An investigation into the role of methylation in mammalian X-chromosome inactivation

T. Ian Simpson

A Thesis submitted for the degree of

Doctor of Philosophy

In the University of Oxford

Linacre College

and the

Genetics Unit

Department of Biochemistry

Oxford

MEDICAL SCIENCE LIBRARY OXFORD
27 APR 2000

Trinity Term 1999

Acknowledgements

I would like to thank Dr. Garry Brown and Ruth Brown for their supervision, encouragement and helpful advice which proved invaluable.

In addition I would like to thank :

The members of the Genetics Unit past and present for making it such a pleasant place to work.

David Hunter for all his help and unequalled problem solving abilities.

Gillian Oldfield for essential coffee and milk supplies and for not counting how many I had.

Professor Ian Craig for the collaboration on the Norrie project and Dr. Sally Boulton for providing retinal tissue samples.

John Elder and Jeremy Rowntree for superb computing support at the Department of Biochemistry.

The staff of the Biochemistry Department Sequencing Facility.

The staff of the HGMP-RC for providing excellent bioinformatics support and the RPCI21, LLNL-X and CpG-island genomic libraries and clones.

Mum and Dad for their tremendous support, love and encouragement which has always allowed me to achieve far more than I could ever have hoped for.

My brother for all his help and support, but also for great laughs and emergency meetings at local hostelries.

Last, but definitely not least, Julie for putting up with my endless ramblings about experiments, for help, advice, and proof-reading. Most of all for her love and understanding which I treasure.

Abstract

An investigation into the role of methylation in mammalian X-chromosome inactivation

T. Ian Simpson, Linacre College, Oxford.

Submitted for the degree of Doctor of Philosophy, Trinity Term 1999

X-chromosome inactivation achieves dosage compensation of X-linked genes between male (XY) and female (XX) mammals. This process involves the down-regulation of most, but not all genes on one of the two X-chromosomes in the nucleus of each female somatic cell. The mechanism of X-inactivation has yet to be elucidated in full, but is known to involve the non-coding transcript of the *Xist* gene, DNA methylation, histone hypo-acetylation and the condensation of higher order chromatin. Recent studies have established mechanisms linking methylation to repressive chromatin structures through methyl-binding proteins and histone deacetylase complexes.

In order to better understand the role of methylation in X-inactivation, the promoters of the human Pyruvate dehydrogenase E1α (PDHA1) and the human and murine Norrie disease protein (NDP/*Ndp*) genes were subjected to direct methylation sequencing, allowing the definition of methylation profiles at nucleotide resolution. The promoter of the PDHA1 gene was found to be hyper-methylated on the inactive X-chromosome and hypo-methylated on the active X-chromosome in agreement with studies at the promoters of other X-linked housekeeping genes. Methylation at the promoters of the NDP/*Ndp* genes was extensively investigated in a range of primary tissues and cell lines. The *Ndp* promoter was found to be methylated on both active and inactive X-chromosomes, but hypo-methylated in the proximal promoter exclusively in tissues that expressed the *Ndp* gene. The NDP promoter was found to be unmethylated on the active X-chromosome and hyper-methylated across the proximal promoter on the inactive X-chromosome in expressing cell lines and human retinal tissues. The novel promoter sequences of the human and murine SMCX/*Smcx* genes were isolated for comparative analysis and to provide a future resource for studying methylation at the promoters of genes which escape the X-inactivation process.

Promoter sequences of the PDHA1, NDP/*Ndp* and SMCX/*Smcx* genes were screened for putative transcription factor binding sites and for conserved CpG-dinucleotide content. Promoter-reporter gene constructs for these genes were transfected into mammalian cells establishing that the sequences studied were functional promoters. Artificial methylation of these constructs was shown to repress their promoter activities.

Table of Contents

Abbreviations *7*

Chapter 1- General Introduction *9*

1.1 X-chromosome inactivation. **9**

1.1.1 Dosage compensation. **9**

1.1.2 The hypothesis of X-chromosome inactivation. **10**

1.1.3 The "X-inactivation centre" (XIC/*Xic*). **12**

1.1.4 The *XIST/Xist* (X-inactive specific transcript) gene. **14**

1.1.5 Refining the *Xic* – through transgenesis. **16**

1.1.6 The blocking factor model of X-inactivation (The Rastan Model). **21**

1.1.7 The *Xist* gene promoter. **24**

1.1.8 Imprinted *Xist* expression **24**

1.1.9 *Xist* RNA stability **26**

1.1.10 Genes escaping X-inactivation. **27**

1.1.11 Chromatin and the inactive X-chromosome **28**

1.1.12 DNA Methylation and the X-chromosome. **32**

1.2 Genomic DNA Methylation **36**

1.2.1 Distribution of methylation in the mammalian genome through development. **36**

1.2.2 Methylation and the control of gene expression **37**

1.2.3 Methods of analysing genomic DNA methylation. **40**

1.3 Outline of this study **46**

Chapter 2 - Materials and Methods *49*

2.1 General Solutions. **49**

2.2 PCR buffers. **49**

2.3 Solutions for hybridisation. **50**

2.4 Solutions for DNA sequencing. **50**

2.5 Loading Buffers **51**

2.6 DNA isolation reagents. **51**

2.7 RNA isolation reagents. **53**

2.8 Bacterial Media and Reagents. **53**

2.9 Vectors. **54**

2.10 Bacterial Strains. **56**

2.11 Libraries. **56**

2.12 Molecular Weight ladders. **57**

2.13 Tissue Culture Reagents and Cell lines. **57**

2.14 Tissue culture cell lines. **58**

2.15 Genomic DNA isolation from tissue culture cells. **59**

2.16 Genomic DNA isolation from peripheral blood leukocytes **60**

2.17 Genomic DNA isolation from tissues **60**

2.18 RNA isolation from tissue culture cells. **61**

2.19 RNA isolation from tissues. **62**

2.20 Isolation of PAC DNA. **62**

2.21 Plasmid Maxi DNA Preparation **63**

2.22 Plasmid DNA miniprep. **64**

2.23 Rapid Miniprep of plasmid DNA. 65

2.24 Restriction enzyme digests. 66

2.25 Agarose gel electrophoresis. 66

2.26 Gel extraction of DNA. 67

2.27 General DNA Clean-up. 68

2.28 Ethanol precipitation of DNA. 68

2.29 Southern blotting from agarose gels. 69

2.30 Multiprime labelling of DNA probes 70

2.31 Spin dialysis – Sephadex G-50. 71

2.32 Hybridisation 71

2.33 Membrane washing, autoradiography and stripping. 72

2.34 Preparation of chemically competent cells. 73

2.35 Chemical transformation of bacterial cells 73

2.36 Preparation of electro-competent cells 73

2.37 Electroporation of the bacterium *Escherichia coli* 74

2.38 Bacterial frozen stocks 74

2.39 Cloning of DNA fragments into plasmid vectors. 75

2.40 PCR screening of bacterial clones. 78

2.41 DNA sequencing. 79

2.42 Fluorescent cycle sequencing of DNA (Automatic sequencing). 81

2.43 Polyacrylamide gel electrophoresis of DNA sequencing reactions. 82

2.44 Polymerase Chain Reaction (PCR). 83

2.45 First strand cDNA synthesis. 86

2.46 RT-PCR reactions. 86

2.47 Frommer bisulphite sequencing. 87

2.48 Artificial methylation of promoter-constructs. 89

2.49 Tissue Culture Methods. 90

2.50 Electroporation of tissue culture cells. 91

2.51 Lipid based transfection with FuGene™6 transfection reagent. 91

2.52 Making 4% Paraformaldehyde (PFA) fixative. 92

2.53 Cold fixation of adherent cells. 92

2.54 Moviol aqueous mountant preparation. 93

2.55 Visualisation of slides 93

2.56 Computer programs. 93

Chapter 3 - Methylation analysis of the Human Pyruvate dehydrogenase E1α gene promoter 96

3.1 Introduction 96

3.1.1 Housekeeping genes on the X-chromosome. 96

3.1.2 The Pyruvate dehydrogenase E1α gene. 99

3.2 Results. 100

3.2.1 CpG-Plots. 100

3.2.2 Methylation sensitive PCR (msPCR). 101

3.2.3 Bisulphite sequencing. 103

3.2.4 Expression analysis. 108

3.2.5 Isolation of mouse PAC clones of the murine *Pdha1* gene. 112

3.3 Discussion. 114

Chapter 4 - Methylation Analysis of the Human and Murine Norrie Disease Protein (NDP/Ndp) Gene Promoters 129

4.1 Introduction 129

4.1.1 Tissue-specific genes on the X-chromosome. 129

4.1.2 The Norrie disease protein (*NDP/Ndp*) gene 131

4.2 Results. 134

4.2.1 CpG-Plots. 134

4.2.2 Bisulphite sequence analysis of the human *NDP* gene promoter 136

4.2.3 Bisulphite sequence analysis of the mouse *Ndp* gene promoter. 141

4.2.4 Expression analysis. 144

4.3 Discussion. 150

Chapter 5 - Isolation of the human and murine SMCX/Smcx gene promoters 162

5.1 Introduction 162

5.1.1 Genomic DNA libraries. 163

5.1.2 Methods for studying X-chromosome inactivation in the mouse. 165

5.1.3 Methods for studying escape from X-chromosome inactivation in the human 166

5.1.4 The *SMCX/Smcx* gene 168

5.2 Results. 170

5.2.1 Generation of the human *SMCX* cDNA probe. 170

5.2.2 Screening the Human LLNL X-only cosmid library. 171

5.2.3 Analysis of the Human *SMCX* upstream region. 173

5.2.4 Screening the Murine RPCI21 genomic PAC library. 174

5.2.5 Analysis of the Murine *Smcx* gene upstream region.....178

5.2.6 Comparison of the human and murine *SMCX/Smcx* upstream sequences.....179

5.2.7 Expression analysis of the *SMCX/Smcx* gene. 180

5.3 Discussion. 185

Chapter 6 - General Discussion 194

6.1 General features of promoter comparison and sequence analysis. 194

6.2 Interpreting methylation profiles. 196

6.3 Assessing the effects of methylation on gene transcription. 202

6.4 Methylation and X-chromosome inactivation. 203

6.5 Future studies. 205

6.6 Conclusion. 206

***Bibliography*.....207**

Abbreviations

5aC	5-aza-cytidine
5aCdr	5-aza-2'deoxy cytidine
BAC	Bacterial artificial chromosome
CIAP	Calf intestinal alkaline phosphatase
CRE	CREB related element
CVS	Chorioinc villus sample
DMS	DNA methylation society
EGFP	Enhanced green fluorescent protein
ES	Embryonic stem
FISH	Fluorescent in-situ hybridisation
GSP	Gene specific primer
HDAC	Histone deacetylase
HGMP-RC	Human genome mapping project resource centre
LMPCR	Ligation mediated PCR
MBD	Methylation binding domain
MBP	Methyl binding protein
MCS	Multiple cloning site
mDNMT	Maintenance DNA methyl transferase
msPCR	methylation sensitive PCR
MSP	Methylation specific PCR
NHCP	Non-histone chromosomal protein
ORF	Open reading frame
PAC	P1 artificial chromosome
PAR	Pseudo autosomal region
PCR	Polymerase chain reaction
RT-PCR	Reverse transcribed PCR
SNuPE	Single nucleotide primer extension
TBP	TATA binding protein

