' flanking sequence. This sequence is sufficient for HPRT expression in cultured cells (Melton et al., 1986).

The predicted outcome of correction, homologous recombination between pDWM101 and the mutant target locus, is shown in Figure 5.1. The corrected gene has the same structure, called type 1, as the wild-type gene except that the first intron is reduced from 10.8kb to 4.1kb. The corrected allele has a restriction map which is distinct from both the wild-type gene and the deletion mutant allele. Hence, correctants can be identified by Southern blot analysis. In the EcoRI restriction map a novel 10.4kb fragment is generated, containing exons 1 and 2, and the 1.3kb band, containing exon 3 is duplicated. In the BamHI restriction map a 14.0kb fragment containing exon 1 and the duplicated exon 3, is generated, and the 9.5kb fragment that contains just exon 3 in the deleted allele is replaced by the wild-type 7.0kb fragment which contains exons 2 and 3. The 7.0kb and the 14.0kb BamHI fragments in the corrected allele, are the products of one BamHI site in the vector sequences and one BamHI site in the chromosomal sequences. While the

# Figure 5.1


WILD-TYPE
1
2
3
X
4
5
6
7
8
9
B
11·5
B
7·0
B
4·9
B
B
11·9
B
R
6·3
R
R
5·5
R
1·3
R
R
5·0
R
9·3
R
DELETION
B
9·5
B
4·9
B
B
11·9
B
X
CORRECTING VECTORS
1
2
3
X
B
4·7
R
1·3
R
pDWM101 (11·7kb)
CORRECTANTS
TYPE 1
3
X
1
2
3
X
4
5
6
7
8
9
B
14·0
B
7·0
B
4·9
B
B
11·9
B
R
1·3
R
10·4
R
1·3
R
R
5·0
R
9·3
R
TYPE 2
1
2
3
X
B
14·0
B
7·0
B
R
10·4
R
1·3
R
TYPE 3
3
X
1
2
3
X
1
2
3
X
B
14·0
B
11·7
B
7·0
B
R
1·3
R
10·4
R
1·3
R
10·4
R
1·3
R


EcoRI
BamHI
A
kb
+
Δ
1
2
3
1
2
3
+
Δ
1
2
3
kb
9·3
10·4
6·3
5·5
5·0
PG
1·3
14·0/PG
11·9/11·7/11·5
9·5
7·0
4·9
B
10·4
14·0
11·7

7.0kb fragment is also present in the wild-type gene, the 14.0kb fragment is diagnostic for targeted correctants. The correcting vector, pDWM101, can only restore HPRT activity to cells in which it undergoes homologous recombination with the mutant allele. pDWM101 has no independent HPRT activity and therefore will not restore HPRT activity by integration into other genomic sites. Consequently, targeted correctants can be isolated by subjecting transformed cells to selection for HPRT activity, that is, in HAT-containing medium.

Three gene targeting experiments using pDWM101 were performed. Electroporations and cell culturing in the first two were carried out by Angela Pow. Analysis of correctant gene structures was carried out by myself. The third experiment was performed by myself.

pDWM101 DNA was linearised at the unique XhoI site in exon 3 before being introduced into E14TG2a cells by electroporation. Electroporation conditions were chosen that resulted in minimal cell death in control plating experiments. As a positive control for transformation E14TG2a cells were electroporated with pDWM100 DNA. pDWM100 was constructed by replacing the 1.25kb XhoI-EcoRI fragment in pDWM101 (Figure 5.1) with the 1.8kb XhoI-EcoRI fragment from pDWM1 (Melton et al., 1986). Thus, pDWM100 contains the entire HPRT coding region and has HPRT activity. It can therefore restore HPRT activity to E14TG2a cells without a requirement for homologous recombination.

In the first experiment 4.6 x $10^7$ E14TG2a cells were electroporated. From these eight $HAT^R$ colonies were obtained (Table 5.1). The frequency of HPRT correctants per cell electroporated is 1.75 x $10^{-7}$. The frequency of $HAT^R$ colonies per cell electroporated with pDWM100 DNA was 2.5 x $10^{-5}$ in this experiment. The ratio of $HAT^R$ colonies obtained from cells electroporated with pDWM101 DNA to $HAT^R$ colonies obtained from cells electroporated with pDWM100 DNA was 0.007. We have called this value the targeting index (Table 5.1). It gives an indication of the number of targeting events relative to the number of random integrations, although it does not take into account integrations of pDWM100 into sites which prevent expression.

Southern hybridisations were carried out to analyse the structure of the HPRT gene in each of the eight $HAT^R$ clones and to confirm that gene targeting had occurred (Figure 5.2). Seven of the eight $HAT^R$ clones had the HPRT gene structure predicted for correctants, which we have called type 1 (Figures 5.1 and 5.2). In the EcoRI digest the new 10.4kb band is visible and the intensity of the 1.3kb band is increased twofold with respect to the equivalent band in the E14TG2a lane. The intensity of the pseudogene band in the EcoRI digests allows comparison of the relative loading in each track, thus providing an internal control. In the BamHI digest the disappearance of the 9.5kb fragment which is characteristic of the deletion, and the regeneration of the 7.0kb wild-type band can be seen. The generation of the 14.0kb fragment, which comigrates with the pseudogene

**Table 5.1 Summary of correction experiments**

| Expt | Number of $HAT^R$ Colonies | Frequency of HPRT Correctants[a] | Positive Control[b] | Targeting Index[c] | Correctant Types 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| 1 | 8 | $1.75 \times 10^{-7}$ | $2.5 \times 1^{-5}$ | 0.007 | 7 | - | 1 |
| 2 | 2 | $4.0 \times 10^{-8}$ | $6.4 \times 10^{-6}$ | 0.006 | 1 | 1 | - |
| 3 | 12 | $2.2 \times 10^{-7}$ | $6.3 \times 10^{-6}$ | 0.035 | 12 | - | - |

[a] Number of $HAT^R$ colonies generated per cell electroporated with correcting vector DNA.

[b] Number of $HAT^R$ colonies generated per cell electroporated with pDWM100 DNA.

[c] Frequency of $HAT^R$ colonies generated per cell electroporated with correcting vector DNA relative to $HAT^R$ colonies generated per cell electroporated with PDWM100 DNA (a/b).

fragment, is indicated by the increased intensity of the band at that molecular weight.

One of the eight $HAT^R$ clones generated in the first electroporation experiment had an HPRT gene with a structure different to the predicted structure. We have called this structure type 3. In this clone two correcting vector molecules have integrated in tandem. The restriction map for this clone can be seen in Figure 5.1, as deduced from the Southern analysis shown in Figure 5.2. In an EcoRI digest the 10.4kb fragment is duplicated and there are 3 copies of the 1.3kb fragment, as opposed to two copies in the type 1 structure and one copy in the deletion mutant. This can be seen in the relative intensity of the 1.3kb band in each lane. In a BamHI digest, in addition to the 14.0kb and 7.0kb fragments seen in type 1 correctants, there is an 11.7kb, vector-sized, fragment. This interpretation was supported when the same hybridisation membrane was reprobed with pUC8 DNA (Figure 5.2B). The pUC8 sequence is present within the 10.4kb EcoRI fragment and consequently the intensity of this band is increased twofold in type 3 DNA relative to type 1. In a BamHI digest of the type 3 DNA the pUC8 sequence is present in both the 14.0kb fragment and the 11.7kb fragment. The equal intensities of the 14.0kb and 11.7kb bands in type 3 show that there are two copies of the vector integrated. If there were more than two copies the intensity of the 11.7kb band would be at least twofold greater than the 14.0kb band. Thus seven of the eight $HAT^R$ clones generated in this experiment were the product of the desired gene targeting

event. One of the eight, although also the product of targeted correction had two copies of the targeting vector integrated in tandem. This structure could have arisen either by concatamerisation of two vector molecules before homologous integration, or by two consecutive homologous recombination events.

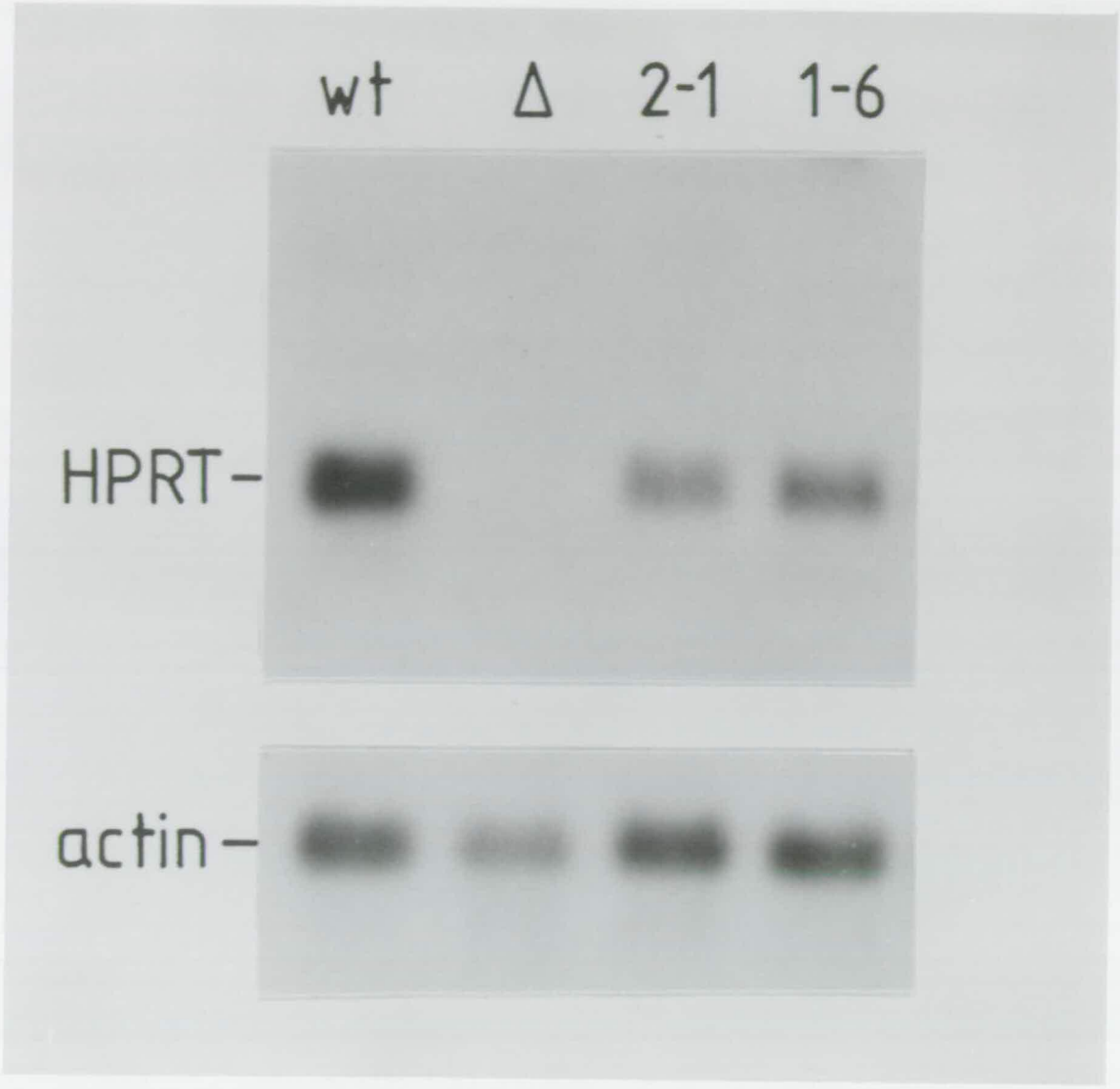
With the intention of producing a targeted correctant clone suitable for injection into blastocysts the above experiment was repeated using lower passage E14TG2a cells. $5 \times 10^7$ E14TG2a cells were electroporated with pDWM101 DNA. Two $HAT^R$ colonies were generated which were analysed by Southern hybridisations. The frequency of $HAT^R$ clones per cell electroporated with pDWM101 DNA was $4.0 \times 10^{-8}$. In this experiment, the frequency of $HAT^R$ clones generated by electroporation with the positive control vector, pDWM100, DNA was $6.4 \times 10^{-6}$. Consequently, the targeting index was 0.006 (Table 5.1), which is the same as for the first experiment (0.007). One $HAT^R$ clone generated in this experiment had the type 1 HPRT gene structure, which is the predicted outcome of homologous recombination. The second of the $HAT^R$ clones had a different HPRT gene structure which we have called type 2 (Figures 5.1 and 5.2). The exon 3 containing 1.3kb EcoRI band has the same intensity in this clone as in E14TG2a. Hence, in this clone, called correctant 2-1, the exon 3 region is not duplicated as it is in type 1 correctants. In addition, the intensity of the 14.0kb BamHI band is lower than in type 1 correctants, although the size of the band is not visibly reduced at this resolution. An

XhoI digest of clone 2-1 showed that the XhoI site in the upstream copy of exon 3 is absent (data not presented). The EcoRI site immediately upstream of the plasmid sequence is present as indicated by the presence of the 10.4kb EcoRI band. The results of Southern analysis of correctant 2-1 are consistent with the targeting vector having integrated correctly, but with an additional event resulting in a small deletion removing the upstream copy of exon 3. The absence of the upstream copy of exon 3 does not affect the HPRT positive status of clone 2-1 because the deletion occurs outside of the functional gene.

To investigate the frequency with which anomalous structures arise in corrected genes, a third electroporation experiment was performed. Twelve $HAT^R$ clones were obtained from $5.5 \times 10^7$ E14TG2a cells electroporated with pDWM101 DNA. The frequency of $HAT^R$ clones per cell electroporated with pDWM101 DNA was $2.2 \times 10^{-7}$. The frequency of $HAT^R$ clones per cell electroporated with pDWM100 DNA was $6.3 \times 10^{-6}$. Therefore, the targeting index in this experiment was 0.035. This is fivefold higher than obtained in the first two experiments, which is within the usual variation seen in cultured mammalian cell transformations. All twelve $HAT^R$ clones generated in this experiment had the type 1 HPRT gene structure as determined by Southern analysis. None of the HPRT correctants produced in the three electroporation experiments had random integrations of the targeting vector in addition to the targeted integration.

**Figure 5.3.**

Northern analysis showing expression from corrected genes in ES cells. Total RNA (30μg) was prepared from ES cells, electrophoresed on formaldehyde-agarose gels, transferred, and hybridised with the HPRT cDNA probe, pHPT4, and subsequently with the actin cDNA probe. RNA: wt, E14 (wild-type); Δ, E14TG2a (HPRT deficient); 2-1, correctant clone 2-1; 1-6, a correctant clone with the type 1 structure. The size of the HPRT mRNA is approximately 1500 nucleotides.


wt
Δ
2-1
1-6
HPRT-
actin-

level in E14 cells. Possible explanations for this are discussed later.

## Screening mice for the HPRT deletion allele

The experiments described in this thesis depended to a large extent on the ability to identify different HPRT alleles by Southern analysis, of mouse tail-biopsy DNA. As described below, this ability facilitated the generation of an HPRT deficient mouse line. To produce an HPRT deficient mouse line it was necessary to distinguish between heterozygous HPRT deficient ($\underline{Hprt}^{b}/\underline{hprt}^{b-m3}$) females and females homozygous for the wild-type gene ($\underline{Hprt}^{b}/\underline{Hprt}^{b}$). An example of how Southern analysis on tail biopsy DNA allowed female mice to be screened is illustrated in Figure 5.4. Genomic DNA was digested with BamHI and hybridised with a full-length HPRT cDNA probe. The restriction fragment pattern has been explained in the context of characterising the E14TG2a mutation earlier in this chapter. The characteristic of a BamHI digest which makes it suitable is the generation of a band at 9.5kb which is specific to the HPRT deletion allele. Whereas the 7.0kb BamHI band is specific to the wild-type allele, or the corrected allele. The first three lanes in Figure 5.3 contain control DNAs for the screening of seven female mice, lanes four to ten. The first lane contains DNA from a heterozygous deficient female as indicated by the presence of both the 7.0kb band which is specific to the wild-type allele, and the 9.5kb band which is specific to the

**Figure 5.4.**

Tracing the HPRT deletion allele (*hprt*$^{b-m3}$) in mice by Southern analysis. Genomic DNA (5 μg), obtained by tail biopsy, was digested with BamHI, electrophoresed, transferred and hybridised with the full-length HPRT cDNA probe. The sizes (in kb) of hybridising bands are indicated. PG indicates pseudogene containing bands. A female (F) mouse which is heterozygous for the *hprt*$^{b-m3}$ allele, and two males (M), one of which is wild-type (+/Y), and the other of which is hemizygous for the deletion allele (Δ/Y), are used as controls to screen seven female mice, which are either wild-type (+/+) or hemizygous for the *hprt*$^{b-m3}$ allele (Δ/+).

mutant allele. The second lane contains DNA from a wild-type male. The third lane contains DNA from an HPRT deficient male as indicated by the presence of the 9.5kb band and the absence of the wild-type specific 7.0kb band. Thus it is possible to determine the genotype of each of the seven females as indicated above each track.

## 6. GERM LINE TRANSMISSION AND EXPRESSION OF THE CORRECTED HPRT GENE

The two correctant clones, 2-1 and the type 1 clone, 1-8, produced from low passage E14TG2a cells, were assayed for the frequency of HPRT deficient cells in their culture populations. To do this cells were grown in the absence of selection for two weeks and then their plating efficiency in medium containing 6-thioguanine (6-TG), relative to their plating efficiency in non-selective medium, was measured. The relative plating efficiency for 2-1 was 0.03%. This is approximately the same as the value for wild-type E14 cells. The plating efficiency for 1-8 was 21.0%. Hence, after 2 weeks of non-selective growth, 21.0% of the cell population in the type 1-8 correctant culture was HPRT deficient. Southern analysis of DNA from $6\text{-TG}^R$ populations derived from the correctant 1-8 culture after twenty four days of non-selective growth revealed an HPRT gene structure identical to that of E14TG2a, the starting deletion mutant (data not shown). There were, at the time, two possibilities for the origin of these cells. They could arise due to the persistence in the HAT culture of E14TG2a cells, surviving by metabolic cooperation (Hooper, 1982). Alternatively, $6\text{-TG}^R$ cells might be generated by the loss of pDWM101 vector sequences from the genome of corrected cells, perhaps due to a reversal of the integrating homologous recombination (correction) event. It was decided that clone 1-8 was unsuitable for injection into blastocysts. Subsequently, by repeating the relative plating efficiency assay on a clonal

isolate of 1-8, it was shown that the presence of 6-TG$^{R}$ cells in the population of correctant 1-8 cells was due to surviving E14TG2a cells and not due to instability of the corrected gene.