TIF	Tagged image file
TRD	Transcription repression domain
X_i	Inactive X-chromosome
X^m	Maternally derived X-chromosome
X^p	Paternally derived X-chromosome
XCE/Xce	X-controlling element
XIC/Xic	X-inactivation centre
YAC	Yeast artificial chromosome

Chapter 1- General Introduction

1.1 X-chromosome inactivation.

X-chromosome inactivation is the method by which dosage compensation of X-linked genes is achieved between male (XY) and female (XX) mammals. This process involves the down-regulation of most, but not all genes on one of the two X-chromosomes in the nucleus of each female somatic cell. The X-inactivation process occurs early in female embryogenesis and there is normally an equal probability that either the maternal (X^m) or paternal (X^p) X-chromosome will be inactivated.

1.1.1 Dosage compensation.

Dosage compensation of X-linked genes is carried out by a variety of organisms to ensure that the different number of X-chromosomes between sexes resulting from primary sex-determination does not result in lethality, developmental delay or reduced fertility. This process has been studied extensively in the fruit fly, the nematode worm and in several mammals including primates, rodents and marsupials. These three evolutionarily distinct classes of organisms achieve X-linked dosage compensation in different, but related ways. The fruit fly *Drosophila melanogaster* compensates by up-regulating expression of the single X-chromosome of the male so that its level of expression is twice that of each female X-chromsome (Baker et al., 1994). Characteristic acetylation patterns of HistoneH4 lysine residues on the up-regulated male X-chromosome suggest that changes in chromatin structure are critical in the compensation process (Bone et al., 1994). In contrast the nematode worm *Caenorhabditis elegans* achieves dosage compensation between hermaphrodites (XX) and males (XO) by lowering the level of expression from both X-chromosomes of the

hermaphrodite by half (Hsu and Meyer, 1993). This process is modulated through a number of SMC (Structural Maintenance of Chromosome) genes (Strunnikov, 1998) including Dpy-27 (dumpy) which product coats the down-regulated hermaphrodite X-chromosomes (Chuang et al., 1994). Mammalian dosage compensation is achieved by a third method, females (XX) silence one of their two X-chromosomes to achieve dosage equivalence with males (XY) (Migeon, 1994). This compensation mechanism is somewhat unique amongst these examples in that X-chromosomes in the same nucleus are treated differently.

1.1.2 The hypothesis of X-chromosome inactivation.

X-chromosome inactivation in mammals was first proposed in landmark papers by Mary Lyon to explain a number of observations relating to gene control on the X-chromosome (Lyon, 1961; Lyon, 1961). The principle predictions of the hypothesis remain true to this day, namely that (1) the condensed X-chromosome (Barr-body) present in the cells of female mammals is genetically inactive, (2) that either the paternal or maternal X-chromosome can be inactivated and (3) that the X-inactivation process occurs early in development.

The hypothesis was primarily supported by two lines of evidence, genetic and cytogenetic. The cytogenetic evidence was gathered from studies looking at "sex-chromatin" or "Barr-bodies" in the cells and tissues of the mouse, rat, opossum and human (Ohno and Hauschka, 1960; Ohno et al., 1959; Ohno and Makino, 1961). Barr-bodies are satellites that are present in the nuclei of female cells and tissues. They were later identified as single condensed X-chromosomes and were proposed to be genetically inactive (Barr and Bertram, 1949).

The cytogenetic evidence was supported by a wealth of genetic evidence. The observation that female mice lacking one X-chromosome (XO) were fertile demonstrated that one active

X-chromosome per female cell was sufficient for normal development (Welshons and Russell, 1959). The second piece of genetic evidence arose from studies of sex-linked coat colour variation in mutant mice including *mottled*, *brindled*, *tortoiseshell* and *dappled* (Dickie, 1954; Fraser et al., 1953; Welshons and Russell, 1959). Heterozygous females of these mutants were found to have mosaic coat colours which were patched in either wild type, mutant or mixed colour patches. These phenotypes were reminiscent of somatic cell mosaicism in mice in which mutant or normal cells were seen either in exclusive or mixed zones (Russell and Major, 1957). Lyon proposed that the coat-colour variation was due to expression of one or other allele of the X-linked coat-colour gene from the single active X-chromosome in each female cell. The presence of wild-type, mutant and mixed colour patches was taken to imply that either of the parental X-chromosomes could be inactivated and that the process occurred early in development. Additional support for the theory came from mice which had translocations of autosomal coat colour genes onto the X chromosome (Cattanach, 1961; Russell and Bangham, 1960). These mice showed coat-colour mosaicism which was not observed when the gene was in its natural autosomal position. This was important because it established that the mosaicism was a feature of the X-chromosomal location of the gene and not due to the nature of the gene itself.

The extension of the X-inactivation hypothesis from mice to humans was of profound importance to the understanding of human sex-linked diseases. Direct evidence to support the expansion of the theory was limited at the time because although human females had been shown to have Barr-bodies there was only one human condition known to have a sex-linked colour phenotype at the time, ocular albinism (Falls, 1951). In this condition the retinal pigment epithelium is absent from males, but present in heterozygous females in mottled patches. This phenotype was considered analogous to the mosaicism seen in sex-linked

mouse coat-colour mutants and was consistent with the proposed effects of X-inactivation in humans.

A critical observation in establishing the theory of X-inactivation in mice was that XO females were phenotypically normal. Problematically, this was not the case in humans in which the XO genotype manifests as Turner syndrome whose typical features include mental retardation, small stature and gonadal dysgenesis (Ford et al., 1959). This discordance between man and mouse did not preclude X-inactivation occurring in humans, but suggested that if it did then the processes were not identical.

The absence of sex-linked phenotypes that could be easily scored for mosaicism in humans led to an alternate approach to gathering data to test the theory. If X-inactivation occurred in humans then the expectation would be that heterozygous manifesting females of X-linked diseases would display phenotypic variation in expression and incomplete penetrance. An analysis of heterozygous individuals suffering from one of sixteen different sex-linked conditions showed variation consistent with this prediction. Along with other cytological data this implied that X-inactivation also occurred in humans (Lyon, 1961).

1.1.3 The "X-inactivation centre" (XIC/*Xic*).

The idea that the X-inactivation process was facilitated by elements present on the X-chromosome itself was first proposed by Russell based on observations of balanced X-autosome translocations in mice (Russell and Bangham, 1960; Russell and Mongomery, 1965). In these events only one of the two resulting translocation products was capable of undergoing X-inactivation. Russell analysed a number of such translocations and began defining a putative "X-inactivation centre" (XIC/*Xic*) as a minimal region required to achieve

X-inactivation. By correlating the ability of a translocation product to X-inactivate with the position of the breakpoint a minimal *Xic* was defined, in the mouse, proximally by Searle's translocation T(X;16)16H (Rastan, 1983) and distally by the deletion breakpoint HD3 (Rastan and Robertson, 1985). This interval represents an estimated physical distance of 8-12Mb and contains at least six genes to date including *Xist* (see section 1.1.4, below) (Borsani et al., 1991; Brockdorff et al., 1991; Brown et al., 1991), *Xpct* (Avner et al., 1998), *Bpx* (Rougeulle and Avner, 1996), *Cdx4* (Horn and Ashworth, 1995), *Brx* (Simmler et al., 1997) and *Tsx* (Simmler et al., 1996) although more genes are likely to be contained within such a large genomic interval.

Similar mapping experiments with rearranged X-chromosomes in human cells and human/mouse somatic cell hybrids resolved an interval on the long arm of the X-chromosome within band Xq13 containing the function of the putative XIC (Willard, 1995). This region is estimated to be between 660 and 1200kb in size, considerably smaller than that of the murine *Xic*. The refined intervals in man and mouse are considered to be syntenic as the conserved markers *Phka, Pgk1, Xist,* and *Cdx4* are found within both regions.

The X-inactivation process is normally random in somatic cells so that each X-chromosome has an equal probability of being selected for inactivation. This feature of X-inactivation in the mouse was found to be biased in female progeny from certain inter sub-species crosses so that one parental X-chromosome was more likely to be inactivated than the other. A locus associated with this biasing effect, the X-controlling element (*Xce*), was genetically mapped to an interval overlapping that of the candidate *Xic* region (Cattanach and Rasberry, 1972). The *Xce* locus has four main alleles Xce^{a}, Xce^{b} and Xce^{c} and Xce^{d} which differentially affect the propensity of the X-chromosome on which they lie to undergo X-inactivation (Simmler et

al., 1993). Attempts to isolate the *Xce* locus itself have been made. Specific hyper-methylated sections of the active X-chromosome are associated with the *Xce* locus and influence the X-inactivation process at a very early stage in development (Courtier et al., 1995). The most recent molecular candidate for the *Xce* locus is a sequence of 34bp direct repeats lying approximately 15kb 3′ to the *Xist* gene whose differential methylation state correlates to *Xce* allele strength (Avner et al., 1998). How this relates to the choosing mechanism of the *Xic* remains unclear.