## The production of chimeric mice

The injection of ES cells into blastocysts and the test breeding of chimeras was carried out by Alan Clarke, Department of Pathology, University of Edinburgh.

Cells from correctant clone 2-1 were injected into C57BL/6/Ola x CBA/Ca/Ola F2 blastocysts. These blastocysts are homozygous for the wild-type alleles at the c and p loci (C/C, P/P) and segregating for the alleles A (agouti) and a (nonagouti). The 2-1 correctant cells are derived from strain 129/Ola and are therefore homozygous for the alleles $c^{ch}$ (chincilla) and p (pink-eyed dilution), which both lighten coat colour, and for $A^{w}$ (white-bellied agouti). Hence, chimeric mice are identifiable by the presence of light colouring in an otherwise dark coat. From 93 blastocysts injected, twenty six mice were born. Fifteen chimeras were obtained, twelve of which were male. These male chimeras were tested for germ line transmission of the ES cell-derived $c^{ch}$ and p markers by crossing with strain 129/Ola females. Offspring, which like strain 129 were light yellow with pink eyes were diagnostic of germ line transmission. One out of eight male chimeras tested showed

germ line transmission to 21 of 28 offspring. The frequency of chimeras per animal born in this experiment was comparable to that obtained with the parental HPRT deficient (E14TG2a) ES cell line (Hooper et al., 1987). However, fewer injected blastocysts survived to term. Nineteen out of 34 E14TG2a-derived male chimeras showed germ line transmission. Correctant 2-1 cells have been shown to have a lower percentage of diploid cells than the parental line E14TG2a and this may account for the difference in the frequency of germ line chimeras generated. In total 75% of offspring inherited ES cell derived markers in our experiment. E14TG2a-derived chimeras have been obtained which transmitted ES cell markers to as many as 100% or to as few as 1% of offspring (Hooper et al., 1987).

The corrected HPRT allele in correctant 2-1 has been designated as $\underline{Hprt}^{b-m4}$. Female offspring of the germ line chimera are heterozygous for the corrected HPRT allele and the wild-type allele. These females were back-crossed to their chimeric father to generate a pure breeding line for expression studies. There were two possible genotypes for both male and female progeny from this cross. Male offspring could be either hemizygous for the wild-type allele or hemizygous for the corrected allele ($\underline{Hprt}^{b-m4}$). Female offspring could be either homozygous for the corrected allele ($\underline{Hprt}^{b-m4}/\underline{Hprt}^{b-m4}$) or heterozygous for the corrected allele ($\underline{Hprt}^{b-m4}/\underline{Hprt}^{b}$). Southern analysis of tail biopsy genomic DNA was used to distinguish between the two possible genotypes for each sex. This is illustrated in

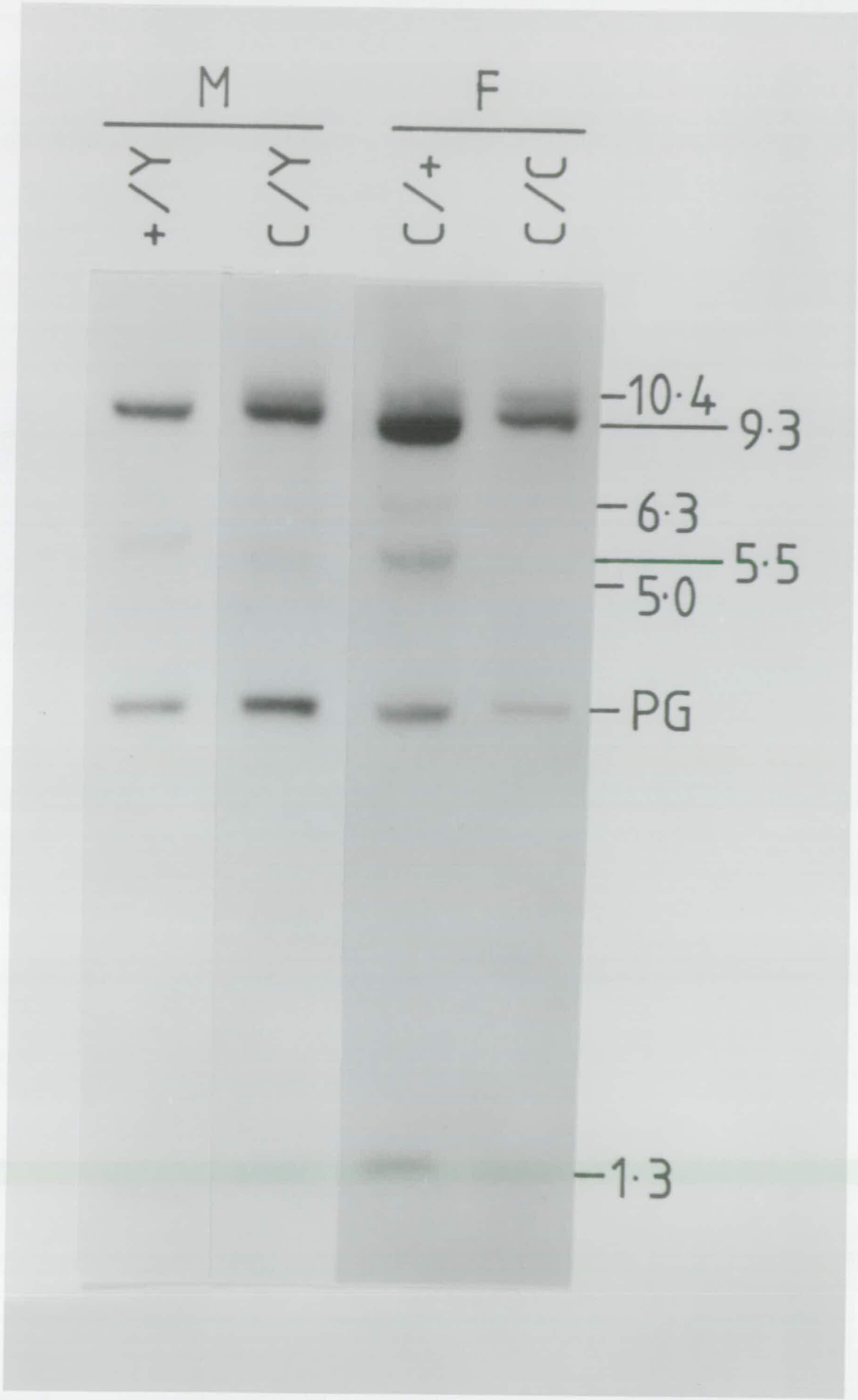

M
F
+/Y
C/Y
C/+
C/C
10·4
9·3
6·3
5·5
5·0
PG
1·3

Figure 6.1. Genomic DNA was digested with EcoRI and hybridised with a full-length HPRT cDNA probe. For males it was necessary to distinguish between the wild-type gene and the corrected gene. In an EcoRI digest the pattern for the corrected gene differs from the wild-type gene in that the 6.3kb and 5.5kb bands are absent and there is an additional band at 10.4kb. This pattern is explained in Chapter 5. Females that are homozygous for the corrected gene have the same Southern pattern as males hemizygous for the corrected gene, in an EcoRI digest. That is, the wild-type bands of 6.3kb and 5.5kb are absent and there is an additional band at 10.4kb. Females that are heterozygous for the corrected gene are indicated by the presence of both the 5.5kb and 6.3kb wild-type fragments and the corrected gene specific band at 10.4kb.

Expression of the corrected HPRT gene in mice

Total RNA was prepared from tissues of females homozygous for the corrected HPRT gene ($\underline{Hprt}^{b-m4}/\underline{Hprt}^{b-m4}$) and males hemizygous for the corrected HPRT gene ($\underline{Hprt}^{b-m4}/Y$). HPRT expression was assayed in each tissue by Northern blot analysis. Figure 6.2 shows a Northern hybridisation of tissue RNAs from a wild-type female and a female homozygous for the corrected gene. The wild-type mice used as controls for expression of the corrected gene are strain 129 and therefore isogenic with the corrected mice, which are also strain 129. The transfer was probed with the full-length

**Figure 6.2.**

Northern analysis of HPRT expression in female mice homozygous for the corrected gene. Total RNA (30 μg) was prepared from mouse tissues, electrophoresed on formaldehyde-agarose gels, transferred, and hybridised with the HPRT cDNA probe and subsequently with the actin cDNA probe. Tissues are analysed from a wild-type female and a female homozygous for the corrected HPRT allele. Tissues: L, liver; B, brain; H, heart; Lu, lung; K, kidney. The size of the HPRT mRNA is approximately 1500 nucleotides. The actin cDNA probe can hybridise to cytoplasmic actin mRNA (2100 nucleotides) and the lower molecular weight muscle actin mRNA (1650 nucleotides).

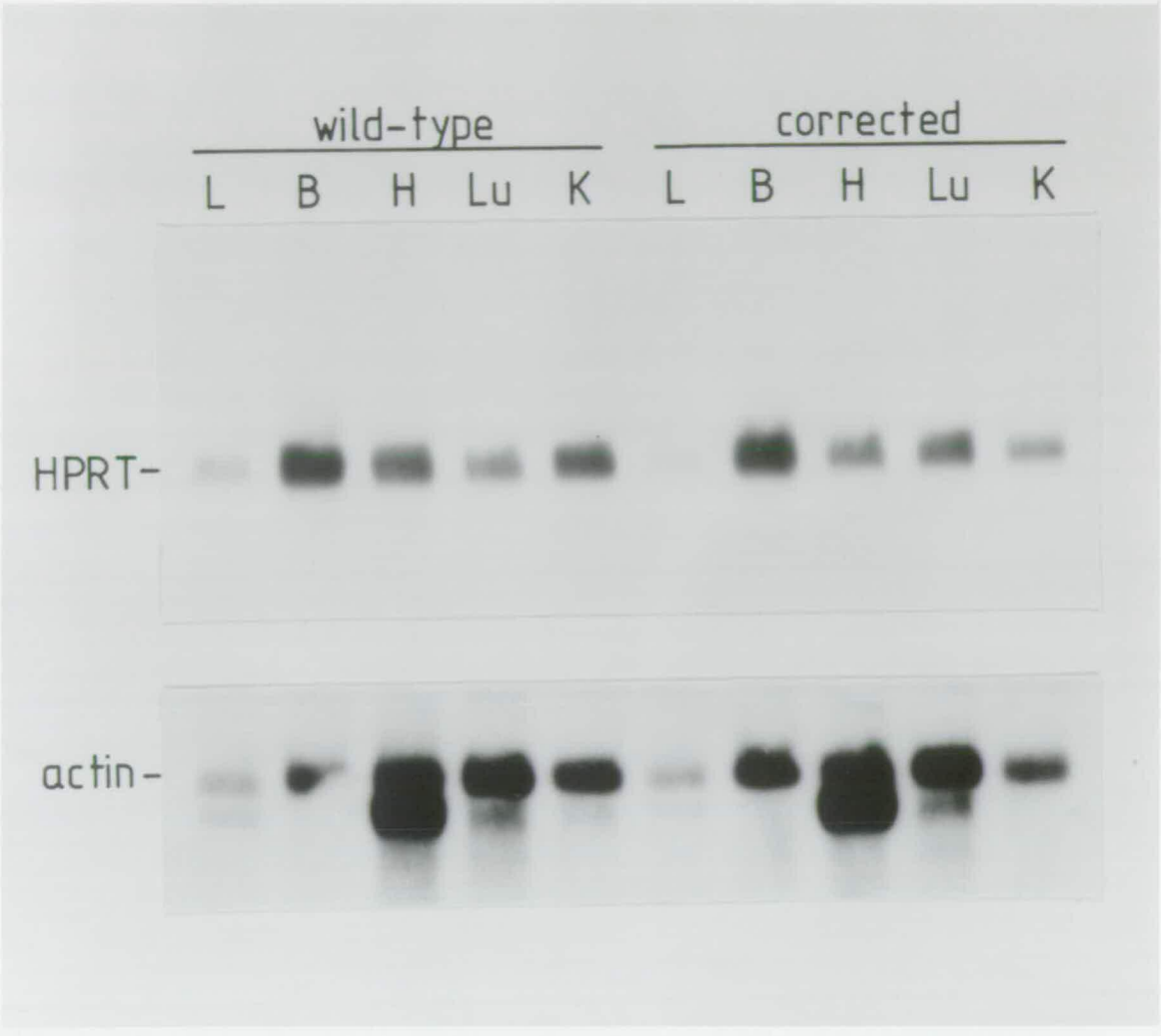

wild-type
corrected
L B H Lu K L B H Lu K
HPRT-
actin-

**Table 6.1 Densitometric analysis of HPRT expression in wild-type and corrected female tissues**

| | | H (HPRT x $10^{-4}$) | A (actin x $10^{-4}$) | H/A | % wild-type |
|---|---|---|---|---|---|
| Wild-type: | Liver | 1.1 | 0.4 | 2.8 | - |
| | Brain | 15.3 | 3.0 | 5.1 | - |
| | Heart | 7.0 | 5.8 | 1.2 | - |
| | Lung | 3.3 | 7.7 | 0.4 | - |
| | Kidney | 7.4 | 3.8 | 1.9 | - |
| Corrected: | Liver | 0.4 | 0.6 | 0.7 | 24 |
| | Brain | 10.7 | 6.2 | 1.7 | 33 |
| | Heart | 5.6 | 10.7 | 0.5 | 43 |
| | Lung | 4.3 | 8.9 | 0.5 | 114 |
| | Kidney | 2.0 | 1.9 | 1.1 | 55 |

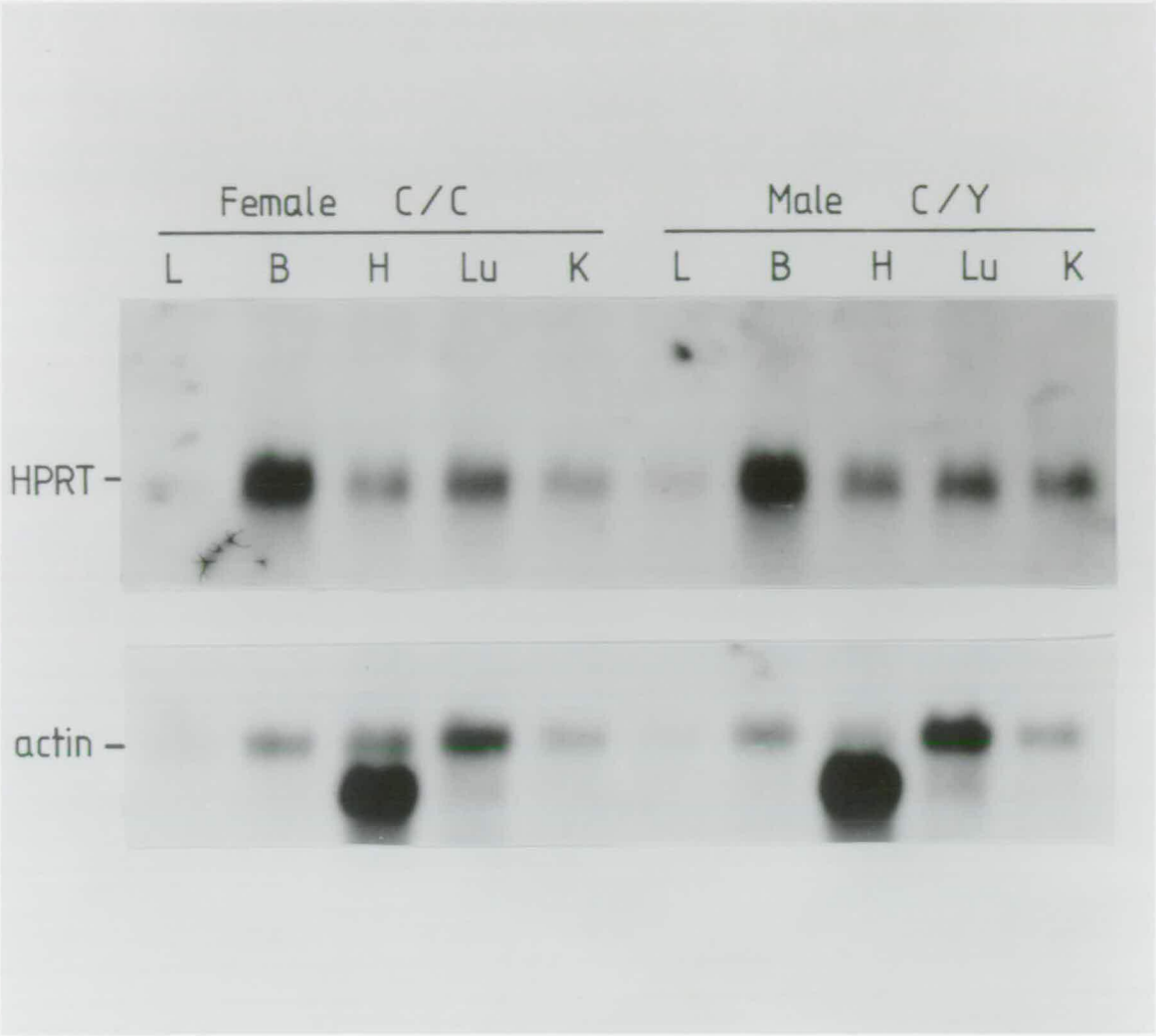
H and A values were obtained from the scanning densitometer and represent the area, in arbitrary units, under the peak corresponding to HPRT and actin mRNA bands, respectively. H/A equals the value H divided by the value A, i.e. it is the standardised HPRT value. % wild-type is the mRNA level in corrected tissues relative to wild-type expressed as a percentage, and equals (H/A corrected ÷ H/A wild-type) x 100.

mouse HPRT cDNA, pHPT4 (Konecki et al., 1982), and subsequently reprobed with the mouse α-actin cDNA (Minty et al., 1981) to allow comparison of mRNA levels for the same tissue in different animals, as explained in Chapter 4. The wild-type female tissues show the characteristic pattern of HPRT expression (see also Figure 4.3). By visual analysis of the autoradiograph the following can be inferred. In a wild-type mouse, liver has the lowest level of HPRT mRNA, while brain has the highest. HPRT expression in heart and kidney is approximately equivalent and at a level intermediate between liver and brain. The level of HPRT mRNA in lung is slightly lower than in heart and kidney. In the female homozygous for the corrected gene HPRT mRNA levels are lowest in liver and highest in brain. The heart and kidney HPRT expression levels are approximately equivalent and apparently at a correct level relative to liver and brain. However, the absolute level of expression in each of liver, brain, heart and kidney is approximately threefold lower than in the equivalent tissues of the wild-type female. In the female that is homozygous for the corrected HPRT gene the level of HPRT mRNA in lung is anomalous, it is higher than in heart and kidney, whereas in the wild-type mouse lung expression is lower than both heart and kidney. Thus, with the exception of lung the corrected gene is expressed at the correct relative level in each of the tissues analysed, but at a reduced absolute level. The actin reprobe shows that for each tissue there is not much variation in loading between mice.