1.1.4 The *XIST/Xist* (X-inactive specific transcript) gene.

The *XIST* gene was identified by positional cloning within the human XIC candidate interval and found to be expressed exclusively from the inactive X-chromosome (Brown et al., 1991). The human *XIST* gene encodes a 17kb transcript that has no open reading frame (ORF), *XIST* RNA is localised to the nucleus and associates with the inactive X-chromosome (Clemson et al., 1996). The murine *Xist* gene was cloned through identity to the human *XIST* gene, its expression was found to be restricted to the murine inactive X-chromosome (Borsani et al., 1991; Brockdorff et al., 1991). The *Xist* gene encodes a 15kb transcript which is localised to the nucleus and contains no conserved ORF (Brockdorff et al., 1992). Comparison of the nucleotide sequences of the human and murine *Xist* transcripts showed a sequence homology of greater than 70% and the presence of several direct tandem repeats. The unique expression of the *XIST/Xist* transcript from the inactive X-chromosome, its localisation in the nucleus and its proposed association with the inactive X-chromosome led to proposals that it was the X-inactivation centre (Brockdorff et al., 1992; Brown et al., 1992).

A number of different experimental approaches were used to characterise the expression of the *Xist* genes in man and mouse. One of the earliest published analyses of *Xist* expression

was that of Kay et al. (1993) in which RT-PCR and RNA FISH techniques were used to determine the nature of *Xist* expression in 2-cell embryos through to E4.5 blastocysts. In the female, expression was first detected in the 8-cell embryo, prior to the overt onset of inactivation which occurs first in extra-embryonic cell lineages around E3.5 (Takagi and Sasaki, 1975). Expression in the male was restricted to germ cells prior to meiosis implying a potential role in silencing the X-chromosome during gametogenesis. Additional evidence relating *Xist* expression to the onset of X-inactivation was gathered from an experimental tissue culture system that used female embryonic stem (ES) cells. These cells undergo random X-inactivation upon differentiation and have been used by a number of groups as an *in-vitro* system for studying the X-inactivation process (Rastan and Robertson, 1985). Prior to differentiation, *Xist* is expressed at low levels from both X-chromosomes, but is up-regulated when the cells are allowed to differentiate confirming that in the *in-vitro* system *Xist* is expressed at a time consistent with a role in the initiation of X-inactivation (Kay et al., 1993).

The first direct evidence of the requirement for *Xist* to achieve X-inactivation was provided by the generation of a null *Xist* allele in a female ES cell line (Penny et al., 1996). The null allele had a 7kb deletion encompassing the first exon and 36bp of the *Xist* gene minimal promoter. Such a large deletion was considered essential to generate the null-allele as the functional components of the *Xist* RNA had not been established. Upon differentiation the mutant ES cells only expressed *Xist* from the non-targeted allele suggesting that the deleted allele was null. Late-replication studies confirmed that the differentiated ES cells had inactive X-chromosomes demonstrating that the disruption of the single *Xist* allele did not affect the ability to count the number of X-chromosomes in the cell. This was of great importance as it implied that *Xist* was not in itself the *Xic*, but one component of the X-inactivation machinery.

1.1.5 Refining the *Xic* – through transgenesis.

Recent transgenic experiments have attempted to further refine the sections of the X-chromosome required to satisfy all of the functions of the *Xic* (Herzing et al., 1997; Lee et al., 1996; Marahrens et al., 1998; Marahrens et al., 1997). These sections would need to be able to achieve the counting, choosing and initiation features of the X-inactivation process.

The first of the experiments introduced a 450kb transgene including the *Xist* gene into male ES cells. *Xist* expression was found to be induced from both endogenous and ectopic loci upon differentiation of the targeted cells (Lee et al., 1996). This experiment demonstrated that the ectopic transgene contained the counting, choosing and initiation functions attributed to the *Xic*. The paper presents evidence consistent with the ectopically expressed *Xist* RNA coating at least part of the autosome on which it is integrated and of inactivating a *lacZ* marker gene present on the transgene. These observations in particular have been contentious as the transgenic cell line used to analyse these two features contained the highest number of copies of the transgene (24) which appeared to express *Xist* at levels directly correlated to the number of transgene copies. The level of *Xist* expression in these ES cells would be vastly greater than that observed in normal differentiating female ES cells and conclusions about the function of the *Xic*s in their autosomal context are difficult to relate to real biology in this system.

Another transgenic construct used a similar approach to that of Lee et al., but with a construct refined to determine whether the *Xist* gene itself was sufficient to satisfy the functions of the *Xic* (Herzing et al., 1997). The construct comprised the *Xist* gene with 9kb of 5′ and 6kb of 3′-flanking sequence along with a *HMG*-βGEO expression/selection cassette. The βGEO

cassette comprises a neomycin and β-galactosidase gene fusion allowing selection with G418 to identify ES cell clones with the construct integrated into the genome in a position that allows the fusion to be expressed. The HMG-CoA reductase (*HMG*) promoter was selected to drive the expression of the βGEO selection cassette as it responds to X-inactivation (Tan et al., 1993) and was thought likely to be inactivated *in-cis* if an inactivation signal was initiated from the ectopic *Xist* locus.

The construct was transfected into male ES cells and clones that were resistant to G418 and expressed the β-galactosidase gene were isolated. Upon differentiation the levels of both *Xist* and β-galactosidase activity were measured to determine whether *Xist* expression was induced and β-galactosidase activity reduced. Essential control ES cell lines were also generated which showed that the *HMG* promoter was not itself down-regulated during differentiation *in-vitro* and that the *Xist* 3′ flanking sequence could not induce the inactivation of the βGEO cassette without the *Xist* gene. Data from two ES cell lines was presented which was consistent with the *Xist* gene inactivating the *HMG*-βGEO cassette *in-cis*. The fact that *Xist* expression was induced upon differentiation and was capable of inactivating a reporter gene *in-cis* strongly supported the argument that *Xist* and the flanking regulatory sequences contained within the Herzing construct were capable of performing the functions attributed to the *Xic*. In order to further address whether the counting function of the *Xic* was contained within the construct the expression of the endogenous *Xist* gene was measured by semi-quantitative RT-PCR using sequence polymorphisms between the two alleles of *Xist* present in the cell lines. Up to 30% of the *Xist* expression in the differentiating ES cell clones was from the endogenous copy of *Xist*. This suggested that the components of the X-counting system were in operation, but as with the *Xic* YAC transgenic described above, multiple copy

integration at the target site makes it impossible to demonstrate that all copies of the ectopic *Xic* are correctly counted so that only one *Xic* remains active per cell.

The next transgenic experiment attempting to identify the elements of the *Xic* involved another *Xist* deletion transgenic similar to that described by Penny et al. (1996). This deletion lay 3′ to that previously described and comprised a deletion of exons 1-5 of the *Xist* gene (Marahrens et al., 1997). The *Xist* gene was knocked out in female ES cells and transgenic mice were generated. When the mutant *Xist* allele was inherited on the paternal X-chromosome embryos died early in embryogenesis, but were normal when the mutant allele was inherited on the maternal X-chromosome. In wild-type mice, X-inactivation is heavily skewed in favour of inactivating the paternal X-chromosome in extra-embryonic tissues presumably due to an imprinting effect at the *Xic* that is not erased in these tissues (Takagi and Sasaki, 1975). Parthenogenetic embryos which inherit both X-chromosomes from their mother (gynogenotes) are unable to inactivate either X-chromosome in extra-embryonic tissues and die early in embryogenesis (Tada and Takagi, 1992). Paternal inheritance of the mutant *Xist* allele in the Marahrens mice had the same effect with neither X-chromosome being inactivated in extra-embryonic tissues due to the imprint on the *Xic* and the null paternal *Xist* allele. The work of Marahrens et al. (1997) agreed with that of Penny et al. (1996) in both cases X-chromosomes carrying the mutant *Xist* allele were never inactivated establishing that *Xist* RNA was required for X-inactivation to occur.

A critical issue to address in the light of the work of Penny et al. (1996) and Herzing et al. (1997) was whether cells with a deleted *Xist* allele were also deficient in the counting and choosing functions of the *Xic*. If they were then the three main features attributed to the *Xic* would be satisfied by the *Xist* gene itself, consistent with the *Xist* gene and the *Xic* being

synonymous. The Penny et al. (1996) knockout experiments concluded that the *Xist* gene did not contain all of the elements of the *Xic* whereas the Herzing et al. (1997) experiments suggested that it did. This issue has been addressed in recent work by Marahrens et al. (1998) in which transgenic mice heterozygous for a null allele of the *Xist* gene were studied. The *Xist* deletion in these mice lay 3′ to that of Penny et al. (1996) comprising a deletion of exons 1-5 (Marahrens et al., 1997). In order to describe the differences in the conclusions of the Penny and Marahrens papers it is necessary to consider the expected differences resulting from deletions in the *Xist* gene either including or excluding deletion of a putative X-chromosome choosing element.

If the choosing element of the X-inactivation process and *Xist* are removed in undifferentiated XX ES cells it would be expected that on differentiation, primary non-random X-inactivation would result (see Figure1.1). If *Xist* was removed, but not the choosing element then it would be expected that secondary non-random X-inactivation would occur. In this case cells would be committed to the choice of which X-chromosome they were to inactivate on differentiation, but those which had chosen to inactivate the X-chromosome carrying the null *Xist* allele would be unable to do so and would be selected against. The end-point for differentiating ES cells with both kinds of deletion should be identical, inactivation of the non-targeted allele in every cell. However, the difference in gene expression between the two X-chromosomes on the way towards this end-point provides a means to distinguish them experimentally.

The work of Penny et al. (1996) demonstrated that the choosing function of the *Xic* was most likely retained in differentiating XX ES cells heterozygous for their null *Xist* allele. Allele-specific Single Nucleotide Primer Extension (SNuPE) was used to quantitate the relative

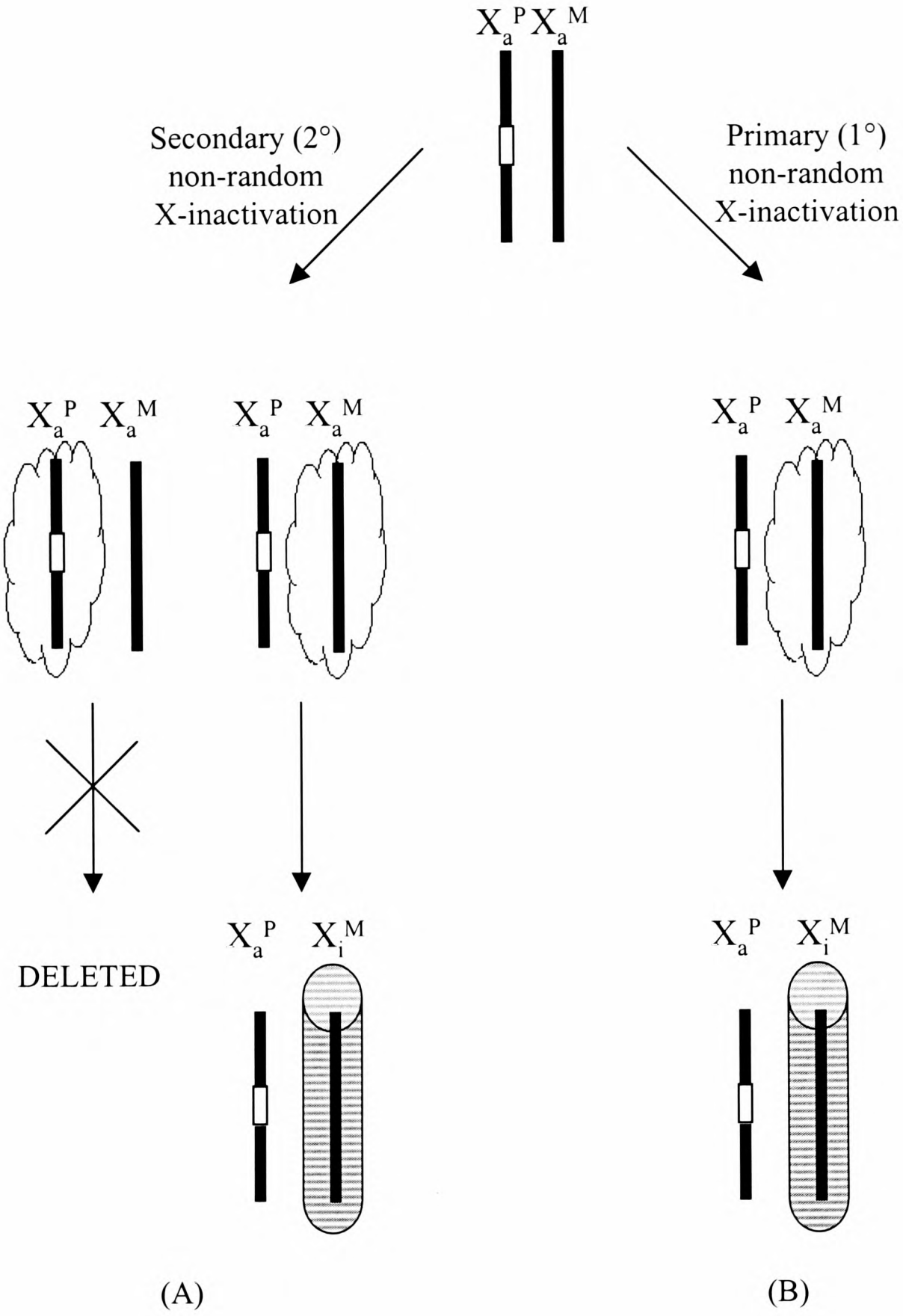


Figure 1.1: Schematic showing primary and secondary non-random X-inactivation. In the diagram the white box represents a deleted *Xic* or *Xist* allele. If the choice element(s) of the *Xic* are not removed then secondary non-random inactivation occurs, a random choice is made as to which X-chromosome to inactivate and cells which chose to inactivate the X-chromosome carrying the mutant *Xic* would be deleted. Those that chose to inactivate the X-chromosome carrying the normal Xic would be able to X-inactivate normally and survive (A).Alternatively if the choice element(s) of the *Xic* are deleted then primary non-random inactivation occurs and the same X-chromosome is inactivated in every cell (B).

level of expression of the *Pgk-1*, *Rps4* and *Smcx* genes from both X-chromosomes. Their relative levels of expression were consistent with secondary non-random X-inactivation in these cells. The authors concluded that the counting and choice elements of the X-inactivation process were likely to lie outside the *Xist* gene. Aggregation chimeras made from these ES cells showed non-random X-inactivation with each cell inactivating the X-chromosome carrying the non-deleted allele. This outcome is consistent with either primary or secondary non-random X-inactivation occurring in these animals.

Marahrens et al. suggested that it was possible to distinguish primary from secondary non-random X-inactivation in mice heterozygous for a null *Xist* allele by weighing them if the process of selection was comparable to that in Searle's mice (see section 5.1.2) (Marahrens et al., 1998). Secondary non-random X-inactivation occurs in Searle's mice because cells which choose to inactivate the *Xic* on the translocated X-chromosome can only inactivate genes on one of the two translocation products. The resulting imbalance of X-linked gene expression is cell-lethal and leads to the complete elimination of all cells that choose to inactivate the translocated X-chromosomes by E12.5, probably around days E7-E8 (McMahon and Monk, 1983). Mice heterozygous for Searle's translocation weigh 21% less as neonates than wild-type litter mates (Lyon et al., 1964). Mice heterozygous for the Marahrens null *Xist* allele weighed the same as their wild-type litter mates consistent with primary non-random X-inactivation in these mice. Analysis of mutant and normal *Xist* RNA expression by RNA fluorescent in-situ hybridisation (FISH) was used in support of these findings showing that at E7.5, E8.5 and E9.5 essentially all cells were expressing the wild-type *Xist* allele. These result are superficially in disagreement with the findings of Penny et al. (1996) whose *Xist* deletion did not affect the choosing function of the *Xic* (i.e. secondary non-random

inactivation was occurring in their chimeras), but the two experiments are difficult to compare as they used different deletions of the *Xist* gene.

The simplest summary of the two *Xist* knockout experiments is to conclude that 1.) *Xist* is unequivocally required for X-inactivation to occur *in-cis*, 2.) deletion of the minimal promoter and exon one ablates *Xist* function, but not the choosing function of the *Xic* and 3.) deletion of exons one to five, but not the minimal promoter ablates *Xist* function and the choosing function of the *Xic*. Marahrens et al. (1997) define a segment of DNA which includes at least part of the choosing element of the *Xic* although it is possible that other components exist.

1.1.6 The blocking factor model of X-inactivation (The Rastan Model).

The blocking factor model was proposed by Sohaila Rastan to explain initiation of the X-inactivation process (Rastan, 1983). The model is based on the concept of a trans-acting blocking factor which is present in cells in a 1:1 stoichiometric ratio to the number of active X-chromosomes to be retained after X-inactivation (see Figure1.2). In a female cell with two X-chromosomes one complement of this factor would protect one X-chromosome from inactivation. In tetraploid female cells there would be two complements as two of the four X-chromosomes remain active per cell (one per complement of autosomes). This n(X):n(autosome) relation of dosage compensation systems is conserved in mammals, *D. melanogaster* and *C. elegans*, however the mechanism by which this is achieved has yet to be elucidated in mammals. The Rastan model must also work in cases of X-chromsome aneuploidy where all but one X-chromosome is inactivated, the so-called "n-1" rule. The Rastan model could be consistent with his if the single blocking factor present per set of