Figure 6.3 shows Northern analysis of tissue RNAs from both a female homozygous for the corrected gene and a male hemizygous for the corrected gene. The level of HPRT expression for each tissue is approximately equivalent in the two animals. Although, the anomalously high expression in lung appears less pronounced in the male. Densitometric analysis was carried out of the HPRT and actin mRNA levels in both Figure 6.2 and Figure 6.3. The results of this analysis for Figure 6.2 are shown in Table 6.1. It has been shown that the level of HPRT mRNA in correctant ES cells in culture is 33% of the wild-type level. With the exception of lung the expression in each of the tissues of the female homozygous for the corrected gene is reduced to a similar level. The level of HPRT mRNA in liver is 24% of wild-type. The level in brain is 33% of wild-type and HPRT mRNA levels in heart and kidney are at 43% and 55% of wild-type, respectively. Densitometric analysis confirmed the anomalous expression in lung. The level of HPRT mRNA in the lung of the female which is homozygous for the corrected gene is 114% of the level in the lung of the wild-type female. Thus, the visual interpretation of these results was confirmed by densitometric analysis. The results of densitometric analysis of the experiment shown in Figure 6.3 are summarised in Table 6.2. For each tissue the H/A value is similar in the male to the female. Thus, HPRT expression in the male hemizygous for the corrected gene is equivalent to in the female homozygous for the corrected gene. The H/A value of 0.5 obtained for male lung compared to the value of 0.7 for female lung supports the suggestion that lung expression is

**Figure 6.3.**

Northern analysis of expression from the corrected HPRT gene in female and male mice. Total RNA (30μg) was prepared from mouse tissues, electrophoresed on formaldehyde-agarose gels, transferred, and hybridised with the HPRT cDNA probe and subsequently with the actin cDNA probe. Tissues are shown from a female homozygous for the corrected gene (C/C) and a male hemizygous for the corrected gene (C/Y). Tissues: L, liver; B, brain; H, heart; Lu, lung; K, kidney. The size of the HPRT mRNA is approximately 1500 nucleotides. The actin cDNA probe can hybridise to cytoplasmic actin mRNA (2100 nucleotides) and muscle actin mRNA (1650 nucleotides).


Female C/C
Male C/Y
L B H Lu K L B H Lu K
HPRT –
actin –

# 7. DISCUSSION

## 7.1 HPRT expression in transgenic mice

The use of cultured mammalian cells in DNA mediated gene transfer experiments has proved to be a valuable system for studying gene expression. Mutational analysis of the transfected genes has allowed the identification of cis-acting DNA elements that regulate expression. This approach has been used extensively to analyse the mouse HPRT promoter (Melton et al., 1986) as described in the Introduction to this thesis. However, the ability to define tissue-specific or temporal control elements using cultured cells is limited.Presently, the most satisfactory and widely used system for studying the tissue-specific and developmental control of gene expression is the transgenic mouse (for reviews see Jaenisch, 1988; Palmiter and Brinster, 1986). The preferred method for producing transgenic mice is by direct microinjection of DNA into one of the two pronuclei in a fertilised egg. The first step in analysing the regulation of expression of a gene is to identify a transgene construct which is expressed at levels comparable to and in the same tissue distribution as the endogenous gene. Subsequent mutational analysis of this 'complete' gene can be used to identify important elements more precisely.

Obviously, it is necessary to monitor the expression of a transgene. Usually, the transgene will be introduced into mice which have endogenous expression of the gene under

investigation. A variety of methods of marking transgene expression have been used. One method is to utilise a reporter gene, to which the promoter region being analysed is fused. The reporter gene may be the homologous gene from a different species, or it may be an unrelated gene which is easily monitored, such as a hormone which visibly affects physiology, or an oncogene which generates tumours. Alternatively, more subtle changes may be preferable. We have used an HPRT transgene which as far as possible represents the endogenous gene. The use of hybrid constructs including non-endogenous genes that grossly affect physiology may result in disturbances to the phenomenon under investigation. In addition sequences controlling gene expression are not confined to the region upstream of the transcription start site, as discussed in the Introduction. Hence, fusing a promoter region to a reporter gene may exclude important regulatory elements.

As well as being used to study gene expression, transgenic mice have been used extensively to study oncogenesis. The system has been invaluable in unravelling the actions and interactions of oncogenes (reviewed Groner et al., 1987). In addition, the microinjection technique has potential applications for the germ line manipulation of other mammalian species. It may be possible to improve livestock animals, (reviewed Pursel et al., 1989) for example, by making transgenic cattle with an improved protein content in their meat. Another approach under investigation is the production of transgenic animals which secrete

pharmaceutically or economically important products in their milk. Already, transgenic sheep which secrete human factor IX in their milk have been produced (Clark et al., 1989).

The factors which affect the efficiency of introducing DNA into fertilised eggs have been studied in detail (Brinster et al., 1985). For optimal integration a few hundred linear molecules of the transgene are injected into the male pronucleus of fertilised eggs. We obtained a single HPRT transgenic mouse line. For reasons considered below this limits what can be inferred from the expression pattern of the transgene. In this line approximately nine copies of the transgene are integrated in a head to tail array at a single autosomal site. Other workers have identified transgenic mice with a single copy or greater than 100 copies of the transgene, usually integrated at a single chromosomal site in a head to tail array (Brinster et al., 1985). In such a multicopy array it is not possible to determine if all the genes are being expressed equivalently or whether just one or a few copies are predominantly expressed. Frequently the level of expression is not copy number dependent. Indeed, it has been shown in mice transgenic for an elastase 1 gene that high copy numbers are detrimental to expression (Davis and MacDonald, 1988). It was observed that the rate of transcription per copy in mice with 250 copies was 100-fold less than in mice with 2 copies. For a copy number below 10 the level of transgene expression increased with increasing copy number. Thereafter, the level of transcription of the transgene declined. This effect is not due to the titration

of essential transcription factors because the level of expression of the endogenous gene remained unaffected by the copy number of the transgene.

We were unable to detect HPRT enzyme activity derived from the transgene in transgenic mice. The level of expression which we subsequently identified by mRNA analyses is probably below the threshold of detection in the enzyme assay. Transgene derived HPRT mRNA was detected in all transgenic tissues analysed. As for endogenous HPRT gene expression, the level of transgene expression was elevated in brain tissue relative to liver, heart and kidney. Thus, the HPRT transgene, pDWM1LS+127/+135 contains the regulatory elements responsible for the brain specific elevation of normal HPRT expression. The responsible signals could regulate expression at the level of the rate of transcription. Such elements could be situated anywhere in the HPRT transgene sequence. While transcriptional enhancers are frequently found upstream of promoters they can also be situated within or downstream of genes. For example the immunoglobulin heavy chain genes have an enhancer situated in an intron (Banerji et al., 1983) and the human β-globin gene has an enhancer situated 3' of the gene which contains multiple binding sites for an erythroid specific protein factor (Wall et al., 1988). Experiments in transgenic mice have also shown that many genes require sequences which are situated many kilobases away from the gene to control expression with the correct tissue specificity. To direct expression to the liver of transgenic

mice, the albumin gene requires an enhancer which is situated 10kb upstream of the promoter (Pinkert et al., 1987). On the other hand, much more limited amounts of flanking sequence are sufficient to give the correct pattern of expression of some genes in transgenic mice. A hybrid gene consisting of the bacterial β -galactosidase coding sequence and sequences from the mouse α -crystallin gene promoter, spanning from 759bp upstream to 45bp downstream of the transcription start site, gave lens specific expression with the appropriate temporal control in transgenic mice (Goring et al., 1987). Furthermore, a hybrid gene containing 213bp of 5' flanking sequence from the mouse elastase 1 gene upstream of the human growth hormone structural gene has the elastase 1 pattern of expression (Ornitz et al., 1985). That is, human growth hormone is detected only in pancreatic acinar cells. Hence, it would appear that there are few constraints on the organisation of cis-acting regulators of transcription.

Alternatively, HPRT expression may be regulated at the level of RNA processing or mRNA stability, in which case the significant elements would be situated within the transcribed sequences. The high level of sequence conservation in the 3' untranslated region of the HPRT gene has already been discussed in relation to a possible role in gene expression in the Introduction. The steady-state level of mRNA from many housekeeping genes is apparently controlled by the rate of mRNA degradation rather than transcription (Carneiro and Schibler, 1984). If the brain-specific elevation in HPRT expression is controlled by

regulated mRNA turnover then it would be expected, perhaps, that very limited flanking sequence would be sufficient. The promoter region would need only to contain general transcription factor binding sites. All the elements required for transcription of an HPRT minigene in Chinese hamster fibroblasts are located within 50bp upstream of the major transcription site (Melton et al., 1986). This sequence contains putative binding sites for the general transcription factors Sp1 and AP2, and may be sufficient for correct HPRT expression in vivo, if regulation is predominantly post-transcriptional.

While pDWM1LS+127/+135 expression was elevated in brain relative to most tissues, the absolute level of expression in each tissue, with the exception of lung, was at least 10-fold below the wild-type level. To facilitate the quantification of transgene expression, mice were produced which were transgenic and which were deficient for endogenous HPRT expression. In this situation it is not necessary to distinguish between mRNA derived from the transgene and mRNA derived from the endogenous gene. However, it is now necessary to generate an internal control for variations in loading. This was achieved by measuring the endogenous actin mRNA (Minty et al., 1981) level in each tissue. Standardising the HPRT mRNA level relative to the actin mRNA level allowed the direct comparison of the HPRT expression in a particular tissue between different mice. However, because actin mRNA levels vary between tissues it does not provide a standard for comparing the level of HPRT mRNA in

one tissue with the level of HPRT mRNA in another tissue. Comparisons between tissues rely on the accuracy with which it is possible to load 30µg of RNA per track. Consequently, the values derived from densitometric analysis for HPRT mRNA levels are intended only to support what visual analysis of each signal indicates. Differences of less than 10% of the wild-type level are not significant.

With the exception of lung and spleen the HPRT mRNA levels in transgenic tissues are considerably lower than wild-type levels. Visual and densitometric analysis indicate that the level of the transgene-derived HPRT mRNA in each of liver, brain, heart and kidney is less than 10% of the wild-type level. There are a few possible reasons for this. The HPRT minigene, while having the necessary control signals for brain-specific elevation of expression, may lack one or more control elements which enhance the absolute level of transcription in all cell-types. Sequences which are located approximately 50kb 5' of and 20kb 3' of the human -globin gene are necessary for wild-type levels of expression of a β -globin transgene (Grosveld et al., 1987). When included in a transgene these sequences confer a position independent, copy number dependent, high level of expression. It has been suggested that these sequences contain nuclear-matrix binding sites and we cannot exclude the possibility of analogous sequences at the HPRT locus. In the pDWM1LS+127/+135 transgene HPRT introns one to six are deleted, only introns seven and eight remain. It is possible that one or more of the missing HPRT introns is required for

correct expression in transgenic mice, although minigenes lacking these introns are efficiently expressed in cultured cells (Melton et al., 1986). The HPRT promoter is situated within a CpG island, as defined by Bird (1986). Because of the absence of the first HPRT intron only half of the CpG island is present in pDWM1LS+127/+135. This could potentially inhibit transcription either by altering the topography or the methylation pattern of the promoter region. Alternatively, specific introns may be required for correct RNA processing or transport.

Brinster et al., (1988) reported a 10- to 100-fold higher transcriptional efficiency for genes containing introns than for the same genes without introns in transgenic mice. It was found that the presence of introns in four transgenes increased both the frequency of obtaining expression in transgenic lines and the level of expression in those lines. There was no difference in the level of expression between the same genes, with or without introns, in cultured cell transformation experiments. Clearly when genes are introduced into the mouse germ line, and consequently experience the influences of development and differentiation, they have additional requirements for expression than when introduced into terminally differentiated cultured cells. Brinster et al., (1980) demonstrated with transcription run-on experiments on isolated nuclei that an increased rate of transcription was responsible for the increased expression of transgenes containing introns. One of the transgenes (human β-globin)

is known to contain an enhancer-like element within an intron. However, no similar elements have been identified in the introns of the other genes used. In these cases, how the presence of introns affects the rate of transcription in vivo but not in cultured cells is not known. Using transgenic mice, elements in the first intron of the human ADA gene have been identified which are essential for appropriate expression of a reporter CAT gene linked to the ADA promoter (Aronow et al., 1989). ADA is a housekeeping gene but exhibits a level of expression in the thymus which is 100-fold higher than most other tissues. An intron region 4 to 10kb downstream of the first exon contains a complex pattern of DNaseI hypersensitive sites which vary according to tissue and cell type. The inclusion of this region in transgene constructs leads to reproducible, high-levels of expression in the appropriate tissues of transgenic mice.

Alternatively, the level of HPRT transgene expression may be low because of an inhibitory effect of the chromosomal environment into which it is integrated. The transgene may have integrated into a transcriptionally silent region. This, the least interesting of the possible explanations, could be tested by generating more HPRT transgenic mouse lines, in which the integration site would be different. Unfortunately only one transgenic line was obtained in this experiment, and so position effects cannot be ruled out.

While the level of HPRT transgene expression in the liver, brain, heart and kidney is less than 10% of wild-type levels,

the level of HPRT transgene expression in lung, and to a lesser extent spleen, is anomalously high. Densitometric analysis of Northern hybridisations indicates that in lung the level of HPRT transgene expression is at least equal to and possibly higher than the wild-type level. The simplest explanation for this is that the transgene has integrated into a position where it is under the influence of an endogenous lung-specific enhancer of transcription. This is not entirely compatible with the idea that transgene expression is low in most tissues because the integration site is in a transcriptionally silent region. In addition, the finding that the expression of the mutant HPRT locus corrected by gene targeting is also slightly high in lung is somewhat provocative. Both the HPRT transgene and the corrected HPRT gene have the same 650bp of 5' flanking sequence. Although, the degree of lung-specific elevation relative to other tissues is considerably less for the corrected gene than for the transgene, it is tempting to try to formulate a single model which would explain both observations.

The transgenic mouse experiment has demonstrated that the regulatory signals responsible for the elevated expression of HPRT in mouse brains are contained within the HPRT minigene. The interpretation of the expression pattern in transgenic mice is complicated by the possibility of position effects and effects due to transgene copy number. These complications are overcome by using gene targeting in ES cells to manipulate and study the expression of the

endogenous HPRT gene.

## 7.2 Targeted correction of a mutant HPRT gene in ES cells

The mutation in the HPRT deficient ES cell line E14TG2a was shown to be a deletion of the promoter and first two exons of the HPRT gene, and hence cannot revert spontaneously. Using gene targeting exons 1 and 2 and 650bp of 5' flanking sequence were replaced, thus restoring HPRT activity. Targeting vectors have been defined as either sequence insertion vectors or sequence replacement vetors (Thomas and Capecchi, 1987). When homologous recombination takes place sequence insertion vectors integrate without the loss of chromosomal sequences and a duplication of the region of homology between the vector and the target sequences results. Homologous recombination with sequence replacement vectors leads to the substitution of genomic sequences for vector sequences with no consequent sequence duplication. The targeting vector, pDWM101, used to correct the E14TG2a deletion is a sequence insertion vector. Thomas and Capecchi (1987) have shown that, when the amount of homology with the target locus is equivalent, both types of vector are equally efficient at gene targeting. It is not possible to construct a replacement vector analogue of pDWM101 without first cloning genomic sequences from upstream of the E14TG2a deletion.

Before introduction into E14TG2a cells the pDWM101 vector

was linearised at a unique restriction site in exon 3, that is within the region of homology. In yeast transformation experiments the generation of free DNA ends within the region of homology increases the frequency of recombination between 10- and 1000-fold, dependent upon the specific vector used (Orr-Weaver et al., 1981). For some vectors linearisation outside the region of homology increases the frequency of recombination 100-fold but for other vectors no enhancement is achieved. In cultured mammalian cells the frequency of recombination between autonomously replicating (non-integrated) plasmids is increased by between 5- and 80-fold if a double strand gap is introduced into the region of homology (Kucherlapati et al., 1984).

I have attempted to obtain a value for the frequency of homologous recombination events relative to random integration events, which we have called the targeting index. In our experiments cells were electroporated with pDWM100 DNA as a positive control for transformation. pDWM100 is derived from pDWM101 but, unlike the targeting vector, pDWM100 has the entire HPRT protein coding region and consequently has independent HPRT activity. pDWM100 restores HPRT activity to E14TG2a cells regardless of the integration site, provided that expression is not inhibited by the chromosomal environment. In three identical electroporation experiments the targeting indexes obtained were 0.007, 0.006 and 0.035. HPRT was also the target gene in experiments described by Thomas and Capecchi (1987). They produced a series of vectors to inactivate the HPRT gene in

wild-type ES cells. The inactivating vectors each contained a modified bacterial neomycin phosphotransferase (neo) gene inserted into the eighth exon of the HPRT gene. The neo gene used, pMC1Neo, is transcribed from the Herpes simplex virus thymidine kinase gene promoter. To improve its expression in ES cells it also contains a duplication of a 65bp sequence from the PyF441 polyoma virus enhancer and a modified translation initiation signal. In addition to disrupting the HPRT coding sequence, pMC1Neo provides a selectable marker for transformation. Cells expressing the neo gene are resistant to the drug G418. The number of G418 resistant colonies generated per cell electroporated with targeting vector DNA in these experiments is analogous to the number of $HAT^R$ colonies generated per E14TG2a cell electroporated with pDWM100 in our experiments. Thomas and Capecchi (1987) identified clones in which the HPRT gene had been inactivated by gene targeting by resistance to 6-TG. In their experiments the number of clones which are both $G418^R$ and $6\text{-}TG^R$ expressed relative to the number which are $G418^R$ is equivalent to what we have called the targeting index. They have shown that this frequency is dependent upon the extent of sequence homology between the vector and the target locus. For sequence replacement vectors frequencies ranged from 1/40,000 for 4.0kb of sequence homology to 1/950 for 9.1kb of homology. Using sequence insertion vectors frequencies ranged from 1/19,000 for a vector with 3.7kb of homology,. to 1/1,100 for a vector with 9.3kb of homology. The highest of these frequencies is 10-fold lower than the frequency obtained with pDWM101, which has between 2.3kb and 4.2kb of

homology with the target HPRT gene.