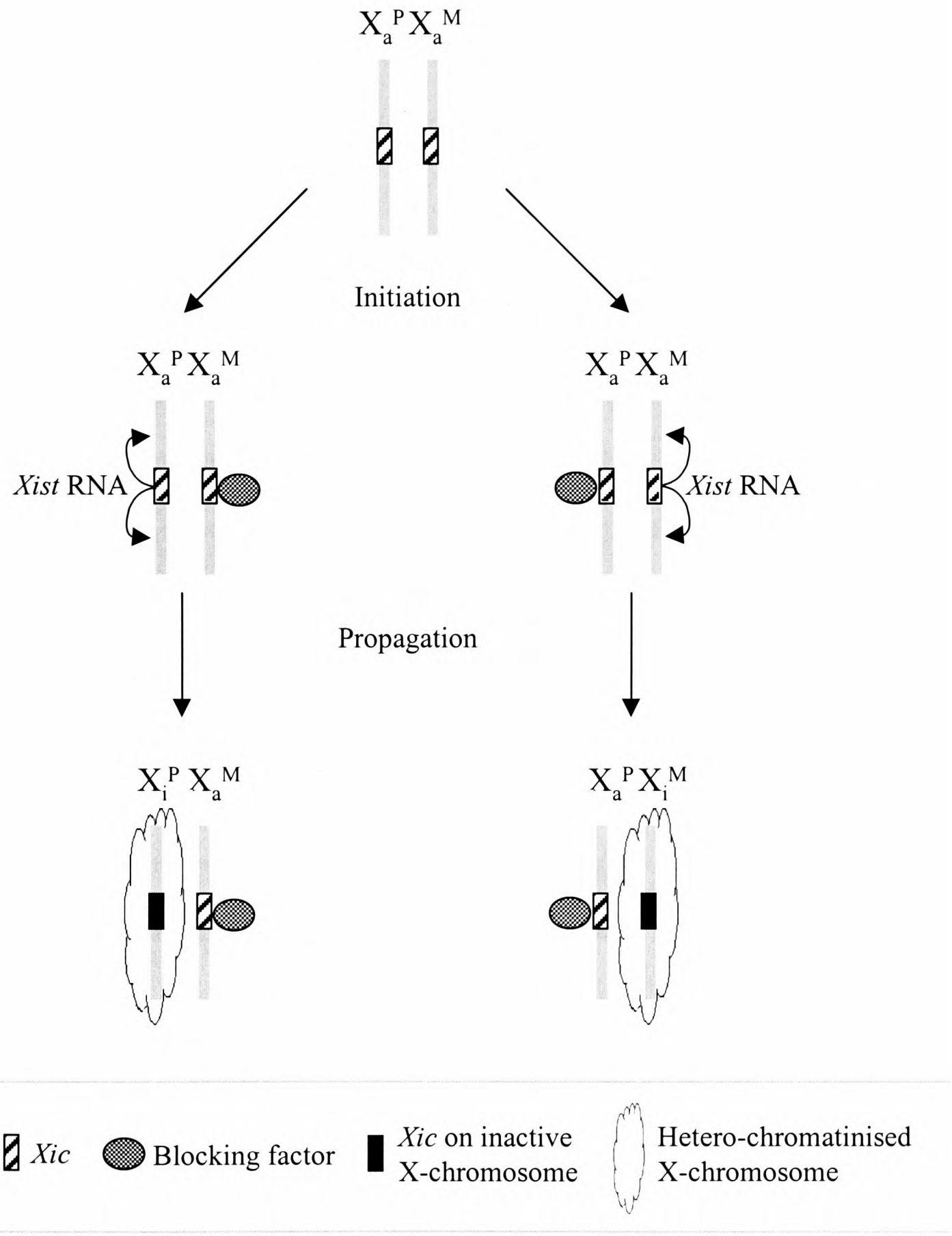


Figure 1.2: Rastan's blocking factor model for X-chromosome inactivation. In this model a unique blocking factor binds to a choice element in the *Xic* preventing initiation of X-inactivation from the X-chromosome to which it is bound. Initiation of X-inactivation occurs from the other X-chromosome and any additional X-chromosomes in the case of X-chromosome aneuploidy. When a cell is polyploid the blocking factor is envisaged to be present in equal quantities to the number of sets of autosomes so that one X-chromosome remains active per set.

autosomes protects a single X-chromosome from inactivation and the extra X-chromosomes are then inactivated by default.

The blocking factor described in Rastan's model could theoretically be the choosing component of X-inactivation. In the model the blocking factor binds an X-chromosome protecting it from inactivation, a choice event by definition. The putative element defined by Marahrens et al. (1998) is a leading candidate for the binding site of such a blocking factor. However, the observations of Marahrens et al. (1998) if correct prove that the blocking factor of Rastan cannot be the choosing factor. The Marahrens null *Xist* allele is incapable of effecting X-inactivation *in-cis* and removes a putative choosing element. If Rastan's blocking factor is the choosing factor then deletion of the choosing element (blocking factor binding site) would mean that the blocking factor would have to bind to the intact *Xist* allele and as a result block inactivation on the only X-chromosome capable of inactivating. This would produce two active X-chromosomes in every cell. Marahrens et al. (1998) observe primary non-random inactivation in the cells of their *Xist* deletion mice proving that X-inactivation occurs successfully in the absence of one choosing factor site.

These observations paved the way for a revised model of inactivation whereby both a blocking and choosing factor were involved in the initiation of X-inactivation (Marahrens et al., 1998). A model was proposed which required a unique blocking factor (akin to that of Rastan) and a ubiquitous initiation factor. In the model the blocking and the initiation factors bind to targets on the X-chromosome in a mutually exclusive way. The blocking factor is proposed to shuttle between *Xic*s until just prior to the onset of inactivation where it finally binds to one, the ubiquitous initiation factor then binds to the other X-chromosome and initiates X-inactivation. The ubiquity of the initiation factor means that any supernumerary

X-chromosomes are bound by the initiation factor and inactivated, but for every complement of the blocking factor one X-chromosome will be protected from inactivation (see Figure1.3). The *Xist* deletion of Marahrens et al. is proposed to remove the binding site for the initiation factor so that the blocking factor must default to the X-chromosome with the deleted initiation factor binding site. As a result the initiation factor binds to the X-chromosome carrying the intact *Xist* allele and initiates inactivation resulting in primary non-random inactivation.

The validity of this model has yet to be established, no molecular candidates exist for either the blocking or initiation factors, although the Marahrens et al. paper defines a candidate region which may contain the binding site of the initiation factor. The model appears capable of satisfying the observed experimental data so long as ;

- binding of both blocking and initiation factors is reversible
- initiation factor is in excess
- binding of the blocking factor reduces the affinity of the initiation factor to its target

Figure1.4 shows the probable steps involved in the inactivation process in the Marahrens *Xist* knockout mouse according to their proposed model. The excess of initiation factor and the presumed irreversibility of the inactivation process itself drives the process to non-random inactivation as described in the figure legend. By their own definition, Marahrens et al. have not necessarily identified a choosing element in their deletion as they suggest. If the interpretation seen in Figure1.4 is correct then their model can allow a choice to be made between X-chromosomes and then reversed in the direction of non-random inactivation by mass-action. These reverse steps would occur before the onset of *Xist* expression so that the

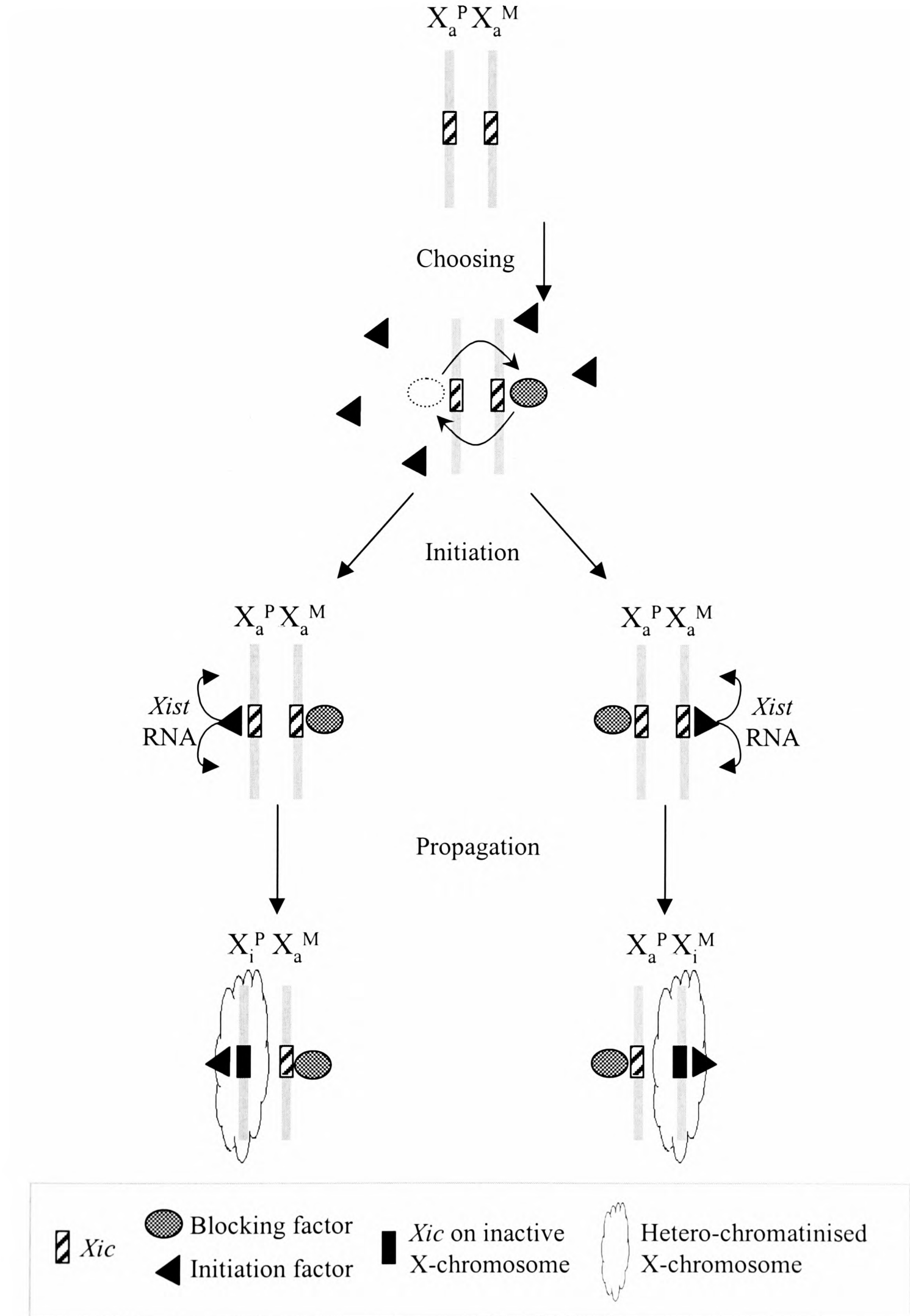


Figure 1.3: The Marahrens model for X-chromosome inactivation.
(see text for explanation)