The experiments discussed above using the HPRT gene demonstrate that gene targeting in ES cells can be achieved at a frequency which makes it applicable to genes the activity of which is not intrinsically selectable. To isolate HPRT correctants we have exploited the systems that are available to select both for and against HPRT activity in cultured cells. Strategies have also been devised to enrich for targeting into non-selectable genes. It is possible to enrich for gene targeting events by using a targeting vector which contains a modified selectable marker that depends on chromosomal sequences at the target site for its expression. Sedivy and Sharp (1989) used a neo gene which lacked both transcription and translation initiation signals to construct targeting vectors for inactivation of a polyoma middle T antigen (pmt) gene in the MT-1.4 cell line. MT-1.4 cells are NIH3T3 cells which contain the oncogenic pmt gene as a stable haploid locus. Targeted integration generates a pmt-neo fusion protein whereas the majority of integrations do not result in neo expression. The neo gene used had the additional modification of a 'bridge' peptide derived from the amino-terminal end of the chloramphenicol acetyltransferase gene. This 'bridge' was intended to provide a spacer between the neo enzyme and the potentially inhibitive influence of the pmt portion of the fusion protein. This strategy gave an approximately 100-fold enrichment for gene targeting by reducing the frequency of random integrations which lead to $G418^R$. In the experiments

reported by Sedivy and Sharp (1989) the frequency of targeted inactivants per $G418^R$ colony was between 1/100 and 1/1000. An additional screening was possible in this instance because inactivation of the pmt gene lead to a reversal of cell morphology to the parental non-transformed phenotype. A significant proviso for this strategy is the requirement that the target gene is expressed, for transcription and translation of the neo gene.

In similar experiments Dorin et al., (1989) achieved targeted integration of a neo gene into the SV40 large T antigen. As in the experiments described previously G418 resistance depended upon endogenous promoter elements and read-through transcription. These experiments were performed in C121, a human-mouse hybrid cell line in which the SV40 large T antigen is constitutively expressed from an SV40 array integrated on human chromosome 7. In this instance the strategy was particularly successful and further analysis of G418 resistant colonies using the polymerase chain reaction (PCR) and Southern hybridisations showed that 3 out of 11 $G418^R$ clones were correctly targeted. As the authors note, this high frequency may be peculiar to this experiment perhaps as a result of the multiple copies of the T antigen gene present in these cells or as a result of the possible recombinogenic function of the SV40 T antigen protein. Subsequently, the neo marker on chromosome 7 of the targeted clones was exploited to generate human-mouse hybrids that contained the target locus and just a small fragment of chromosome 7 in chromosome-mediated gene

transfer experiments. This may provide a general strategy for isolating new regions of the human genome. Jasin and Berg (1989) have also achieved enrichment for gene targeting by using a defective gene as the positive selectable marker. Their target gene was the single integration of the SV40 early region transcription unit in COS1 cells. As the selectable marker the targeting vectors contained the bacterial xanthine-guanine phosphoribosyltransferase (gpt) gene, which can be selected for with mycophenolic acid. The gpt gene in the targeting vectors has the SV40 early region promoter but is transcriptionally impaired due to the absence of an enhancer. Homologous integration into the SV40 locus provides the gpt gene with the SV40 early region enhancer and thus restores transcriptional activity. Hence, only clones in which homologous recombination had occurred or in which the gpt gene had fortuitously integrated into another region with activating properties, would become mycophenolic acid resistant. The use of the transcriptionally defective gpt gene reduced the number of mycophenolic acid resistant clones with random integrations to approximately 100-fold less than obtained with a fully active gpt gene. Southern blot analysis of mycophenolic acid resistant clones obtained in six calcium phosphate transfection experiments revealed that 25 of 57 were homologous recombinants. To study its role in mammalian development Joyner et al., (1989) have inactivated the En-2 gene in mouse ES cells. En-2 is the mouse homologue of the Drosophila gene engrailed, which is thought to have a role in establishing spatial domains in the developing brain. The

strategy they used combined selection for the neo gene with screening by PCR. The neo gene Joyner et al., (1989) used had independent transcriptional activity due to the presence of 500bp from the human β-actin promoter, and therefore acted as a marker for transformation. In the targeting vector, a sequence replacement vector, the neo gene was flanked by 3kb of homology with the target gene on one side and 700bp on the other. The vector was introduced into ES cells by electroporation and selection for the neo gene was applied. PCR was used to identify targeted clones in pools of G418 resistant colonies. The two oligonucleotide primers for PCR were designed so that amplification could occur only if targeted integration had occurred. That is, one primer was positioned in the neo gene while the other was positioned in the target En-2 locus, outside the region of homology. In targeted clones PCR generates an 800bp fragment. In any integration other than homologous recombination the primers will not be in the correct arrangement to initiate the PCR amplification.

It is generally believed that ES cells are more likely to retain their pluripotency if <u>in vitro</u> manipulations are kept to a minimum. Joyner et al., (1989) used a strategy which avoids the large number of rounds of colony pooling necessary in standard sib-selection schemes. Electroporated cells were plated at a dilution to produce approximately ten $G418^R$ colonies per dish. A few hundred cells were picked from each colony, pooled, subjected to PCR and analysed for the presence of the 800bp fragment diagnostic of gene targeting.

Having identified a positive plate colonies were screened individually by PCR, and positives were checked by Southern analysis. The frequency of homologous recombination in three experiments was one event per 260 $G418^R$ ES cell colonies. The highest frequency obtained by Thomas and Capecchi (1987) in HPRT inactivation experiments using a transcriptionally active neo gene was 1/1000. The frequency of random integrations leading to $G418^R$ may be lower in the En-2 experiment due to the absence of polyadenylation signals in the neo gene. Also the human β-actin promoter is weaker than the promoter in pMC1Neo. Zimmer and Gruss (1989) used PCR alone to screen for targeted clones among cells into which vector DNA had been introduced by nuclear microinjection. They inactivated the homoeobox Hox 1.1 gene in ES cells with the aim of producing mice deficient for this function, to elucidate its role in development. The targeting vector used contained a minimal amount of non-homology with the target Hox 1.1 gene, consisting of a genomic subclone with a 20bp oligonucleotide introduced at one site. The oligonucleotide has two functions. Firstly, it changes the reading frame of the Hox 1.1 protein coding region and consequently leads to premature translation termination. Secondly, it provides a site for a PCR primer. The second PCR primer is unique to the target gene. Only the gene structure which results from homologous recombination can give an amplified fragment of 1.1kb from PCR. The average frequency of gene targeting that Zimmer and Gruss (1989) obtained was 1 event per 150 microinjected cells. The authors believe that this represents a ratio of homologous to

illegitimate recombination of 1:30, assuming that 20% of microinjected cells become stably transformed (Capecchi, 1980). Although the microinjection procedure is labour intensive, it would seem to be practical due to the high targeting frequency it apparently generates. A different strategy to enrich for gene targeting events has been described by Mansour et al., (1988). In addition to enabling selection for cells which have integrated vector DNA, this procedure allows selection against cells in which vector DNA has integrated non-homologously. The positive selection is provided by pMC1Neo (Thomas and Capecchi, 1987). This neo gene construct which has independent activity is positioned in the targeting vector so as to be flanked by sequences with homology to the target locus and disrupt the coding sequence of the target gene. $G418^R$ can result from both random and homologous integrations. To enrich for gene targeting a negative selectable marker is built into the targeting vector outside the region of homology with the target locus. Mansour et al., (1988) used the Herpes simplex virus thymidine kinase (HSV-tk) gene, with the same promoter and enhancer combination used in pMC1Neo, as the negative marker. Cells which express HSV-tk are selectively killed in medium containing $2 \times 10^{-6}$M gancyclovir (GANC). Because the HSV-tk gene is outside the region of homology with the target locus when homologous recombination occurs the HSV-tk gene does not integrate into the genome. Consequently, cells in which gene targeting has occurred do not express HSV-tk and are not sensitive to GANC. If targeting vector DNA integrates randomly the HSV-tk gene becomes stably

integrated into the genome, and cells in which this occurs can be selectively killed by GANC. Thus gene targeting events are enriched for by selection of transformed cells in G418 and GANC. Mansour et al., (1988) tested this strategy by inactivating the HPRT gene which can be directly selected for using 6-TG. Out of 24 clones isolated by G418 and GANC selection, 19 were shown to be homologous recombinants. Thus, the positive-negative selection procedure provides a 2000-fold enrichment relative to positive selection for pMClNeo alone (Thomas and Capecchi, 1987). The same strategy was used to inactivate the int-2 proto-oncogene (Mansour et al., 1988). Four clones were identified with targeted inactivation from 81 $G418^R$, $GANC^R$ clones tested. Cells with random integrations can survive GANC selection if the site of integration inhibits HSV-tk expression or if the HSV-tk gene acquires a mutation. The authors claim that the addition of a large stretch of non-homology at one end of the integrating vector does not affect the efficiency of homologous recombination, which conflicts with evidence discussed earlier from transformation experiments in yeast (Orr-Weaver et al., 1981) and mammalian cells (Kucherlapati et al., 1984).

Comparisons between targeting frequencies obtained in different experiments are complicated by different selection systems, different extents of homology, different cell types and so forth. However, it is clear that there are locus specific variations. It is possible that the transcriptional activity of a gene can influence its

accessibility to targeting vector DNA. Recent work described by Johnson et al., (1989) suggests this is not the case. Using the positive-negative selection system (Mansour et al., 1988) they compared the frequencies of obtaining targeted inactivation of three genes. The three genes were c-fos, which is expressed at a low level in ES cells, and adipsin and adipocyte P2 ( aP2 ) which are not expressed at detectable levels in ES cells. No difference was found in the frequency of homologous recombination between these genes. The number of targeted inactivants per $G148^R$ clone for each gene was comparable to that for HPRT (Thomas and Capecchi, 1987). Of more practical significance between 5 and 10% of the colonies which were both $G418^R$ and $GANC^R$ were targeted. This result is particularly important because it means that targeted manipulation of the mouse germ line is not confined to genes that are expressed in ES cells.

I have presented the results of the structural analysis of the HPRT gene in 22 $HAT^R$ clones obtained in the targeted correction experiments. In twenty of these clones the HPRT gene had the predicted structure, which we have called type 1. One example was found of each of two unpredicted structures, called type 2 and type 3. The type 3 structure has two copies of the correcting vector integrated in tandem. In the type 2 structure there has been a small deletion of the upstream copy of the duplicated exon 3. This deletion is in the region of homology between the targeting vector and the chromosomal gene where recombination is initiated. Small deletions have also been observed in HPRT inactivation

experiments (Doetschman et al., 1988) using a replacement vector with a neo gene inserted into the third exon. Southern analysis of six targeted inactivants showed that in two an EcoRI site at the upstream end of vector sequences was absent. The structure of all six targeted clones was analysed further by amplifying the upstream boundary between vector and chromosome by PCR and subsequent sequencing. The two clones lacking the EcoRI site had small deletions (14bp and 37bp) which were immediately 5' to the region of homology. These small deletions and the type 2 structure we have characterised were only identified because they deleted one of the restriction sites used in Southern analysis. It is possible that such deletions occur at a high frequency but go undetected. The deletions observed by Doetschman et al., (1988) may be the result of the low amount of homology between the vector and the target gene at this end of the vector (132bp).

Clearly the technology now exists for introducing defined mutations into any gene in ES cells provided a genomic clone is available. We have demonstrated the feasibility of introducing targeted modifications into the mouse germ line. This provides a powerful approach for studying mammalian gene expression. The function of putative gene regulatory sequences can be tested directly by targeting mutations to them. Gene targeting used in conjunction with transgenic mice should allow very sophisticated gene expression studies. Of course, it is also possible to produce mice with mutations in any gene which has been identified as defective

in a human genetic disease. Thus, mouse models for inherited diseases of man can be generated to provide a system for testing new therapies. In particular, the development of somatic gene replacement therapies should benefit from both the availability of mouse models for diseases and the ability to make targeted gene corrections.

In addition, the targeted inactivation of a gene is the best way to determine its function if this is not known. This strategy is being attempted for many genes thought to have a role in development. Recently, the generation of mice with a germ line mutation of the c-abl gene has been reported (Schwartzberg et al., 1989). Targeted mutation of the c-abl gene in ES cells was achieved using a replacement vector containing a promoterless neo gene. On average, 1 in 34 $G418^R$ clones were homologous recombinants. Mice heterozygous for the c-abl mutation show no mutant phenotype. The c-abl gene product is believed to be involved in signal transduction. The usefulness of these mice may be greatly limited if homozygosity of the mutation is lethal early in development. Disruption of the β2-microglobulin (β 2-m) gene has also been achieved and introduced into the mouse germ line (Zijlstra et al., 1989). The β 2-m gene product is a component of MHC class 1 molecules which have a number of functions in the immune system and probably a number of other unidentified reactions. Zijlstra et al., (1989) used a targeting vector containing a neo gene, which lacked polyadenylation signals, to inactivate the β2-m gene. PCR was used to isolate ES cell clones with the targeted

inactivation. One out of twenty five of the $G418^R$ clones obtained were targeted inactivants. This frequency is high and may reflect the presence of a recombination hotspot. In addition the lack of polyadenylation signals in the neo gene enriches for targeting events approximately fourfold. Six male chimeras were obtained which transmitted the disrupted β2-m gene to their offspring. Mice heterozygous for the mutation are phenotypically normal. Homozygotes have yet to be generated.

## 7.3 Expression of the corrected HPRT gene

The expression of the corrected HPRT gene has been analysed both in cultured ES cells and in male and female mice. The quantification of HPRT mRNA and problems associated with it have been discussed in the context of HPRT expression in transgenic mice. In cultured ES cells containing the corrected gene the level of HPRT mRNA is approximately threefold lower than in ES cells containing the wild-type gene. In mice with the corrected HPRT gene the level of HPRT mRNA in liver, brain, heart and kidney is between 2- and 3-fold lower than in wild-type mice. However, while the absolute level of HPRT mRNA in these tissues is low the wild-type pattern of relative levels is maintained. That is liver has the lowest HPRT mRNA level, kidney and heart have higher levels, and HPRT mRNA is most abundant in brain. Like the HPRT transgene pDWM1LS+127/+135 the corrected gene contains the regulatory information for elevated levels of expression

in brain. Both the transgene and the corrected gene have 635bp of 5' flanking sequence. Both genes also have the natural HPRT 3' untranslated region.

As for the transgene, there are a number of possibilities why the corrected gene is expressed at a lower level than the wild-type gene. Firstly an important enhancer of transcription may be absent. The region from 635bp to greater than 12kb upstream of the wild-type gene is deleted in the corrected allele. Furthermore, 6.7kb of sequence is deleted from the first intron of the corrected gene. A second possibility is that as a result of the upstream deletion new sequences are brought into proximity with the HPRT promoter and these may have an inhibitory effect on transcription. The corrected gene has 2.7kb of prokaryotic vector DNA upstream of the 635bp of HPRT 5' flanking sequence. The expression of certain genes in transgenic mice is strongly inhibited by prokaryotic sequences (Jaenisch 1988).

While the level of HPRT mRNA from the corrected gene is low in liver, brain, heart and kidney, in lung the level of HPRT mRNA is close to the wild-type level. The degree to which lung tissue is different from other tissues in corrected mice is small in comparison with the lung specific elevation in transgenic mice. The corrected gene is in the natural HPRT location on the X-chromosome and therefore is not subject to position effects to the same extent as a transgene. However, as a result of the deletion in E14TG2a new sequences are in

proximity with the promoter of the corrected gene and these could be responsible for the reduced absolute level of expression in most tissues and the higher level of expression in lung. If the same mechanism is causing the elevated HPRT expression in the lung of both transgenic and corrected mice it must be a property of the 635bp of 5' flanking sequence. This is the only characteristic which is different from the wild-type gene but common to both the corrected gene and the transgene. Perhaps the end point of the upstream flanking sequence is within a regulatory region. Enhancers are modular (Dynan, 1989), they consist of multiple overlapping binding sites for different transcriptional activator proteins with different tissue specificities. The loss of part of a regulatory region may alter its transcription enhancing properties. Alternatively, there may be a lung specific repressor of transcription upstream of -635bp, which is not present in either the transgene or the corrected gene. However, the difference in the degree of the lung-specific elevation of HPRT expression between the corrected gene and the transgene perhaps implies that the causal mechanisms are different, and that their coincidence is merely a red herring.

## 7.4 Concluding Remarks

We have demonstrated that the brain-specific elevation in HPRT enzyme activity in mice (Lo and Palmour, 1979) correlates with an elevation in levels of HPRT mRNA. We have

also shown that no more than 635bp of 5' flanking sequence is required to achieve the elevated expression in brain tissue. The interesting possibility that the variations in HPRT activity between tissues are regulated by the HPRT 3' untranslated region invites further investigation. Using gene targeting in ES cells it will be possible to make precise alterations to the HPRT 3' untranslated region and subsequently observe the effect on the tissue dependent variations in HPRT mRNA levels. It will also be interesting to analyse the rate of HPRT transcription and HPRT mRNA degradation in different tissues. We would predict that targeted manipulations of the 3' untranslated region would affect mRNA stability rather than the rate of transcription. Very sophisticated manipulation and investigation of HPRT gene expression is now possible.

## REFERENCES


Aronow, B., Lattier, D., Silbiger, R., Dusing, M., Hutton, J., Jones, G., Stock, J., McNeish, J., Potter, S., Witte, D. and Wiginton, D. (1989) Genes Dev. **3**, 1384-1400.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) Cell **27**, 299-308.

Banerji, J., Olson, L. and Schaffner, W. (1983) Cell **33**, 729-740.

Baumeister, A.A. and Frye, G.D. (1985) Neurosci. Biobehav. Rev. **9**, 169-178.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. **8**, 127-151.

Benoist, C. and Chambon, P. (1981) Nature **290**, 304-310.

Bergmann, I.E. and Brawerman, G. (1977) Biochem. **16**, 259-264.

Berk, A.J. and Sharp, P.A. (1977) Cell **12**, 721-732.

Bienz, M. and Pelham, H.R.B. (1986) Cell **45**, 753-760.

Bird, A.P. (1986) Nature **321**, 209-213.

Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. **7**, 1513-1523.

Boyd, E.M., Dolman, M., Knight, L.M. and Sheppard, E.P. (1965) J. Physiol. Pharmacol. **43**, 995-998.