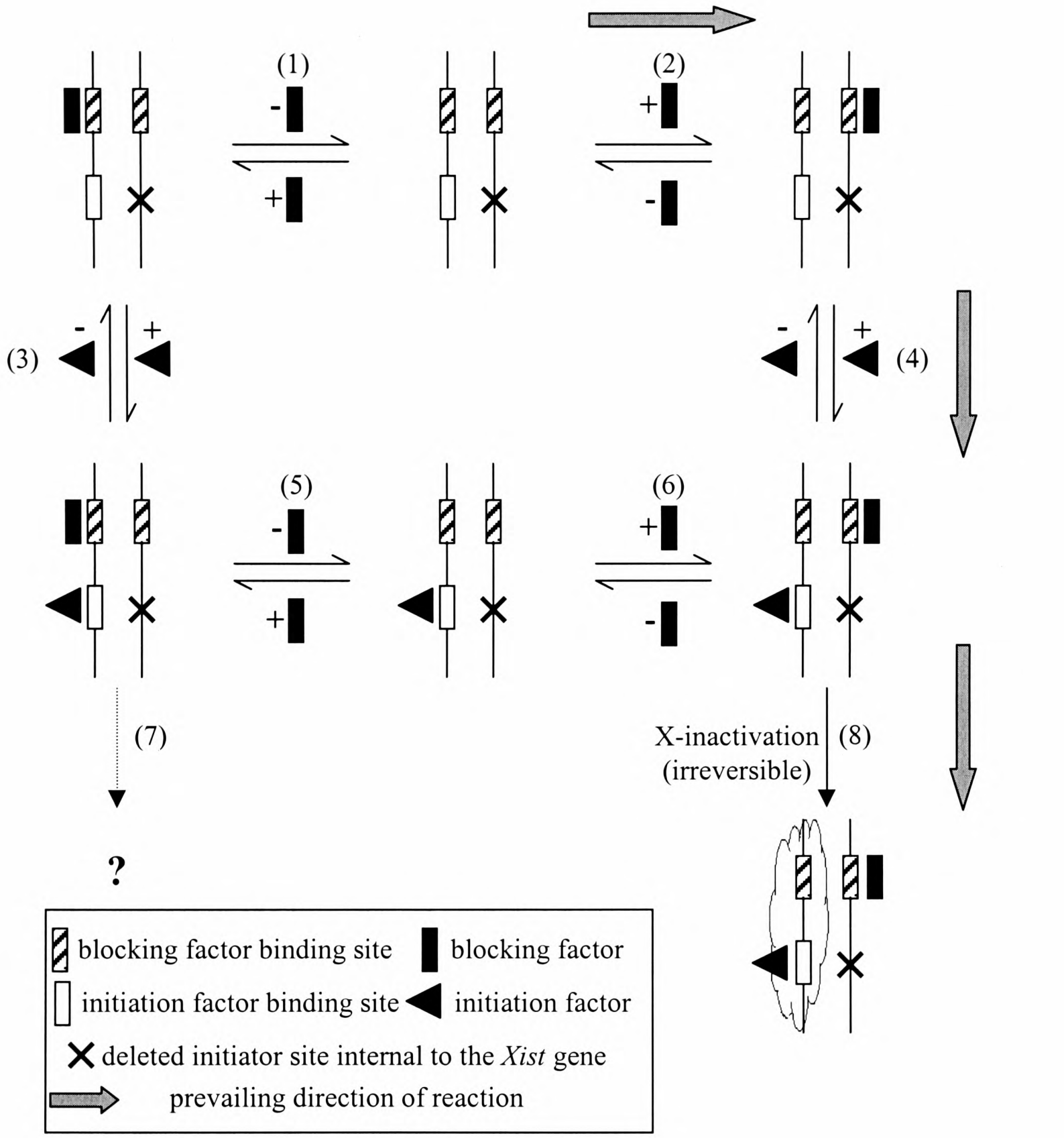


Figure 1.4: Schematic showing the putative equilibrium steps of the Marahrens model of X-chromosome inactivation.
(See text for detailed explanation). Briefly the irreversibility of the inactivation process and the proposed excess of the initiation factor would mean that the various chemical equilibria would be driven through the path as shown by the red arrow (2, 4, 8). This model is consistent with the known features of the X-inactivation process, but at present no molecular candidates exist for the blocking or initiation factors.

RNA FISH studies of Marahrens et al. would not have allowed them to conclude that a choice element had been deleted. The element they have defined would more simply be described as a binding site for an initiation factor, the choosing element is likely be that which the blocking factor binds to as originally proposed by Rastan.

1.1.7 The *Xist* gene promoter.

To understand more about the control of the *Xist* gene it was essential to define the minimal promoter and to identify regulatory elements within it. The method of choice for dissecting out the elements of a putative promoter remains the transfection of deletion construct vectors into suitable cell lines. The promoter fragment is cloned upstream of a reporter gene so that the level of expression initiated by the promoter fragment can be assayed. Three different groups used such a system to dissect out the components of the *Xist* gene minimal promoter (Hendrich et al., 1997; Huntriss et al., 1997; Sheardown et al., 1997a). These studies identified a minimal promoter of 120-400bp and described several putative transcription factor sites. Repeat elements and other sequences up to 25kb upstream of the *Xist* gene were identified as having significant effects on the level of *Xist* transcription. The studies of both Hendrich et al. (1997) and Sheardown et al. (1997a) found that the *Xist* gene was expressed at similar levels in undifferentiated and differentiated male and female ES cells and female somatic cells.

1.1.8 Imprinted *Xist* expression.

The way in which the *Xist* gene is controlled appears critical to the initiation and propagation phases of the X-inactivation process. RNA FISH studies revealed that early increases in *Xist* expression were restricted to a single *Xist* allele in each cell (Kay et al., 1993). RT-PCR exploiting polymorphisms between paternal and maternal *Xist* transcripts determined that the

increased levels of *Xist* RNA were predominantly from the paternal *Xist* allele. This imprinting effect on the *Xist* gene disappeared prior to the onset of random X-inactivation around the time of gastrulation (E4.5). Preferential paternal X-inactivation is observed in extra-embryonic tissues which are the first to undergo X-inactivation at around E3.5. This is approximately a day prior to the onset of random X-inactivation in somatic tissues and led to the suggestion that skewed inactivation in extra-embryonic tissues was the result of imprinted paternal *Xist* expression (see Figure 1.5).

Subsequent studies of the *Xist* gene promoter centred on analysis of DNA methylation at the promoter of the *Xist* gene (Ariel et al., 1995; Norris et al., 1994; Zuccotti and Monk, 1995). Methylation is the leading candidate for the epigenetic tag that imprints a gene although it remains controversial as to whether methylation is a cause or consequence of the imprinting process (see section 1.2.2, below). Norris et al. (1994) used specific restriction endonucleases to map the methylation status across the *Xist* gene promoter by exploiting the fact that certain endonucleases are unable to cut their target sites when they are methylated (see section 1.2.3, below). The studies found that in somatic cells the *Xist* promoter was hyper-methylated on the active and hypo-methylated on the inactive X-chromosomes. In extra-embryonic tissues where preferential paternal inactivation occurs the paternal *Xist* allele was hypo-methylated and the maternal hyper-methylated establishing a direct relationship between methylation and expression status of the gene. Further, methylation at the *Xist* promoter was found to undergo a developmentally programmed demethylation during spermatogenesis resulting in mature spermatozoa with hypo-methylated *Xist* alleles. This agreed with earlier work which had established that *Xist* was expressed in both human and murine testis and suggested that the paternal X-chromosome was either inactive in spermatocytes or in a primed state that predisposed the chromosome to skewed inactivation in extra-embryonic tissues (Salido et al.,

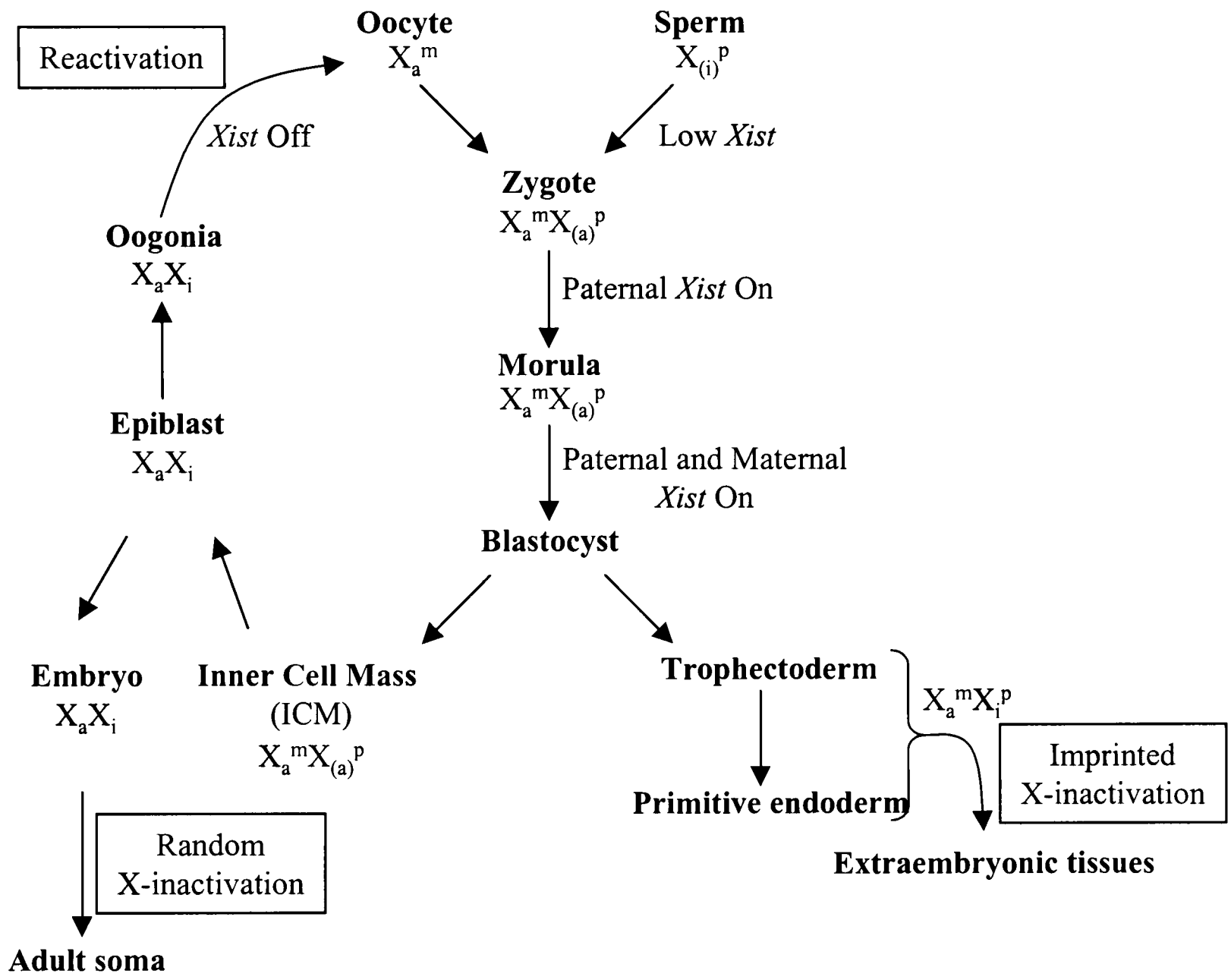


Figure 1.5: Developmental regulation of the X-inactivation process.
In the diagram X_a (active X-chromosome), X_i (inactive X-chromosome), X^m (maternally derived X-chromosome) and X^p (paternally derived X-chromosome). Developmentally *Xist* cycles through imprinted and random expression in a manner controlled by defined site-specific methylation at the promoter of the Xist gene. (see text for detailed explanation).