Bradley, A., Evans, M., Kaufman, M.H. and Robertson, E. (1984) Nature **309**, 255-256.

Brennand J., Chinault, A.C., Konecki, D.S., Melton, D.W. and Caskey, C.T. (1982) Proc. Natl. Acad. Sci. USA **79**, 1950-1954.

Brinster, R.L., Chen, H.Y., Trumbauer, M.E., Yagle, M.K. and Palmiter, R.D. (1985) Proc. Natl. Acad. Sci. USA **82**, 4438-4442.

Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E. and Palmiter, R.D. (1988) Proc. Natl. Acad. Sci. USA **85**, 836-840.

Brown, J.R., Daar, I.O., Krug, J.R. and Maquat, L.E. (1985) Mol. Cell. Biol. **5**, 1694-1706.

Buratowski, S., Hahn, S., Guarente, L. and Sharp, P.A. (1989) Cell **56**, 549-561.

Capecchi, M.R. (1980) Cell **22**, 479-488.

Carneiro, M. and Schibler, U. (1984) J. Mol. Biol. **178**, 869-880.

Chapman, V.M., Kratzer, P.G., Siracusa, L.D., Quarantillo, B.A., Evans, R. and Liskay, R.M. (1982) Proc. Natl. Acad. Sci. USA **79**, 5357-5361.

Clark, A.J., Bessos, H., Bishop, J.O., Brown, P., Harris, S., Lathe, R., McClenaghan, M., Prowse, C., Simons, J.P., Whitelaw, C.B.A. and Wilmut, I. (1989) Bio/Technol. **7**, 487-492.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedlinger, C. and Chambon, P. (1980) Science **209**, 1406-1414.

Davis, B.P. and MacDonald, R.J. (1988) Genes Dev. **2**, 13-22.

Davison, B.L., Egly, J., Mulvihill, E.R. and Chambon, P. (1983) Nature **301**, 680-686.

Derman, E., Krauter, K. Walling, L., Weinberger, C., Ray, M. and Darnell, J.E. (1981) Cell **23**, 731-739.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S. and Smithies, O. (1987) Nature **330**, 576-578.

Doetschman, T., Maeda, N. and Smithies, O. (1988) Proc. Natl. Acad. Sci. USA **85**, 8583-8587.

Dorin, J.R., Inglis, J.D. and Porteous, D.J. (1989) Science **243**, 1357-1360.

Dush, M.K., Sikela, J.M., Khan, S.A., Tischfield, J.A. and Stambrook, P.J. (1985) Proc. Natl. Acad. Sci. USA **82**, 2731-2735.

Dynan, W.S. (1989) Cell **58**, 1-4.

Dynan, W.S. (1989a) Trends Genet. **5**, 35-36.

Edwards, N.L., Jeryc, W. and Fox. I.H. (1984) Adv. Exp. Med. Biol. **165**, 23-26.

Eisenmann, D.M., Dollard, C. and Winston, F. (1989) Cell **58**, 1183-1191.

Epstein, C.J. (1970) J. Biol. Chem. **245**, 3289-3294.

Evans, M.J. and Kaufman, M.H. (1981) Nature **292**, 154-156.

Farnham, P.J., Abrams, J.M. and Schimke, R.T. (1985) Proc. Natl. Acad. Sci. USA **82**, 3978-3982.

Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6-13.

Finger, S., Heavens, R.P., Sirinathsinghji, D.J.S., Kuehn, M.R. and Dunnett, S.B. (1988) J. Neurol. Sci. **86**, 203-215.

Fire, A., Samuels, M. and Sharp, P.A. (1984) J. Biol. Chem. **259**, 2509-2516.

Fromm, M. and Berg, P. (1983) Mol. Cell. Biol. **3**, 991-999.

Fuscoe, J.C., Fenwick, R.G., Ledbetter, D.H. and Caskey, C.T. (1983) Mol. Cell. Biol. **3**, 1086-1096.

Georgiev, O. and Birnstiel, M.L. (1985) EMBO J.**4**, 481-489.

Gidoni, D., Dynan, W.S. and Tjian, R. (1984) Nature **312**, 409-413.

Goldstein, M., Kuga, S., Kusano, N., Meller, E., Dancis, J. and Schwarcz, R. (1986) Brain Res. **367**, 114-117.

Goring, D.R., Rossant, J., Clapoff, S., Breitman, M.L. and Tsui, L.C. (1987) Science **235**, 456-458.

Graham, F.L. and van der Eb, A.J. (1973) Virol. **52**, 456-467.

Graves, J.A.M. (1982) Exptl. Cell Res. **141**, 99-105.

Groner, B., Schonenberger, C-A. and Andres, A.C. (1987) Trends Genet. **3**, 306-308.

Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA **77**, 1432-1436.

Grosveld, G.C., de Boer, E., Shewmaker, C.K. and Flavell, R.A. (1982) Nature **295**, 120-126.

Grosveld, F., Blom van Assendelft, G., Greaves, D.R. and Kollias, G. (1987) Cell **51**, 975-985.

Gutensohn, W. and Guroff, G. (1972) J. Neurochem. **19**, 2139-2150.

Hahn, S., Buratowski, S., Sharp, P.A. and Guarente, L. (1989) Cell **58**, 1173-1181.

Handyside, A.H., O'Neill, G.T., Jones, M. and Hooper, M.L. (1989) Roux Arch. Dev. Biol. **198**, 48-56.

Herr, W. and Gluzman, Y. (1985) Nature **313**, 711-714.

Herr, W. and Clarke, J. (1986) Cell **45**, 461-470.

Hooper, M.L. (1982) Biochim. Biophys. Acta **651**, 85-103.

Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987) Nature **326**, 292-295.

Horikoshi, M., Wang, K.C., Fujii, H., Cromlish, J.A., Weil, P.A. and Roeder, R.G. (1989) Nature **341**, 299-303.

Isamat, M., Macleod, K.F., King, A., McEwan, C. and Melton, D.W. (1988) Somat. Cell Mol. Genet. **14**, 359-369.

Jaenisch, R. (1988) Science **240**, 1468-1475.

Jasin, M. and Berg, P. (1988) Genes Dev. **2**, 1353-1363.

Jinnah, H.A., Gage, F.H. and Friedmann, T. (1989) (personal communication, submitted).

Johnson, R.S., Sheng, M., Greenberg, M.E., Kolodner, R.D., Papaioannou, V.E. and Spiegelman, B.M. (1989) Science **245**, 1234-1236.

Jolly, D.J., Okayama, H., Berg, P., Esty, A.C., Filpula, D., Bohlen, P., Johnson, G.G., Shively, J.E., Hunkapillar, T. and Friedmann, T. (1983) Proc. Natl. Acad. Sci. USA **80**, 477-481.

Jones, N.C., Rigby, P.W.J. and Ziff, E.B. (1988) Genes Dev. **2**, 267-281.

Joyner, A.L., Skarnes, W.C. and Rossant, J. (1989) Nature **338**, 153-156.

Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) Trends Biochem. **11**, 20-23.

Kadonaga, J.T., Carner, K.R., Masiarz, S.R. and Tjian, R. (1987) Cell **51**, 1079-1090.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M. and Beato, M. (1984) Nature **308**, 513-519.

Keith, D.H., Singer-Sam, J. and Riggs, A.D. (1986) Mol. Cell. Biol. **6**, 4122-4125.

Kelley, W.N. and Wyngaarden, J.B. (1979) In: The Matabolic Basis of Inherited Disease, 4th ed. New York, McGraw Hill, 1011-1036.

Kelley, W.N. and Wyngaarden, J.B. (1983) In: The Metabolic Basis of Inherited Disease, 5th ed. New York, McGraw Hill, 1115-1143.

Kim, S.H., Moores, J.C., David, D., Respess, J.G., Jolly, D.J. and Friedmann, T. (1986) Nucl. Acids Res. **10**, 6763-6775.

Konecki, D.S., Brennand, J., Fuscoe, J.C., Caskey, C.T. and Chinault, A.C. (1982) Nucl. Acids Res. **10**, 6763-6775.

Korn, E.D., Remy, C.N. and Buchanan, J.M. (1955) J. Biol. Chem. **217**, 875-885.

Kornberg, A., Lieberman, I. and Sims, E.S. (1955) J. Biol. Chem. **215**, 417-427.

Kratzer, P.G., Chapman, V.M., Lambert, H., Evans, R.E. and Liskay, R.M. (1983) Cell **33**, 37-42.

Kucherlapati, R.S., Eves, E.M., Song, K-Y., Morse, B.S. and Smithies, O. (1984) Proc. Natl. Acad. Sci. USA **81**, 3153-3157.

Kuehn, M.R., Bradley, A., Robertson, E.J. and Evans, M.J. (1987) Nature **326**, 295-298.

Lavia, P., Macleod, D. and Bird, A. (1987) EMBO J.**6**, 2773-2779.

Lee, C.C., Wu, X., Gibbs, R.A., Cook, R.G., Muzny, D.N. and Caskey, C.T. (1988) Science **239**, 1288-1291.

Lee, W., Haslinger, A., Karin, M. and Tjian, R. (1987) Nature **325**, 368-372.

Lehrach, H., Diamond, D., Wosney, J.M. and Boedtker, H. (1977) Biochem. **16**, 4743-4746.

Lesch, M. and Nyhan, W.L. (1964) Am. J. Med. **36**, 561-570.

Lester, S.C., Korn, N.J. and DeMars, R. (1982) Somat. Cell Genet. **8**, 265-284.

Linton, J.P., Yen, J.J., Selby, E., Chen, Z., Chinsky, J.M., Liu, K., Kellems, R.E. and Crouse, G.F. (1989) Mol. Cell. Biol. **9**, 3058-3072.

Liskay, R.M. and Evans, R.J. (1980) Proc. Natl. Acad. Sci. USA **77**, 4895-4898.

Littlefield, J.W. (1964) Science **145**, 709-710.

Lo, Y-F.V. and Palmour, R.M. (1979) Biochem. Genet. **17**, 737-745.

Lock, L.F., Melton, D.W., Caskey, C.T. and Martin, G.R. (1986) Mol. Cell. Biol. **6**, 914-924.

Lock, L.F., Tagaki, N. and Martin, G.R. (1987) Cell **48**, 39-46.

Lyon, M.F. (1972) Biol. Rev. **47**, 1-35.

Malter, J.S. (1989) Science **246**, 664-666.

Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) Science **236**, 1237-1245.

Mansour, S.L., Thomas, K.R. and Capecchi, M.R. (1988) Nature **336**, 348-352.

Martin, G. (1981) Proc. Natl. Acad. Sci. USA **78**, 7634-7638.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. and Bird, A.P. (1989) Cell **58**, 499-507.

Melton, D.W. (1981) Somat. Cell Genet. **7**, 331-344.

Melton, D.W., Konecki, D.S., Ledbetter, D.H., Hejtmancik, J.F. and Caskey, C.T. (1981) Proc. Natl. Acad. Sci. USA **78**, 6977-6980.

Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T. (1984) Proc. Natl. Acad. Sci. USA **81**, 2147-2151.

Melton, D.W., McEwan, C., McKie, A.B. and Reid, A.M. (1986) Cell **44**, 319-328.

Melton, D.W. (1987) In: Oxford Surveys on Eukaryotic Genes (Edited by N. Maclean ) Vol. **4**, 34-76.

Miller, A.D., Jolly, D.J., Friedmann, T. and Verma, I.M. (1983) Proc. Natl. Acad. Sci. USA **80**, 4709-4713.

Miller, A.D., Eckner, R.J., Jolly, D.J., Friedmann, T. and Verma, I.M. (1984) Science **225**, 630-632.

Minty, A.J., Caravatti, M., Robert, B., Cohen, A., Daubas, P., Weydert, A., Gros, F. and Buckingham, M.E. (1981) J. Biol. Chem. **256**, 1008-1014.

Mitchell, P.J., Carothers, A.M., Han, J.H., Harding, J.D., Kas, E., Venolia, L. and Chasin, L.A. (1986) Mol. Cell. Biol. **6**, 425-440.

Mohandas, T., Sparkes, R.S. and Shapiro, L.J. (1981) Science **211**, 393-396.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981) Nucl. Acids Res. **9**, 6047-6068.

Nyhan, W.L., Parkman, R., Page, T., Gruber, H.E., Pyati, J., Jolly, D. and Friedmann, T. (1986) Adv. Exp. Med. Biol. **195**, 167-170.

Ondek, B., Gloss, L. and Herr, W. (1988) Nature **333**, 40-45.

Ornitz, D.M., Palmiter, R.D., Hammer, R.E., Brinster, R.L., Swift, G.H. and MacDonald, R.J. (1985) Nature **313**, 600-602.

Orr-Weaver, T.L., Szostak, J.W. and Rothstein, R. (1981) Proc. Natl. Acad. Sci. USA **78**, 6354-6358.

Osbourne, T.F., Gil, G., Brown, M.S., Kowal, R.C. and Goldstein, J.L. (1987) Proc. Natl. Acad. Sci. USA **84**, 3614-3618.

Palella, T.D., Silverman, L.J., Schroll, C.T., Homa, F.L.,

Levine, M. and Kelley, W.N. (1988) Mol. Cell. Biol. **8**, 457-460.

Palmiter, R.D., Chen, H.Y. and Brinster, R.L. (1982) Cell **29**, 701-710.

Palmiter, R.D. and Brinster, R.L. (1986) Ann. Rev. Genet. **20**, 465-499.

Patel, P.I., Framson, P.E., Caskey, C.T. and Chinault, A.C. (1986) Mol. Cell. Biol. **6**, 393-403.

Pellicer, A., Wigler, M., Axel, R. and Silverstein, S. (1978) Cell **14**, 133-141.

Pinkert, C.A., Ornitz, D.M., Brinster, R.L. and Palmiter, R.D. (1987) Genes Dev. **1**, 268-276.

Pursel, V.G., Pinkert, C.A., Miller, K.F., Bolt, D.J., Campbell, R.G., Palmiter, R.D., Brinster, R.L. and Hammer, R.E. (1989) Science **244**, 1281-1288.

Reynolds, G.A., Basu, S.K., Osbourne, T.S., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) Cell **38**, 275-285.

Rosenbloom, F.M., Kelley, W.N., Miller, J., Henderson, J.F. and Seegmiller, J.E. (1967) JAMA **202**, 103-105.

Ross, J. (1988) Mol. Biol. Med. **5**, 1-14.

Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) Nature **334**, 218-224.

Schwartzberg, P.L., Goff, S.P. and Robertson, E.J. (1989) Science **246**, 799-803.

Sedivy, J.M. and Sharp, P.A. (1989) Proc. Natl. Acad. Sci. USA **86**, 227-231.

Seegmiller, J.E., Rosenbloom, F.M. and Kelley, W.N. (1967) Science **155**, 1682-1684.

Shaw, G. and Kamen, R. (1986) Cell **46**, 659-667.

Singer-Sam, H., Keith, D.H., Tani, K., Simmer, R.L., Shively, L., Lindsay, S., Yashida, A. and Riggs, A.D. (1984) Gene **32**, 409-417.

Solter, D. (1987) Trends Genet. **3**, 23-27.

Southern, E.M. (1975) J. Mol. Biol. **98**, 503-517.

Stout, J.T., Chen, H.Y., Brennand, J., Caskey, C.T. and Brinster, R.L. (1985) Nature **317**, 250-252.

Stout, J.T. and Caskey, C.T. (1989) In: The Metabolic Basis of Inherited Disease, 5th ed. New York, McGraw Hill, 1007-1028.

Swain, J.L., Stewart, T.A. and Leder, P. (1987) Cell **50**, 719-727.

Swanson, L.W., Simmons, D.M., Arriza, J., Hammer, R., Brinster, R.L., Rosenfeld, M.G. and Evans, R.M. (1985) Nature **317**, 363-366.

Thomas, K.R. and Capecchi, M.R. (1987) Cell **51**, 503-512.

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L. and Melton, D.W. (1989) Cell **56**, 313-321.

Toniolo, D., D'Urso, M., Martini, G., Persico, M., Tufano, V., Battistuzzi, G. and Luzzatto, L. (1984) EMBO J.**3**, 1987-1995.

Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkvens, T.M., van der Voorn, L., van Ormondt, H. and van der Eb, A.J. (1985) EMBO J.**4**, 437-443.

van Dyke, M.W., Roeder, R.G. and Sawadogo, M. (1988) Science **241**, 1335-1338.

Venolia, L. and Gartler, S.M. (1983) Nature **302**, 82-83.

Vieira, J. and Messing, J. (1982) Gene **19**, 259-268.

Wahl, G.M., Hughes, S.H. and Capecchi, M.R. (1975) J. Cell Physiol. **85**, 307-320.

Wall, L., de Boer, E. and Grosveld, F. (1988) Genes Dev. **2**, 1089-1100.

Wilson, J.M., Tarr, G.E., Mahoney, W.C. and Kelley, W.N. (1982) J. Biol. Chem. **257**, 10978-10985.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T. and Migeon, B.R. (1984) Proc. Natl. Acad. Sci. USA **81**, 2806-2810.

Wolf, S.F., Dintzis, S., Toniolo, D., Persico, G., Lunnen, K.D., Axelman, J. and Migeon, B.R. (1984a) Nucl. Acids Res. **12**, 9333-9348.

Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. and Jaenisch, R. (1989) Nature **342**, 435-438.

Zimmer, A. and Gruss, P. (1989) Nature **338**, 150-153.

## Appendix

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L. and Melton, D.W. (1989) Cell **56**, 313-321.

Cell, Vol. 56, 313–321, January 27, 1989, Copyright © 1989 by Cell Press

# Germ Line Transmission and Expression of a Corrected HPRT Gene Produced by Gene Targeting in Embryonic Stem Cells

**Simon Thompson,* Alan R. Clarke,† Angela M. Pow,* Martin L. Hooper,† and David W. Melton***
* Department of Molecular Biology
University of Edinburgh
Mayfield Road
Edinburgh EH9 3JR
Scotland
† Department of Pathology
University of Edinburgh
Teviot Place
Edinburgh EH8 9AG
Scotland

## Summary

**The deletion mutation in the HPRT-deficient mouse embryonic stem (ES) cell line E14TG2a has been corrected by gene targeting. The presence of plasmid sequences in the correcting vector DNA did not affect the frequency of correction. We have characterized three different HPRT gene structures in correctants. Cells from one corrected clone have been introduced into mouse blastocysts, and germ line transmission of the ES cell-derived corrected gene has been achieved. The corrected gene has the same pattern of expression as the wild-type gene, with the characteristic elevated level of expression in brain tissue. Hence, we have demonstrated the feasibility of introducing targeted modifications into the mouse germ line by homologous recombination in ES cells.**

## Introduction

A strategy enabling precise modifications to be made to the mammalian genome would be of benefit both to biological and medical research. It would become possible to manipulate the expression of genes by targeting changes to their control sequences. This would be of value to the study of gene expression. It could also be of potential commercial value if used, for example, in livestock animals to increase output, or produce novel materials. In addition, genes could be inactivated to create animal models for human genetic diseases or to study the action of developmental genes, while the ability to correct mutant genes has implications for gene therapy. An advantage of being able to target modifications is that genes are manipulated in their natural chromosomal environment, whereas the use of conventional methods for introducing DNA sequences into the germ line (Jaenisch, 1988) allows no control over the chromosomal site of integration or the number of copies introduced. At the very least, this complicates the interpretation of gene expression studies and may result in insertion, either into sites that are inappropriate for expression or into essential genes, with deleterious consequences.

Several recent advances have made gene targeting in the mammalian genome possible. Here we report the generation of mice that show germ line transmission of a targeted gene modification. This has been achieved using homologous recombination in mouse embryonic stem (ES) cells.

Mouse ES cells can be cultured in vitro and still retain their pluripotency (Evans and Kaufman, 1981; Martin, 1981). These cells, when reintroduced into mouse blastocysts, can contribute to the germ line of the resultant chimeras (Bradley et al., 1984). ES cells deficient for the enzyme hypoxanthine–guanine phosphoribosyltransferase (HPRT; IMP:pyrophosphate phosphoribosyltransferase; EC 2.4.2.8) have been used to produce mice that transmit the same enzyme deficiency through their germ line (Hooper et al., 1987; Kuehn et al., 1987). The pathology of this deficiency is interesting because in humans the result is the severe neurological disorder Lesch–Nyhan syndrome (Lesch and Nyhan, 1964; reviewed in Stout and Caskey, 1988). Consequently, it is noteworthy that, in normal individuals, HPRT expression is elevated in brain relative to other tissues. Hooper et al. (1987) selected for spontaneous HPRT-deficient ES cells in culture. Kuehn et al. (1987) used multiple retroviral infection of cultured ES cells to promote insertional mutation, and subsequently selected for HPRT-deficient clones. Neither procedure abolished the ability of the mutant ES cells to contribute to germ line chimerism.

Gene targeting, mediated by recombination between introduced vectors and homologous chromosomal sequences, has been employed in ES cells both to inactivate gene function (Thomas and Capecchi, 1987) and to correct the activity of a mutant gene (Doetschman et al., 1987). In both cases the target gene was HPRT, which has the advantage that it is possible to select either for, or against, activity while the ES cells are in culture. Additional advantages of the system are that the gene is X-linked and is therefore hemizygous in male cells, and the structure of the gene is well characterized (Melton et al., 1984). In their gene inactivation experiments, Thomas and Capecchi (1987) used vectors containing a modified neomycin resistance ($neo^R$) gene inserted into the eighth exon of the HPRT gene. The $neo^R$ gene has the dual function of disrupting the HPRT coding sequence and providing an additional selectable marker. Homologous recombination between such a vector and the chromosomal HPRT gene results in the inactivation of the chromosomal gene. Thomas and Capecchi (1987) described targeting vectors as either sequence replacement or sequence insertion vectors, depending on their mechanism of integration, and found that both types worked with similar efficiencies.

We have previously described the targeted correction of an HPRT-deficient ES cell line, E14TG2a, using homologous recombination (Doetschman et al., 1987). E14TG2a was used by Hooper et al. (1987) to produce HPRT-deficient mice. The characterization of the mutation in E14TG2a is presented here. We describe the use of sequence insertion vectors to correct the HPRT deficiency

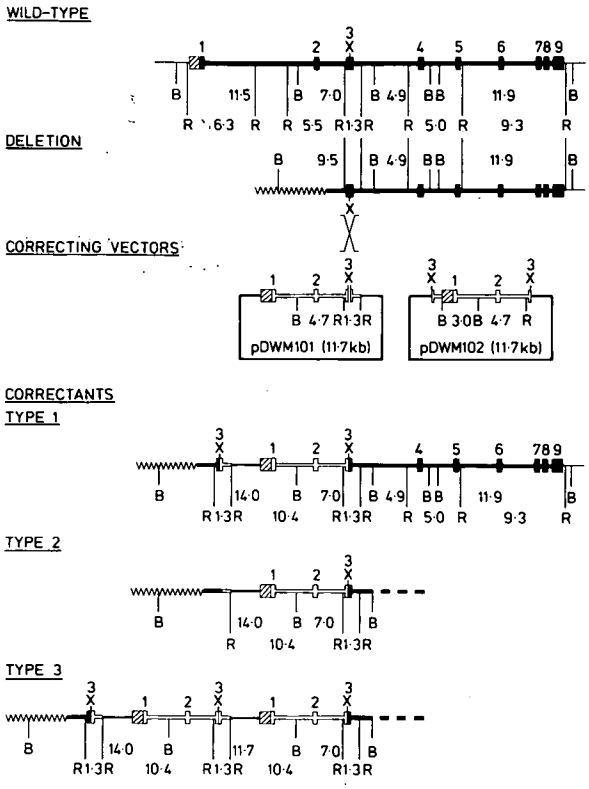



Figure 1. Targeted Correction of the Deletion in E14TG2a Cells and Correctant Structures

The structures of the wild-type HPRT gene, the E14TG2a deletion, the two correcting vectors, and the three correctant types are shown schematically. Closed boxes, endogenous exons; thick closed lines, endogenous introns; cross-hatched boxes, promoter regions; thin lines, HPRT flanking sequences; jagged line, distant flanking sequence brought into proximity as a result of the deletion in E14TG2a; open boxes, vector-derived exons; open thick lines, vector-derived introns; closed intermediate thickness lines, plasmid sequence. The number of each exon is shown directly above it. Selected restriction sites are shown: R, EcoRI; B, BamHI; X, XhoI. The sizes (in kb) of all EcoRI and BamHI restriction fragments containing exon elements are shown between the restriction sites. The thick closed broken line in the type 2 and 3 correctants denotes that beyond exon 3, they have the same structure as type 1 correctants. With the exception of the plasmid in the correcting vectors and the exon elements themselves, the same scale is used throughout.

in E14TG2a, and the identification of three types of corrected clone. Cells from one corrected clone have been injected into mouse blastocysts, and have successfully contributed to the germ line of a chimeric mouse. We have crossed this germ line chimera with HPRT-deficient mice in order to analyze expression from the corrected gene. The corrected gene shows levels of expression equivalent to the normal gene, with the same characteristic elevation in brain.

## Results

### Characterization of the E14TG2a Mutation

The structure of the wild-type mouse HPRT gene is shown schematically in Figure 1. The coding sequence of the gene is divided into nine exons that span approximately 33 kb of the X chromosome (Melton et al., 1984). We have used Southern blot hybridization to characterize the mutation in the male ES cell line E14TG2a, which is deficient for HPRT enzyme activity. The restriction enzyme sites used in this analysis are shown in Figure 1. Southern hybridization of DNA from the wild-type ES cell line E14 and from E14TG2a, using a full-length mouse HPRT cDNA probe, is shown in Figure 2A. The sizes of the bands are indicated, and the exon elements they contain can be seen in Figure 1. In the EcoRI-digested E14TG2a lane the 6.3 kb band, which contains exon 1, and the 5.5 kb band, which contains exon 2, are both absent. In the BamHI digest of E14TG2a the doublet band, consisting of the 11.5 kb fragment, which contains exon 1, and the 11.9 kb fragment, which contains exons 5 to 9, is reduced in intensity relative to the wild-type band. The 7.0 kb BamHI fragment, which contains exons 2 and 3 of the wild-type gene, is replaced by a 9.5 kb band containing exon 3. From these results we infer that the mutation in E14TG2a is a deletion spanning exons 1 and 2 of the HPRT gene. Using a PvuII digest (data not shown) we have mapped the 3′ end of the deletion to between 1.0 kb and 2.9 kb upstream of exon 3. We have used probes from the region flanking the HPRT gene to show that the 5′ end of the deletion is at least 10 kb upstream of the gene (data not shown).

Since the wild-type allele present in strain 129 mice is *Hprt*$^{b}$, in accordance with the rules of the International Committee for Standardized Genetic Nomenclature for Mice (Mouse News Letter 72, 2–27, 1985) we designate the mutant allele present in E14TG2a as *hprt*$^{b\text{-}m3}$.

### Correction of the E14TG2a Mutation by Gene Targeting

To correct the mutation in E14TG2a we constructed the sequence insertion vector pDWM101 (see Figure 1 and Experimental Procedures for details of construction). pDWM101 contains mouse exons 1, 2, and 3 and the mouse HPRT promoter. The exon 3 region provides homology with the target locus. The extent of homology between the targeting vector and the genome is between 2.3 kb and 4.2 kb, depending on the precise 3′ endpoint of the deletion. pDWM101 contains 650 bp of upstream flanking sequence, which is sufficient for HPRT expression in cultured cells (Melton et al., 1986). The anticipated outcome of correction has the same structure as the wild-type gene, except that the first intron is reduced from 10.8 kb to 4.1 kb (see Figure 1). The targeted locus has a restriction map distinct from both the wild-type gene and the deletion mutant, and thus enables correctants to be identified by Southern analysis. We have called the predicted structure type one (see Figure 1). In the EcoRI restriction map a novel 10.4 kb fragment is generated, containing exons 1 and 2, and the 1.3 kb band, containing exon 3, is duplicated. In the BamHI restriction map a 14.0 kb fragment, containing exon 1 and the duplicated exon 3, is generated and the 9.5 kb band, containing exon 3, is replaced by the wild-type 7.0 kb band which contains exons 2 and 3. The 7.0 kb and the 14.0 kb bands are the products of one BamHI site in the vector and one BamHI site in the ge-

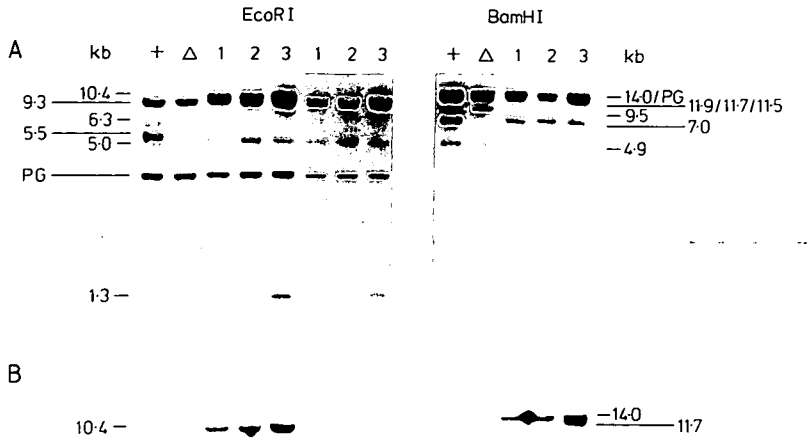



Figure 2. Southern Analysis of E14TG2a and Correctant DNA

Genomic DNA (10 μg) was restricted with EcoRI or BamHI, electrophoresed, transferred, and hybridized. Probes: In (A), a full-length HPRT cDNA probe was used. In (B), the same transfer was reprobed with pUC8. DNA: +, E14; Δ, E14TG2a; 1, a type 1 correctant; 2, type 2-1; 3, type 3-1. Both a long and a short exposure of EcoRI-digested correctant DNA are included in (A). The sizes (in kb) of hybridizing bands are shown adjacent to each panel. The fragments containing the HPRT pseudogene sequences are indicated by PG. The exon elements present in each band can be determined by consulting Figure 1.

nome. While the 7.0 kb band is also present in the wild-type gene, the 14.0 kb band is diagnostic for targeted correctants.

To promote recombination, the vector was linearized at the unique XhoI site in exon 3 prior to being introduced into E14TG2a cells by electroporation. The conditions chosen were those that resulted in minimal cell death in control plating experiments. Correctants were selected in HAT medium. Cells only become $HAT^R$ if the correct targeting event takes place between pDWM101 and the defective HPRT gene. As a positive control E14TG2a cells were electroporated with pDWM100, a fully functional HPRT minigene derivative of pDWM101, which can rescue HPRT-deficient cells without the requirement for homologous recombination.

In the first experiment, which used late-passage stem cells, eight $HAT^R$ colonies were generated from $4.6 \times 10^7$ cells electroporated (see Table 1). The ratio of $HAT^R$ colonies obtained from cells electroporated with pDWM101 to $HAT^R$ colonies obtained from cells transformed with pDWM100 was 0.007 (see Table 1, targeting index). This frequency gives an indication of the number of targeting events relative to the number of random integrations, although it does not take into account integrations of pDWM100 into sites that inhibit expression. Southern hybridization showed that seven of the eight $HAT^R$ clones had the structure predicted for type 1 correctants (see Figure 2). In the EcoRI digest the new 10.4 kb band is visible, and the intensity of the 1.3 kb band is doubled with respect to the equivalent band in the E14TG2a lane. The intensity of the pseudogene band (Isamat et al., 1988) in the EcoRI digests allows comparison of the relative loading of each lane, thus acting as an internal control. In the BamHI digest the disappearance of the 9.5 kb fragment, characteristic of the deletion, and the regeneration of the 7.0 kb wild-type band can be seen. The presence of the 14.0 kb fragment, which comigrates with the pseudogene fragment, is indicated by the increased intensity of the band at that molecular weight.

One $HAT^R$ clone had the structure we have called type 3 (see Figure 1). This clone has two correcting vectors integrated in tandem. The effect on the restriction map is that, in an EcoRI digest, the 10.4 kb fragment is duplicated and there are three copies of the 1.3 kb fragment compared with two copies in type 1 correctants and one copy in the deletion mutant. This is evident in the relative intensities of these bands in Figure 2A. In a BamHI digest the 14.0 kb and 7.0 kb fragments, seen in type 1 correctants, are present but there is an additional 11.7 kb, vector-sized fragment. This can be seen more clearly in Figure 2B, which shows a pUC8 reprobe of the transfer in Figure 2A. The pUC8 sequence is present within the 10.4 kb EcoRI band, and consequently the intensity of this band is doubled in type 3 DNA relative to type 1. In a BamHI digest

Table 1. Summary of Correction Experiments

| Experiment | Correcting Vector | Number of $HAT^R$ Colonies | Frequency of HPRT Correctants[a] | Positive Control[b] | Targeting Index[c] | Correctant Types 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|
| 1 | pDWM101 | 8 | $1.75 \times 10^{-7}$ | $2.5 \times 10^{-5}$ | 0.007 | 7 | – | 1 |
| 2 | pDWM101 | 2 | $4.0 \times 10^{-8}$ | $6.4 \times 10^{-6}$ | 0.006 | 1 | 1 | – |
| 3 | pDWM101 | 12 | $2.2 \times 10^{-7}$ | $6.3 \times 10^{-6}$ | 0.035 | 12 | – | – |
| | pDWM102 | 8 | $1.5 \times 10^{-7}$ | $6.3 \times 10^{-6}$ | 0.024 | 8 | – | – |

[a] Number of $HAT^R$ colonies generated per cell electroporated with correcting vector DNA.
[b] Number of $HAT^R$ colonies generated per cell electroporated with pDWM100 DNA.
[c] Frequency of $HAT^R$ colonies generated per cell electroporated with correcting vector DNA relative to $HAT^R$ colonies generated per cell electroporated with pDWM100 DNA (a/b).

the pUC8 fragment is present in both the 14.0 kb band and the 11.7 kb band. This is diagnostic for the type 3 structure. The equal intensities of the 14.0 kb and 11.7 kb bands in the type 3 lane show that there are two copies of the vector integrated. If there were more than two copies the intensity of the 11.7 kb band would be at least twice that of the 14.0 kb band. There was no evidence either from these correctants, or from any of the others we have examined subsequently, for any random integrations of the correcting vector, which would have been detected as additional bands hybridizing to the pUC8 probe.

To generate a corrected ES cell line suitable for reintroduction into mouse blastocysts, a second experiment was performed using low-passage E14TG2a cells. pDWM101 was used as the targeting vector; electroporation of $5 \times 10^7$ cells produced two $HAT^R$ clones (see Table 1). Southern analysis was used to determine the structure of the HPRT gene in these correctants. One clone had a structure indistinguishable from the seven type 1 correctants generated in the first experiment, and so was called type 1-8. The second $HAT^R$ clone had a subtly different pattern, which we have called type 2 (see Figures 1 and 2). In this clone, the exon 3–containing 1.3 kb EcoRI band has the same intensity as in the E14TG2a lane, indicating that exon 3 is not duplicated. In addition, the intensity of the 14.0 kb BamHI band is lower than in type 1 correctants, although the size of the band is not visibly different at this resolution. An XhoI digest for this clone shows that the XhoI site in the upstream copy of exon 3 is absent (data not shown). The presence of the 10.4 kb EcoRI band indicates that the EcoRI site immediately upstream of the plasmid sequence is present. From this we infer that this clone, correctant type 2-1, has the targeting vector inserted correctly, but a small deletion has occurred to remove the upstream copy of exon 3. Therefore in this correctant, unlike type 1 correctants, there is no exon 3 duplication. This does not affect the HPRT-positive status of the clone because the deleted exon is not part of the functional gene.

### Frequency of 6-Thioguanine Resistant Cells in Correctant Populations

To assay the frequency of $HPRT^-$ cells in the two correctant populations (type 1-8 and type 2-1) generated in the second experiment, cells were grown in the absence of selection for 2 weeks, and then their plating efficiency in medium containing 6-thioguanine (6-TG), relative to their plating efficiency in nonselective medium, was measured. The relative plating efficiency for type 2-1 was 0.03%, which is approximately the same as the figure obtained for wild-type ES cells. However, the relative plating efficiency for type 1-8 was 21.0%; that is, after 2 weeks of growth in nonselective medium, 21.0% of the cell population in the type 1-8 culture was HPRT deficient.

To analyze further the presence of $6\text{-}TG^R$ cells in the type 1-8 population, we used Southern hybridization to examine DNA from cells grown under different selection conditions. The structure of the gene in DNA prepared from the type 1-8 culture grown nonselectively for 24 days was indistinguishable from the E14TG2a deletion structure. Thus, when selection is removed from the type 1-8 culture, cells containing the corrected gene are rapidly replaced by $6\text{-}TG^R$ cells which are $6\text{-}TG^R$ because their HPRT gene has the E14TG2a structure. This could be due to the persistence in the $HAT^R$ culture of E14TG2a cells, surviving, since the original isolation of the correctant clone, by metabolic cooperation (Hooper, 1982). Alternatively, $6\text{-}TG^R$ cells might arise by loss of vector sequences from corrected cells, due to a reversal of the original homologous recombination (correcting) event. These possibilities are under investigation. Whatever the reason, it was deemed inappropriate to use the type 1-8 culture as a source of cells for blastocyst injections. Hence, the correctant type 2-1 was chosen for introduction into mouse blastocysts (see below).