1992). The work of Ariel et al. (1995) and Zuccotti et al. (1995) confirmed the work of Norris et al. (1994), but with the addition that they also studied the methylation of *Xist* during oogenesis. The maternal X-chromsome was found to be hyper-methylated during maturation of the oocyte, methylation persisted in the pre-implantation embryo, but is presumably erased prior to random X-inactivation around the time of gastrulation.

1.1.9 *Xist* RNA stability.

The relative transcription rate of the *Xist* gene has been measured in different ES cells and embryos by nuclear run-on assays. No significant difference in expression was found between male ES cells and female somatic cells, yet *Xist* RNA accumulated only in differentiated female cells (Panning et al., 1997; Sheardown et al., 1997b). The reason for this accumulation of *Xist* RNA was proposed by both groups to be due to stabilisation of the RNA either by the binding of a stabilising factor or removal of a destabilising factor. Both groups also proposed the possibility that a switch of expression to a different *Xist* promoter may result in the generation of an alternate stable transcript, but no such promoter has been described.

The series of events initiating the X-inactivation process are complex, they involve imprinted *Xist* expression prior to the onset of random X-inactivation, stabilisation of *Xist* RNA at one of the two X-chromosomes and suppression of low level expression at the other. Propagation of the stabilised *Xist* RNA in a bi-directional manner *in-cis* has to be facilitated apparently without contacting the DNA of the X-chromosome (Clemson et al., 1996). In addition the current models proposed to explain initiation of the X-inactivation process require many factors which have yet to be isolated, these include, blocking, initiation and stabilisation/destabilisation factors. The issue is further complicated by a recent report of

anti-sense transcripts across the *Xist* locus (Lee et al., 1999). The transcript, *Tsix*, associates with both X-chromosomes prior to inactivation, but then associates specifically with the active X-chromosome until the *Xist* gene is turned off. The biological relevance of this putative gene remains unclear, but it threatens to add yet another layer of complexity to the X-inactivation process.

1.1.10 Genes escaping X-inactivation.

X-inactivation results in the silencing of thousands of genes on the inactive X-chromosome, but it does not inactivate every gene. The first evidence to suggest that some genes escaped the X-inactivation process in humans came from the presence of a disease phenotype (Turner syndrome) in females with a single X-chromosome (Ford et al., 1959). The only genetic difference between Turner and normal females was the inactive X-chromosome, it was therefore expected that some genes were normally expressed from the inactive X-chromosome, escaping inactivation. The first gene demonstrated to escape X-inactivation was the XG blood group gene (Fialkow, 1970) which has since been mapped to the boundary of a pseudo-autosomal region (PAR) of the X-chromosome.

There are two main PARs on the human X-chromosome one of approximately 2.5Mbp (X_p/Y_p or PAR1) and the other of 320kb (X_q/Y_q or PAR2) (Rappold, 1993). The PARs represent excellent candidate regions for genes involved in the pathogenesis of Turner syndrome as they are not subject to inactivation so the XO phenotype would result in haplo-insufficiency of PAR genes. Recent work has, however, complicated this assumption because it is now clear that other genes not within PARs can escape the inactivation process, thereby increasing the number of candidates for Turner (Disteche, 1995; Distesch, 1997).

To date in excess of 20 genes have been shown to escape X-inactivation in humans (Brown et al., 1997; Disteche, 1995; Distesch, 1997) with one estimate suggesting that up to 25% of X-linked genes could escape the inactivation process (Brown et al., 1997). The majority of escaping genes are clustered within the two pseudo-autosomal regions or within Xp11 (Miller and Willard, 1995) although genes on the long arm have been shown to escape such as RPS4X (Xq13) (Fisher et al., 1990). A number of systems have been established to determine whether a gene is subject to inactivation in man and mouse (see sections 5.1.2 and 5.1.3). To date only *SMCX/Smcx* (Agulnik et al., 1994; Wu et al., 1994), *EIF2S3X/Eif2s3x* (Ehrmann et al., 1998) and *UTX/Utx* (Greenfield et al., 1998), have been shown to escape X-inactivation in man and mouse suggesting that there are important differences in the X-inactivation process between mammals.

Studies of genes escaping inactivation are of great importance not just because they may help us to understand Turner syndrome, but because they also necessitate a re-evaluation of the clinical management of X-linked diseases. In addition the apparent clustering of escaping genes has important implications for the mechanism of propagation and stability of the inactivation process itself. The relationship between nuclear architecture, *Xist* RNA, chromatin and DNA methylation must ultimately be directed from the level of DNA sequence. Differences between inactivated and escaping sections of DNA could prove to be the key that unlocks the mechanism of the inactivation process.

1.1.11 Chromatin and the inactive X-chromosome.

The nature of "sex-chromatin" has been the focus of research since it was first described by Barr and Bertram in 1949. The inactive X-chromosome adopts a condensed structure occupying the periphery of the nucleus of female cells popularly called the "Barr-body".

Several groups have used FISH techniques with X-chromosome specific probes to determine the relative positions of different parts of the active and inactive X-chromosomes. One of the first such studies concluded that the inactive X-chromosome formed a looped structure with telomeres closely opposed in contrast to the more linear active X-chromosome (Walker et al., 1991). A semi-quantitative study by another group used FISH and serial sectioning with a laser scanning confocal microscope to calculate the volume and territory surface area of the active and inactive X-chromosomes (Eils et al., 1996). A statistical analysis between the active and inactive X-chromosomes and autosomal controls showed that the active and inactive X-chromosomes occupied similar volumes but were very differently shaped. The active X-chromosome was found to be flat and rough whereas the inactive was rounder and smoother. Interestingly, the active X-chromosome topology was found to be indistinguishable from that of a similar sized autosome (chromosome 7).

Structural analysis of the active and inactive X-chromosome has recently been directed to the nature of their chromatin components. Chromatin is made up of histones and non-histone chromosomal proteins (NHCPs). Two molecules each of the core histone proteins H2A, H2B, H3 and H4 form to make a histone octamer (Eickbusch and Moudrianakis, 1978) around which >160bp of DNA wraps to form a nucleosome (Luger et al., 1997). A further histone, H1, is involved in linking nucleosomes together to form higher order chromatin structure (Widom, 1989). Histone proteins can be subject to a variety of covalent modifications which have been implicated in modifying their structural and functional roles, these include phosphorylation, ubiquitination, methylation, ADP-ribosylation and acetylation reviewed in (Wolffe and Hayes, 1999).

In a landmark paper Jeppersen and Turner described a striking lack of acetylated histone-H4 on the inactive X-chromosome (Jeppesen and Turner, 1993). Human and mouse metaphase chromosomes were immunolabeled with antibodies raised to specific acetylated isoforms of histones. All chromosomes were labelled with the antibody in bands corresponding to gene-rich regions (R-band) except for the inactive X-chromosome which appeared very lightly labelled. Highly acetylated regions of the genome are associated with actively transcribing or potentially transcribable chromatin (Hebbes et al., 1992) so the apparent lack of acetylation on the inactive X-chromosome correlated to its known genetic inactivity. When cells were cultured in the presence of a deacetylase inhibitor (sodium butyrate) three small bands of hyper-acetylated histone H4 appeared on the inactive X-chromosome corresponding to regions of the X-chromosome that escape inactivation (pter-22.2, p11.3-11.2 and q13).

Further light has been cast onto the relationship between histone modification, *Xist* RNA and X-inactivation in a recent paper (Duthie et al., 1999). In this study, RNA FISH was used to detect the *Xist* transcript in a number of primary fibroblast cultures from a variety of rodent sources. In particular *Xist* distribution was analysed across the inactive X-chromosome of two species of the common vole *Microtus transcaspicus* and *M. rossiaemeridionalis*. The vole inactive X-chromosome has a strikingly different structure to that of man and mouse in that approximately 50% of the chromosome is composed of large blocks of constitutive heterochromatin. In over 60% of vole fibroblast metaphases studied *Xist* RNA was distributed in a banded pattern and was absent from the large block of heterochromatin around the centromere. Further study in other rodent species revealed a strong correlation between late-replicating (Giemsa-dark) bands and the absence of *Xist* RNA.

In order to further investigate this novel finding, the group went on to study the distribution of *Xist* RNA on the inactive X-chromosome of fibroblasts derived from two strains of mice with different X-autosome translocations. Antibodies to acetylated histone-H4 were used to identify areas of the chromosome as actively transcribing or transcription permissive and this was correlated to the positions of autosomal material and *Xist* RNA. The study showed that *Xist* RNA did not spread well into the autosomal material and that regions of histone-H4 acetylation were predominantly autosomal and not coincident with the position of *Xist* RNA.

These findings are consistent with a model for the spreading of the X-inactivation signal first proposed by Riggs (1985). In the model nucleating sites for the inactivation mechanism are spread along the X-chromosome and seed heterochromatin formation. In their paper Duthie et al. (1999) propose a detailed model based on their findings and the Riggs model. *Xist* RNA is proposed to bind the *Xic* at high affinity sites resulting in local changes in chromatin conformation. These high affinity sites are reduced in number or absent in regions escaping inactivation and Giemsa-dark bands. Other sites in Giemsa-light bands recruit the *Xist* RNA and other proteins and start to form heterochromatin which spreads and co-operatively interacts with other such nucleation sites. The inactivating chromosomal regions are then proposed to loop out and localise to the periphery of the nucleus away from the transcriptional machinery. Pivotal to the validation of such a model is the isolation of the high affinity binding sites. It has been suggested that these sites could be LINE elements in which the X-chromosome is particularly rich (Lyon, 1998). Advances in the understanding of chromatin structure and its regulation have provided new avenues for investigating the X-inactivation process and are likely to facilitate the next step in the elucidation of its mechanism.