### The Effect of Plasmid Sequences on Targeting Efficiency

The use of the targeting vector pDWM101 results in the integration of plasmid sequences into the genome. To determine whether the incorporation of plasmid sequences was detrimental to the efficiency of gene targeting, a third experiment was performed (see Table 1). In this experiment E14TG2a cells were electroporated with either pDWM101 or pDWM102 DNA (see Figure 1 and Experimental Procedures for details of construction). pDWM102 contains precisely the same HPRT gene-derived sequences as pDWM101, but they are rearranged so that when pDWM102 DNA is restricted with XhoI the pUC8 sequences are excised. Consequently, correctants generated with this vector do not have pUC8 sequences integrated into the genome. The predicted, type 1, restriction map for pDWM102 correctants differs from that for pDWM101 correctants in the following ways: In the EcoRI restriction map the 10.4 kb and the duplicated 1.3 kb fragments seen in pDWM101 correctants are replaced by a fragment of 9.0 kb, which is equal to their combined size minus the size of pUC8 (2.7 kb), due to the absence of the EcoRI site immediately upstream of the promoter (see Figure 1). In a BamHI digest the loss of pUC8 and the gain of a BamHI site upstream of the promoter results in the replacement of the 14.0 kb fragment by a 3.0 kb internal fragment and an 8.3 kb fragment extending into the upstream flanking sequence. A comparison of the Southern hybridization patterns for BamHI-digested DNA from type 1 correctants, generated with pDWM101 and pDWM102, is shown in Figure 3.

In the third experiment, 12 $HAT^R$ colonies were generated from cells electroporated with pDWM101 and eight correctants were generated from cells electroporated with pDWM102 (see Table 1). The frequency of gene targeting relative to random integrations, calculated as for previous experiments, was 0.035 for pDWM101 and 0.024 for pDWM102. Thus we conclude that the presence of plasmid sequences in the targeting vector does not affect the frequency of gene targeting. All correctants generated in this experiment had the type 1 gene structure.

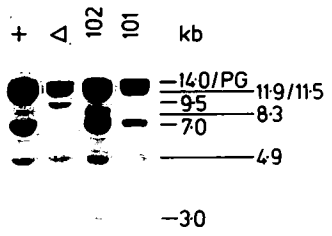



Figure 3. Comparison of Type 1 Correctants Produced by Correcting Vectors pDWM101 and pDWM102

Genomic DNA (10 μg) was restricted with BamHI and probed with an HPRT cDNA probe. DNA: +, E14; △, E14TG2a; 102, a pDWM102 type 1 correctant; 101, a pDWM101 type 1 correctant. The sizes (in kb) of the hybridizing bands are shown. The pattern is explained in the text.

## Production of Chimeric Mice

Cells from the correctant clone type 2-1 were injected into F2 (C57BL/6/Ola × CBA/Ca/Ola) blastocysts. The latter are homozygous for the wild-type alleles at the *c* and *p* loci (*C/C, P/P*) and segregating for the alleles *A* (agouti) and *a* (nonagouti), whereas the 129/Ola-derived type 2-1 cells are homozygous for the alleles $c^{ch}$ (chinchilla) and *p* (pink-eyed dilution), which both lighten the coat color, and for $A^w$ (white-bellied agouti). Hence, chimeric mice are identifiable by the presence of light coloring in an otherwise dark coat. Twenty-six mice were born from 93 blastocysts injected. Fifteen chimeric mice were obtained, 12 of which were male. Male chimeras were tested for germ line transmission of the $c^{ch}$ and *p* markers, derived from ES cells, by crossing with strain 129/Ola females. Offspring which, like strain 129/Ola, are light yellow with pink eyes are diagnostic of germ line transmission (see Table 2). One out of eight males tested showed germ line transmission to 12 of 16 offspring (see Figure 4). Although type 2-1 cells have a modal chromosome number of 40 and those metaphase spreads with 40 chromosomes have no detectable abnormalities in G-banding, only eight out of 20 metaphases examined had the modal number. Nevertheless, when the male germ line chimera was mated to an F1 (C57BL/6/Ola × CBA/Ca/Ola) female, the resulting 11 fertilized eggs, studied at the one-cell stage, all had 40 chromosomes (data not shown).

## Germ Line Transmission and Expression of the Corrected Gene

To study the expression of the corrected gene in mice, the male chimera showing germ line transmission of cell line–derived markers was crossed to homozygous HPRT deficient ($hprt^{b\text{-}m3}/hprt^{b\text{-}m3}$) females, bred from the series of chimeras produced with E14TG2a (see Experimental Procedures). The gene pool of these females is derived entirely from strain 129/Ola and from the outbred strain MF1. The latter is a strain of mice whose precise genetic status has not been characterized, but coat colors observed in the progeny of crosses are consistent with their being homozygous for the albino allele (*c/c*) and for the nonagouti allele (*a/a*). White homozygous $hprt^{b\text{-}m3}/hprt^{b\text{-}m3}$ females were chosen for mating to chimeras; if the genotype deduced above for MF1 is correct, these homozygotes must either be of genotype *c/c* or carry the gene combination $c^{ch}/c^{ch}$, *p/p*, *a/a* (Silvers, 1979). In both cases, host blastocyst–derived offspring are agouti or black, while cell line–derived offspring are lighter than agouti (Table 2). The cell line–derived marker was transmitted to nine of 12 progeny (see Figure 4). Cell line–derived female offspring inherit one X chromosome carrying the corrected gene, which we designate $Hprt^{b\text{-}m4}$, derived from their father, and the other X chromosome, carrying deletion $hprt^{b\text{-}m3}$, from their mother. Two such females were screened for the corrected gene by Southern hybridization of tail DNA (see Figure 5).

EcoRI digests of DNA from these two mice show the 10.4 kb band diagnostic for the corrected gene. The relative intensity of this band is reduced in this hybridization, with respect to the corrected cell line DNA, because the ES cell line is male and therefore contains just one corrected HPRT allele per cell, whereas in the female mice only one of the two alleles is corrected. The absence of a wild-type HPRT allele in these mice is indicated by the lack of the 5.5 kb and 6.3 kb bands. DNA from a heterozygous $hprt^{b\text{-}m3}/Hprt^{b}$ mouse is shown in Figure 5, for comparison. In BamHI-digested DNA from mice carrying the corrected gene, heterozygosity is indicated by the presence of both the 9.5 kb band, containing exon 3 from the deletion, and the 7.0 kb band, containing exons 2 and 3 from the corrected allele. These two bands are also present in DNA from heterozygous $hprt^{b\text{-}m3}/Hprt^{b}$ mice, but

Table 2. Expected Phenotypes of Progeny Mice from Test Crosses of Chimeras

| Host Blastocyst F2 (C57BL/6/Ola × CBA/Ca/Ola) | Stem Cells (Strain 129/Ola) | Tester Female[a] | Coat Color | |
|---|---|---|---|---|
| | | | Host Progeny | Stem Cell Progeny |
| *C/C* | $c^{ch}/c^{ch}$ | (1) $c^{ch}/c^{ch}$ | agouti | light yellow[b] |
| *P/P* | *p/p* | *p/p* | | |
| | $A^w/A^w$ | $A^w/A^w$ | | |
| | | (2) *c/c* | agouti or black | light chinchilla |
| | | (3) $c^{ch}/c^{ch}$ | agouti or black | light yellow[b] |
| | | *p/p* | | |
| | | *a/a* | | |

[a] Tester females: (1), strain 129/Ola; (2) and (3), differing possible genotypes of white homozygous $hprt^{b\text{-}m3}/hprt^{b\text{-}m3}$ females (see text).
[b] These animals are predicted to have pink eyes, the remaining four classes dark eyes.

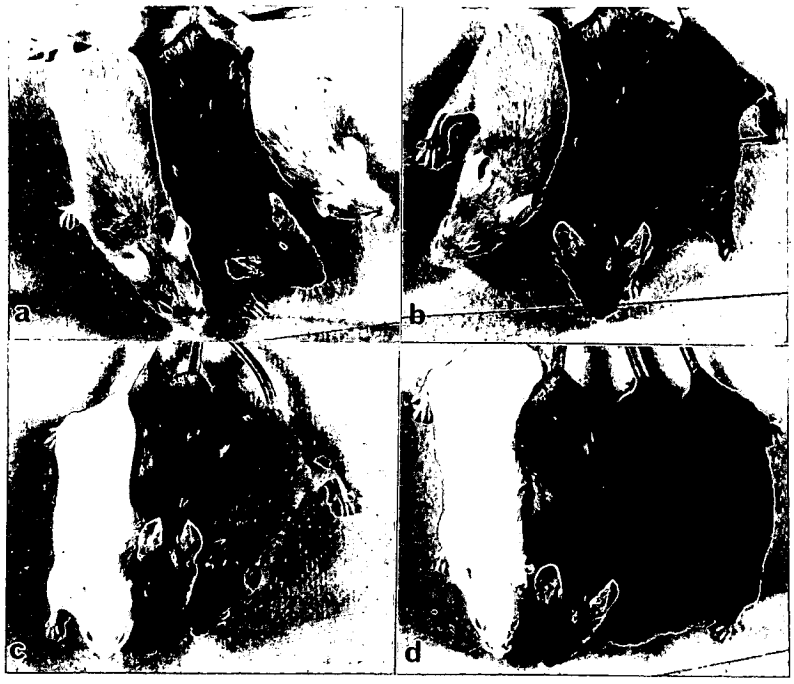



Figure 4. Germ Line Transmission of Coat Color Markers from Type 2-1 Chimeras

Mice are listed from left to right in each case. (a) 129/Ola female; male type 2-1 germ line chimera; cell line–derived light yellow offspring. (b) First two animals as in (a); host blastocyst–derived agouti offspring. (c) White *hprt*$^{b\text{-}m3}$/*hprt*$^{b\text{-}m3}$ female; male type 2-1 germ line chimera; cell line–derived light chinchilla offspring. (d) First two animals as in (c); host blastocyst–derived black offspring; host blastocyst–derived agouti offspring.

the *hprt*$^{b\text{-}m3}$/*Hprt*$^{b\text{-}m4}$ mice are distinguishable by the EcoRI digests.

A Northern analysis of HPRT mRNA from stem cells and mouse tissues is shown in Figure 6. The level of the 1500 nucleotide HPRT mRNA in the corrected cells is indistinguishable from the level in wild-type E14 cells. No HPRT transcripts are detected in RNA from E14TG2a cells. In normal mice the level of HPRT mRNA in brain tissue is elevated approximately 5-fold relative to liver tissue (see Figure 6). Because HPRT is X-linked, some cells in heterozygous mice will be functionally HPRT deficient, because of random inactivation of one X chromosome in each cell. Hence to provide a positive control for HPRT expression in the two heterozygotes containing one deleted and one corrected allele, RNA from two heterozygous *hprt*$^{b\text{-}m3}$/*Hprt*$^{b}$ mice was included. The pattern of expression in mice heterozygous for the corrected allele is essentially the same as in mice heterozygous for the wild-type allele. All heterozygotes showed the characteristic low level of HPRT mRNA in liver and elevated levels in brain. Differences in the degree of mosaicism between HPRT-expressing cells and cells in which the wild-type or corrected HPRT gene is on the inactive X chromosome probably account for the slight variations observed in the absolute levels of expression between mice. Thus we conclude that the corrected gene functions in the same way

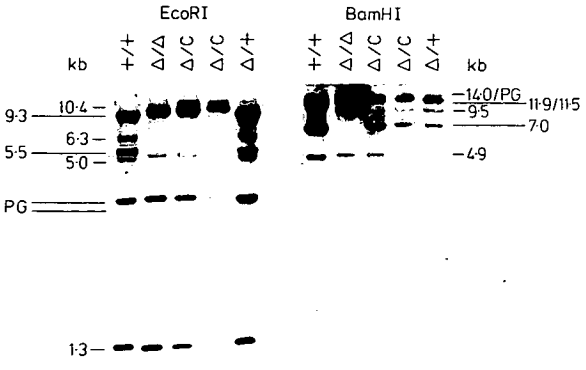



Figure 5. Southern Analysis to Show Germ Line Transmission of the Corrected HPRT Gene

Tail DNA prepared from female mice was digested with either EcoRI (left) or BamHI (right) and probed with the full-length HPRT cDNA. Lanes: +/+, homozygous *Hprt*$^{b}$/*Hprt*$^{b}$; Δ/Δ, homozygous *hprt*$^{b\text{-}m3}$/*hprt*$^{b\text{-}m3}$; Δ/C, heterozygous *hprt*$^{b\text{-}m3}$/*Hprt*$^{b\text{-}m4}$; Δ/+, heterozygous *hprt*$^{b\text{-}m3}$/*Hprt*$^{b}$. The size (in kb) of each hybridizing band is shown. The fragments containing the HPRT pseudogene are labeled PG. One of the heterozygous *hprt*$^{b\text{-}m3}$/*Hprt*$^{b\text{-}m4}$ mice shows a previously characterized pseudogene polymorphism (Isamat et al., 1988).


Figure 6. Northern Analysis Showing Expression from the Corrected Gene in ES Cells and Mouse Tissues

Total RNA (30 μg) was prepared from ES cells or mouse tissues, electrophoresed on formaldehyde–agarose gels, transferred, and probed with full-length HPRT cDNA. Cell RNA: +, E14; Δ, E14TG2a; C, correctant type 2-1 grown nonselectively. Tissue RNA: Liver (L) and brain (B) RNAs are shown from two heterozygous *hprt*$^{b\text{-}m3}$/*Hprt*$^{b}$ mice, Δ/+; two heterozygous *hprt*$^{b\text{-}m3}$/*Hprt*$^{b\text{-}m4}$ mice, Δ/C; and one homozygous *Hprt*$^{b}$/*Hprt*$^{b}$ mouse, +/+. The size of the HPRT mRNA is approximately 1500 nucleotides.

as the wild-type gene in mice. That is, the function of the mutant gene in E14TG2a has been fully corrected by gene targeting.

## Discussion

In this study we have demonstrated the feasibility of making targeted modifications to the mouse genome by homologous recombination in ES cells. We have used targeting vectors of the type defined as insertion vectors (Thomas and Capecchi, 1987) to correct an HPRT deficiency in the ES cell line E14TG2a, which has a deletion of the first two exons of the HPRT structural gene. Insertion of the targeting vector pDWM101 into the genome by homologous recombination corrects the deletion mutation. Three electroporation experiments generated a total of 30 $HAT^R$ colonies. To give an indication of the number of gene targeting events relative to random integrations, we have calculated the ratio of the number of $HAT^R$ colonies per cell electroporated with pDWM101 to the number of $HAT^R$ colonies per cell electroporated with pDWM100, a fully functional minigene derivative of pDWM101, and denoted this the targeting index (see Table 1). All electroporations were carried out under identical conditions. The 4-fold variation in targeting index observed between experiments could not be correlated with the state of the stem cell cultures at the time of electroporation. The highest targeting index we observed was 0.035, which is more than 10-fold higher than the frequency reported by Thomas and Capecchi (1987). Neither calculation allows for random integrations into regions of the genome that prevent expression of the introduced gene. However, a relatively high frequency of targeting events is implied by our observations, and those of Thomas and Capecchi (1987) and Doetschman et al. (1987), that no random integrations, in addition to the targeting event, have been found in any homologous recombinants analyzed so far. This suggests that it should prove possible to make targeted genome alterations without the risk of deleterious consequences arising from random integrations.

Southern analysis of the HPRT gene in correctants shows that the majority (28/30) have the predicted (type 1) gene structure. However, we have identified two correctants that have different gene structures. The type 2 structure has a deletion of the upstream copy of the exon 3 region, which may have occurred during, or after, the integration event. This could be equivalent to the deletions reported by Doetschman et al. (1988) during their HPRT inactivation experiments. The type 3 structure has two copies of the correcting vector inserted in tandem. We have shown that there is no significant difference in the frequency of correction between a vector that does result in the integration of plasmid sequences, pDWM101, and one that does not, pDWM102. This is helpful, because generally it is easier to construct insertion vectors that do result in the integration of plasmid sequence. In both targeting vectors, pDWM101 and pDWM102, the extent of homology with the target locus is the same (between 2.3 kb and 4.2 kb). Others have shown that the greater the extent of homology, the higher the efficiency of gene targeting (Thomas and Capecchi, 1987).

Cells from the correctant clone type 2–1, in which the corrected gene is essentially as stable as a wild-type gene, were introduced into mouse blastocysts. The frequency of chimeras per animals born was comparable to the value reported previously for the parental (E14TG2a) stem cells (Hooper et al., 1987). We believe that the lower frequency of blastocysts surviving to term in the experiments described here reflects difficulties associated with establishing the technique in our hands, rather than a reduction in the ability of targeted stem cells to contribute to the developing embryo. One chimeric male, from eight tested, was shown to transmit the corrected gene through the germ line. This is a lower proportion than for E14TG2a-derived male chimeras, of which 19 out of 34 showed germ line transmission (Hooper et al., 1987). The reduced ratio may be a consequence of the lower fraction of diploid cells in type 2–1. In total, the chimera described here transmitted corrected cell line–derived markers to 21 of 28 offspring. Chimeras produced with E14TG2a have shown transmission of stem cell–derived markers to 100% of offspring and to as few as 1% of offspring (Hooper et al., 1987). The ES cells used here for blastocyst injections were subjected to the mildest electroporation conditions possible, and their time in culture was kept to a minimum. This may have contributed to the successful production of germ line chimerism from the corrected stem cells.

By crossing to HPRT-deficient females, we have produced female mice in which the corrected gene is the only functional HPRT gene. This enabled us to study expression of the corrected gene in the absence of a wild-type gene. We chose to analyze expression in these females because males with a corrected gene would not be available until the next generation. Mosaicism in the female mice, caused by random X inactivation, did not complicate the interpretation of the expression data. HPRT is a member of the housekeeping class of genes, which are generally expressed constitutively at low levels in all cell types. However, a characteristic of HPRT expression is the elevation of mRNA levels in brain relative to liver and other tissues. The deletion in E14TG2a extends at least 10 kb upstream of the HPRT gene. Correction with pDWM101 replaces just 650 bp of 5′ flanking sequence, which we show is sufficient to control HPRT expression with the characteristic liver and brain mRNA levels. Hence, the mechanism by which expression is elevated in brain can have no requirement for sequences located more than 650 bp upstream of the transcription initiation point. In cultured cells 50 bp of 5′ sequence is sufficient for a wild-type level of expression from an HPRT minigene (Melton et al., 1986). Although the minigene experiments were subject to random integration and variable copy number, the corrected gene in mice is present in only one copy per cell, in its natural chromosomal environment. By targeting changes both to the promoter and the 3′ untranslated region of HPRT, we hope to determine the mechanism by which expression is elevated in brain tissue.

The brain-specific elevation of HPRT expression is particularly noteworthy because a deficiency of HPRT in hu-

mans is the molecular basis of Lesch–Nyhan syndrome, which is characterized by learning difficulties, spasticity, choreoathetosis, and compulsive self-injurious behavior. Mice deficient for HPRT do not demonstrate the behavioral patterns characteristic of Lesch–Nyhan syndrome. This suggests that mice, unlike humans, can tolerate the HPRT deficiency, perhaps by utilizing other pathways to maintain the levels of purines and neurotransmitters in the brain. However, the levels of striatal dopamine are reduced by approximately 20% in HPRT-deficient mice relative to their normal littermates, which compares to a 70% to 90% reduction in Lesch–Nyhan patients (Finger et al., 1988). Hence, we would not expect to observe altered behavior in mice with a corrected HPRT gene, although we would predict that the levels of striatal dopamine would be restored to the wild-type level.

Although in these experiments the detection of targeting events is facilitated by the selection systems available for HPRT activity, strategies for targeting into nonselectable genes, using vectors containing selectable markers (Jasin and Berg, 1988; Mansour et al., 1988; Doetschman et al., 1988), or the detection of targeted clones by the polymerase chain reaction (Kim and Smithies, 1988), are being developed. Consequently, it should soon be possible to target modifications to any chosen gene in the mouse genome.

## Experimental Procedures

### Production of Homozygous *hprt*$^{b\text{-}m3}$/*hprt*$^{b\text{-}m3}$ Female Mice

Mice of inbred strains 129/Ola, C57BL/6/Ola, and CBA/Ca/Ola and outbred strain MF1 were obtained from Olac 1976 Ltd., Bicester, Oxon. A male E14TG2a chimera showing 100% transmission of the $c^{ch}$ allele (Hooper et al., 1987) was mated to a wild-type MF1 female. Female offspring, heterozygous for the null *hprt*$^{b\text{-}m3}$ allele (generation 1), were mated to MF1 males, and male offspring (generation 2) carrying the null allele were identified by enzyme assay on blood samples. These HPRT-deficient males were mated to MF1 females, and embryos were transferred to foster mothers to improve their health status. Heterozygous female offspring (generation 3) were mated to strain 129/Ola males. HPRT-deficient male offspring were identified by enzyme assay on blood samples, and heterozygous female offspring were identified by Southern blot analysis of DNA obtained by tail biopsy (generation 4). HPRT-deficient males from generation 4 were mated to heterozygous females from either generation 3 or 4. Homozygous *hprt*$^{b\text{-}m3}$/*hprt*$^{b\text{-}m3}$ females were identified by enzyme assay on blood samples, and their status was confirmed by tail blots. Segregation of coat color markers in these animals made it possible to choose white homozygotes for mating to chimeras produced with corrected stem cells.

### Embryonic Stem Cells

The wild-type male ES cell line E14 was derived from strain 129/Ola blastocysts by trypsinization of immunosurgically isolated inner-cell masses. The dissociated cells were cultured on STO fibroblast feeder layers in Buffalo rat liver cell (BRL)-conditioned medium (A. H. Handyside, G. T. O'Neill, M. Jones, and M. L. Hooper, submitted). The HPRT-deficient E14TG2a cell line was isolated as a spontaneous 6-TG$^R$ derivative of E14 (Hooper et al., 1987). For all the experiments reported here ES cells were grown in BRL-conditioned medium without a feeder layer. The medium used was Glasgow modified Eagle's medium (Flow Laboratories) supplemented with 1× nonessential amino acids and 1 mM sodium pyruvate (Flow Laboratories), and 5% (vol/vol) of both fetal calf serum and newborn calf serum (Sera-Lab, Sussex). Batches of serum were chosen for their ability to give good ES cell plating efficiencies (typically 10%–30%) without affecting the ability of ES cells to differentiate in vitro. Cells were grown in the absence of antibiotics in 37°C incubators with 5% $CO_2$ in air.

BRL-conditioned medium was made as follows: BRL cells were grown to confluence in 75 $cm^2$ tissue culture flasks. The medium was discarded, and 15 ml of fresh medium was added. This was collected after 2 days and replaced with another 15 ml of fresh medium. In total, 45 ml of conditioned medium, representing 6 days of collection, was obtained from each flask. The conditioned medium was filtered (pore size, 0.2 μm) and stored at −20°C.

ES cells were grown on gelatin-coated tissue culture flasks in 60% BRL-conditioned medium, 40% fresh medium, containing 0.1 mM β-mercaptoethanol. Typically, the medium was changed daily and cultures were split every 3–4 days. Under these conditions the cell number doubled every 24 hr and the majority of cells in the culture had the typical stem cell morphology with little spontaneous in vitro differentiation. We considered that the likelihood of obtaining germ line transmission from corrected stem cells would be enhanced by the use of early-passage cells and by minimizing their time spent in culture. For this reason correctant clones generated in the first electroporation experiment from late-passage stem cells were not used for blastocyst injections. For the second experiment, E14TG2a cells (passage 25 since the isolation of the E14 line) were grown up and the targeting procedures and analysis were carried out as rapidly as possible. The cumulative passage number of correctant type 2-1 cells at the time of blastocyst injection had risen to passage 36.

### Vectors

The correcting vector used in these experiments, pDWM101, is similar in structure to the vector described by Doetschman et al. (1987) except that, in pDWM101, all correcting sequences are derived from the mouse HPRT gene (see Figure 1). pDWM101 was constructed by first cloning a 4.7 kb BamHI–EcoRI restriction fragment, containing exon 2, and derived from the mouse HPRT genomic clone, λHPT1 (Melton et al., 1984), into BamHI- and EcoRI-cut pUC8 (Vieira and Messing, 1982). A 3.0 kb BglII–BamHI fragment containing the HPRT promoter and exon 1, derived from λHPT32, was then inserted into the BamHI site, and a 1.35 kb EcoRI fragment, containing exon 3, isolated from λHPT1, was cloned into the EcoRI site. The resulting 11.7 kb plasmid, pDWM101, has the same organization as the equivalent part of the HPRT gene itself except that the first intron is reduced from 10.8 kb in the gene to 4.1 kb in the vector. The correcting vector can only rescue HPRT-deficient stem cells by homologous recombination with the defective gene. As a positive control for these correction experiments, an HPRT minigene derivative of pDWM101 was constructed by replacing the 1.25 kb XhoI–EcoRI fragment from pDWM101 with the 1.8 kb XhoI–EcoRI fragment from the minigene pDWM1 (Melton et al., 1986), which contains the remainder of the coding sequence and the 3′ untranslated region. This minigene derivative, pDWM100, is fully functional and can rescue HPRT-deficient stem cells without the requirement for homologous recombination. A rearranged version of the correcting vector, pDWM102, in which an XhoI digest is used to excise the pUC8 sequences prior to recombination, was also constructed: A 7.8 kb SalI (pUC8 polylinker site to the 5′ side of correcting sequence)–XhoI (site within exon 3) fragment from pDWM101 was first cloned into a pUC8 derivative, where an XhoI linker had been inserted into the SmaI site within the polylinker. This plasmid was then linearized at a SalI site immediately upstream of the HPRT promoter region, and a 1.25 kb fragment, containing the remainder of exon 3 and the following intervening sequence, from pDWM101, was inserted to generate the 11.7 kb plasmid, pDWM102, with the structure shown in Figure 1.

### Electroporation and Selection for Correctants

Correcting vector DNA was introduced into ES cells by electroporation (Bio-Rad Gene Pulser). Approximately 5 × $10^6$ cells were resuspended in 0.8 ml of HEPES-phosphate buffered saline (pH 7.05) (Graham and van der Eb, 1973) with 20 μg of XhoI-linearized vector DNA. The cells were given a single pulse (800 V, path length 0.4 cm, 3 μF), then allowed to stand at room temperature for 10 min before plating in nonselective medium. HAT selection (Littlefield, 1964) was imposed after 24 hr.

### Southern and Northern Hybridizations

High molecular weight genomic DNA was prepared by the method of Pellicer et al. (1978). Total RNA was prepared from cultured cells, or mouse tissues, as described previously (Melton et al., 1981). GeneScreen

Plus (New England Nuclear Products) was used as the hybridization membrane. The conditions for transfer, hybridization, washing, and rehybridization were those recommended by the supplier. For Southern hybridizations the probe was either the purified insert from the mouse HPRT cDNA recombinant pHPT5 (Konecki et al., 1982) or plasmid pUC8. For Northern hybridizations the probe was the purified insert from pHPT4 (Konecki et al., 1982). Probes were labeled by the random priming method (Feinberg and Vogelstein, 1983).

### Blastocyst Injections and Production of Chimeric Mice

Cells for microinjection were washed, trypsinized and resuspended in 10 mM HEPES-buffered Glasgow modified Eagle's medium containing 10% (vol/vol) newborn calf serum. F2 (C57BL/6/Ola × CBA/Ca/Ola) blastocysts were collected on the day of injection by flushing uterine horns 3.5 days postcoitum with Glasgow modified Eagle's medium containing 10% (vol/vol) newborn calf serum. Approximately 12–15 cells were injected into each blastocyst, using a Leitz micromanipulator and needles prepared with a Campden Instruments puller and microforge. Injected blastocysts were allowed to re-expand for 1–2 hr following manipulation and were then implanted into the uterine horns of pseudopregnant C57BL/6/Ola mice, mated 2.5 days previously to vasectomized F1 (C57BL/6/Ola × CBA/Ca/Ola) males.

### Acknowledgments

We are grateful to Carolanne McEwan, Mary Jones, Angus King, and John McKeating for invaluable technical assistance; to Professor Matt Kaufman for the use of his micromanipulation facility; to Dr. David Whittingham for carrying out the first three generations of breeding from E14TG2a chimeras; to Gerry O'Neill for G-banding of type 2-1 metaphase spreads and chromosome counts on one-cell embryos; to the staff of the Faculty of Medicine Animal Area for animal care; to Annie Wilson, Graham Brown, and Craig Walker for artwork and photography; to Janet Panther for preparing the manuscript; and to Paul Simons for the use of his electroporator. S. T. was supported by a training award from the Science and Engineering Research Council. The work was funded by the Cancer Research Campaign (M. L. H.), and Agricultural and Food Research Council, and the Medical Research Council (D. W. M.).

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

Received December 8, 1988.

### References

Bradley, A., Evans, M., Kaufman, M. H., and Robertson, E. (1984). Formation of germline chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255–256.

Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576–578.

Doetschman, T., Maeda, N., and Smithies, O. (1988). Targeted mutation of the *Hprt* gene in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA *85*, 8583–8587.

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *132*, 6–13.

Finger, S., Heavens, R. P., Sirinathsinghji, D. J. S., Kuehn, M. R., and Dunnett, S. B. (1988). Behavioural and neurochemical evaluation of a transgenic mouse model of Lesch-Nyhan syndrome. J. Neurol. Sci. *86*, 203–215.

Graham, F. L., and van der Eb, A. J. (1973). New technique for assay of infectivity of human adenovirus 5 DNA. Virology *52*, 456–467.

Hooper, M. L. (1982). Metabolic co-operation between mammalian cells in culture. Biochim. Biophys. Acta *651*, 85–103.

Hooper, M. L., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature *326*, 292–295.

Isamat, M., Macleod, K. F., King, A., McEwan, C., and Melton, D. W. (1988). Characterization, evolutionary relationships and chromosome location of processed mouse HPRT pseudogene. Somatic Cell Mol. Genet. *14*, 359–369.

Jaenisch, R. (1988). Transgenic animals. Science *240*, 1468–1475.

Jasin, M., and Berg, P. (1988). Homologous integration in mammalian cells without target gene selection. Genes Dev. *2*, 1353–1363.

Kim, H.-S., and Smithies, O. (1988). Recombinant fragment assay for gene targeting based on the polymerase chain reaction. Nucl. Acids Res. *16*, 8887–8903.

Konecki, D. S., Brennand, J., Fuscoe, J. C., Caskey, C. T., and Chinault, A. C. (1982). Hypoxanthine phosphoribosyltransferase genes of mouse and Chinese hamster: construction and sequence analysis of cDNA recombinants. Nucl. Acids Res. *10*, 6763–6775.

Kuehn, M. R., Bradley, A., Robertson, E. J., and Evans, M. J. (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature 326, 295–298.

Lesch, M., and Nyhan, W. L. (1964). A familial disorder of uric acid metabolism and central nervous system function. Am. J. Med. *36*, 561–570.

Littlefield, J. W. (1964). Selection of hybrids from matings of fibroblasts in vitro and their presumed recombinants. Science *145*, 709.

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 strategy for targeting mutations to non-selectable genes. Nature *336*, 348–352.

Martin, G. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. USA *78*, 7634–7638.

Melton, D. W., Konecki, D. S., Ledbetter, D. H., Hejtmancik, J. F., and Caskey, C. T. (1981). *In vitro* translation of hypoxanthine phosphoribosyltransferase mRNA: characterization of a mouse neuroblastoma cell line that has elevated levels of hypoxanthine phosphoribosyltransferase protein. Proc. Natl. Acad. Sci. USA *78*, 6977–6980.

Melton, D. W., Konecki, D. S., Brennand, J., and Caskey, C. T. (1984). Structure, expression and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA *81*, 2147–2151.

Melton, D. W., McEwan, C., McKie, A. B., and Reid, A. M. (1986). Expression of the mouse HPRT gene: deletional analysis of the promoter region of an X-chromosome linked housekeeping gene. Cell *44*, 319–328.

Pellicer, A., Wigler, M., Axel, R., and Silverstein, S. (1978). The transfer and stable integration of the HSV thymidine kinase gene into mouse cells. Cell *14*, 133–141.

Silvers, W. K. (1979). The Coat Colors of Mice: A Model for Mammalian Gene Action and Interaction (New York: Springer-Verlag).

Stout, J. T., and Caskey, C. T. (1988). The Lesch-Nyhan syndrome: clinical, molecular and genetic aspects. Trends Genet. *4*, 175–178.

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

Vieira, J., and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259–268.

## Information for Contributors (1989)

### Scope of Articles

*Cell* publishes reports of research of exceptional significance in any area of biology. Papers will be considered for publication only if they report novel results of interest to a wide audience. *Papers should be as brief as possible and should occupy between 2 and 10 published pages; authors should not expect to present more than one paper in an issue.*

### Editorial Offices

Authors in the United States should submit to the Principal Office, and authors in Europe to the European Office.

### Time of Publication

We try to let authors have reviewers' comments within two to three weeks. Papers that do not conform to the general criteria for publication in *Cell* will be returned immediately to authors, without detailed review, to avoid unnecessary delay in submission elsewhere. Articles will be published within three months of acceptance. When papers are accepted subject to revision, only a single revised version will be considered, so long as it is submitted within three months of the acceptance in principle.

### Submission of Manuscripts

One original and two copies of each manuscript should be submitted. Where appropriate, original photographs should be provided with each copy. When a submitted manuscript is closely related to papers that are in press or have been submitted elsewhere, it will be considered for publication in *Cell* only after copies of these papers also have been provided. *Under no circumstances will any paper be considered for* Cell *that contains any data that have been or will be submitted for publication elsewhere (including symposium volumes).* Nor will any paper appear in *Cell* if the authors communicate the results to the national press or comparable media before the expected date of publication.

### Organization of Manuscripts

The Summary must be a single paragraph of no more than 120 words. The Results and Discussion sections may be subdivided by further headings or may be combined. Experimental Procedures follow the Discussion. Any tables and legends to figures should be typed as separate sections, following the style of the journal. Footnotes should not be used and will be transferred to the text to appear in parentheses. The entire manuscript should be typed with double spacing (including table legends and references). Ten published pages of *Cell* correspond to a principal text (title through Discussion) of 17 typed pages, accompanied by <7 single-column figures. Manuscripts typed on dot-matrix printers are not acceptable. In addition to the full title, a running title of less than 50 characters should be supplied.

Only genetic loci should be italicized; journal names and Latin phrases and names will be set in Roman type. Any nonstandard abbreviations should be defined in the text when first used.

### Typesetting from Disk

It will speed publication when manuscripts can be typeset directly from authors' floppy disks. It is preferable if the disk is formatted under MS-DOS (PC-DOS), but we can also handle some other formats, including MacIntosh. We can handle files saved in most word processor formats. Please indicate which word processor was used to generate the file, the type of computer, and the operating system. With a MacIntosh computer, please save the file in the MS-DOS compatible option, if one is available.

### References

Only articles that have been published or are in press may be included in the references. Authors are encouraged to rely as far as possible upon citation of articles published in primary research journals. Unpublished results (including articles submitted for publication) or personal communications should be cited as such within the text. Personal communications should be substantiated by a letter of permission. Abstracts of papers presented at meetings may not be cited.

### Illustrations and Cover Art

A set of original drawings or photographs should be provided for reproduction of the figures (six or seven, maximum). Original photographs should preferably be cropped to the final intended width, and be no larger than 12 inches by 10 inches. *All figures will be set to a width of precisely 3 inches unless the authors indicate that a greater width should be used* (these distances include any lettering on the vertical axes). Nucleic acid sequences should be presented as glossy prints or original typescripts. It is preferable for magnifications to be indicated by means of a bar on the photograph.

Black and white or color prints (no slides), 5 inches by 7 inches or larger, will be considered for reproduction on the cover of *Cell*. Cover art should accompany submitted manuscripts, and a short descriptive summary should be included.

### Proofs

Page proofs should usually be received about eight weeks after acceptance of an article. *It is necessary to regard submitted manuscripts as final texts to which no changes should be made at the proof stage apart from correction of printer's errors.*

### Distribution of Material

Publication of a research article in *Cell* is taken to imply that the authors are prepared to distribute freely to interested academic researchers for their own use any clones of cells or DNA or antibodies, or other similar materials, used in the experiments that have been reported. Authors may be required to make primary data available to the Editor in cases of dispute. Nucleic acid and protein sequences published in *Cell* should be deposited in the appropriate data bank.