1.1.12 DNA Methylation and the X-chromosome.

Methylation was first proposed as a direct candidate to mediate the X-inactivation process by Riggs and Holliday twenty four years ago (Holliday and Pugh, 1975; Riggs, 1975).

The X-inactivation process has a number of features that could potentially be explained by functions attributed to DNA methylation. Firstly, X-inactivation is clonal; the same X-chromosome is found to be inactive in all daughter cells derived from the same parental cell. This requires an epigenetic mark to be present on the X-chromosome that instructs the daughter cells to inactivate it and not the other X-chromosome. Methylation could satisfy this criterion because methylation patterns are clonally inherited through mitosis presumably as a result of semi-conservative replication and the action of the maintenance DNA methyltransferase (mDNMT) (Pfeifer et al., 1990; Wigler et al., 1981) (see section 1.2.1, below). Secondly, gene inactivation on the X-chromosome has to be achieved across an entire chromosome, this would require a ubiquitous transcription suppression system. DNA methylation at the promoters of genes has been correlated to the inactivity of genes *in-cis* (Bird, 1986) and is ubiquitous throughout the genome (Solage and Cedar, 1978). Thirdly X-inactivation has to be reversible through gametogenesis to allow the expression of genes from the male X-chromosome and allow random inactivation to occur during the development of female somatic tissues. Methylation is reversible, there are both ill-defined de-methylating activities in the cell as well as a mode of passive demethylation whereby removal of mDNMT results in sequential removal of methylation through DNA replication (Weiss et al., 1996).

Initial characterisations of methylation on the active (Xa) and inactive (Xi) X-chromosomes were carried out using methylation sensitive restriction enzymes (Wolf et al., 1984). These

analyses were performed at the promoters of the *HPRT* (Lock et al., 1987) and *PGK* (Keith et al., 1986) genes which have large CpG-islands at their 5′-ends rich in the CpG sites that are the main target of genomic DNA methylation (Woodcock et al., 1987). These early studies found that the promoters on the Xi were hyper-methylated and those on the Xa were hypo-methylated. More recent methylation analysis using similar techniques at a number of other X-linked loci have supported these original findings (Norris et al., 1991; Tribioli et al., 1992).

Semi-quantitative measurement of global methylation across both the Xi and Xa using nick-translation (Adolph and Hameister, 1990) and anti methyl-cytosine antibodies (Bernardino et al., 1996) found that the methylation content of the chromosomes was identical. Although other studies disagree with these findings (Prantera and Ferraro, 1990; Viegas-Pequignot et al., 1988).

The non-methylatable cytosine analogues 5-aza-2′-deoxycytidine (5aCdr) and 5-aza-cytidine (5aC) have been used by several groups to reactivate inactive X-chromosomes in tissue culture systems (Graves, 1983; Hansen and Gartler, 1990; Mohandas et al., 1981). This reactivation of the Xi by demethylation is suggestive of a critical role for methylation in the stability of the inactive X-chromosome. This is supported by experiments in marsupials in which instability of the inactive X-chromosome has been correlated to the absence of differential methylation at the promoters of X-linked genes (Cooper et al., 1993; Loebel and Johnston, 1996).

Recent techniques including bisulphite (Frommer et al., 1992) and ligation mediated PCR sequencing (Pfeifer et al., 1989) (see section 1.2.3, below) have allowed the dissection of methylation profiles at the nucleotide level. Several groups have used these techniques to

profile methylation at the promoters of genes on the active and inactive X-chromosomes (Hornstra and Yang, 1994; Park and Chapman, 1994; Pfeifer et al., 1990; Stoger et al., 1997). These studies further support the findings that gene promoters are hyper-methylated on the Xi and hypo-methylation on the Xa.

The promoters of genes escaping the X-inactivation process are commonly subject to methylation analysis as their promoters and would be predicted to be unmethylated on both active and inactive X-chromosomes. Methylation analysis has been carried out on the promoters of the *MIC2* (Goodfellow et al., 1988), *ANT3* (Slim et al., 1993), *STS* (Li et al., 1996), *UBE1* and *PCTK1* (Carrel et al., 1996) genes and also at a transgene that escapes inactivation (Goldman et al., 1998). All of these studies report hypo-methylation at the gene promoter on both the active and inactive X-chromosomes, consistent with the concept that promoter methylation is an implicit part of the X-inactivation process.

The results of the studies outlined above are consistent with a role for DNA methylation in the initiation of inactivation and maintenance of X-inactivation. However, there has been much debate as to whether methylation is actually required for X-inactivation to take place. An early study of methylation at the promoter of the *Hprt* gene found that methylation of the promoter did not occur prior to inactivation of the gene (Lock et al., 1987). Whereas studies of the *G6pd* and *Pgk1* genes found methylation at a time coincident with inactivation (Grant et al., 1992). It is important to note that neither study used techniques which looked at large numbers of methylatable sites in the promoters and as such would be less likely to identify low levels of methylation were they to exist.

One of the most important studies to look at the role of methylation in the X-inactivation process was that of Panning et al. (1996) in which *Xist* expression and the expression status of two X-linked genes was studied in male mice and ES cells in which the *Dnmt* gene was knocked-out (Panning and Jaenisch, 1996). The study found that upon differentiation of *Dnmt*(-/-) male ES cells and during early development of male *Dnmt*(-/-) embryos the *Xist* gene was up-regulated from and associated with the single X-chromosome. The *Xist* RNA expression was found to be indistinguishable from that of wild-type female *Dnmt*(+/+) ES cells and embryos although only a small proportion of the cells in the male embryo showed increased *Xist* expression. Analysis of *Pgk1* and *Mecp2* (genes normally subject to X-inactivation) expression in the *Dnmt*(-/-) male ES cells and embryos suggested that they were inactivated on the single X-chromosome. These findings were consistent with *Xist* expression being sufficient for X-inactivation in differentiating cells and tissues in the absence of DNA methylation.

In summary, DNA methylation is critical for the control of *Xist* expression in embryogenesis and for the maintenance of differential *Xist* expression between active and inactive X-chromosomes in somatic tissues. Methylation is also essential for the maintenance of the inactive X-chromosome and the clonal inheritance of inactive status to daughter cells through mitosis. Evidence is contradictory as to whether DNA methylation is involved in the local control of X-linked genes and the propagation of the inactivating signal. Knock-out experiments of the *Dnmt* gene suggest that methylation is not required for the propagation of inactivation, however these experiments look at the activity of only two X-linked genes. Furthermore, knocking out the *Dnmt* gene has consequences for gene expression across the entire genome resulting in massive cell-death both *in-vitro* and *in-vivo*. The perturbation caused by the *Dnmt* knockout may well be too severe to allow an interpretation that has any

bearing on the biology *in-vivo* especially when strong cell selection is occurring. Experiments to determine the precise timing of methylation at the promoters of genes on the inactivating X-chromosome and their expression status *in-vivo* will more subtly test this question.

1.2 Genomic DNA Methylation.

Historically DNA has been considered to be made up of four bases, adenine, thymine, guanine and cytosine. In reality many prokaryotes and eukaryotes have additional bases which are generally chemically modified forms of the original four. The only modified base present in the mammalian genome is 5-methylcytosine which is formed by the covalent addition of a methyl group to the 5-carbon atom of the pyrimidine ring. An estimated 30 million modified cytosines are thought to be present in the diploid genome (Bestor and Tycko, 1996) comprising approximately 5% of the total cytosine content (Ehrlich et al., 1982). DNA methylation has long been postulated to be involved in the control of gene expression (Holliday and Pugh, 1975; Riggs, 1975) but only in recent years have the molecular mechanisms by which this is achieved been established.

1.2.1 Distribution of methylation in the mammalian genome through development.

Methylation occurs predominantly, but not exclusively at CpG-dinucleotides (Woodcock et al., 1987) which are distributed differentially throughout the genome. In most regions of the genome they are present at a concentration of about 1% and are almost totally methylated (Sanford et al., 1984), but in GC-rich isochores their concentration is about 5% and they are almost totally unmethylated (Bernardi, 1989). The low density CpG regions of the genome generally contain repetitive elements, satellite DNA and tissue-specific genes whereas the high density regions contain CpG-islands and housekeeping genes (Bird, 1986).

The precise definition of a CpG-island is a region of DNA greater than 200bp in length with a (G+C) content greater than 50% and a CpG-dinucleotide relative abundance greater than 0.6 for its entire length (Gardiner-Garden and Frommer, 1987). CpG-islands are associated with the 5′-end of most housekeeping genes and are normally unmethylated with the exception of those of the inactive X-chromosome (see section 1.1.12, above). Up to 40% of tissue specific genes have CpG-islands at their 5′-ends which are thought to be predominantly methylated in cells and tissues in which the gene is not expressed (Yievin and Razin, 1993).

Throughout development methylation patterns change dynamically (Jaenisch, 1997) (see Figure 1.6). In the pre-implantation embryo there is a dramatic reduction in genomic methylation which is thought to reset gene expression so that early embryonic cells are totipotent (Monk et al., 1987; Rougier et al., 1998). Consistent with this idea is the observation that subsequent differentiation of embryonic cells in the post-gastrulation embryo coincides with genome-wide *de-novo* methylation (Jahner and Jaenisch, 1984). Tissue specific genes are subsequently found to undergo demethylation in cells and tissues in which they are to be expressed (Razin and Cedar, 1991). Once differentiation is complete, methylation profiles are preserved by the action of the maintenance DNA methyltransferase (mDNMT) which has a strong preference for hemi-methylated DNA (Wigler et al., 1981) and is targeted to sites of DNA replication (Leonhardt et al., 1992).

1.2.2 Methylation and the control of gene expression.

Numerous studies both *in-vitro* and *in-vivo* have correlated the methylation of gene promoters and flanking regions with lack of expression, but it has proven almost impossible to formally demonstrate that methylation itself is the cause of the inactivity rather than the consequence. The scientific community is divided as to the role of methylation in the genome between

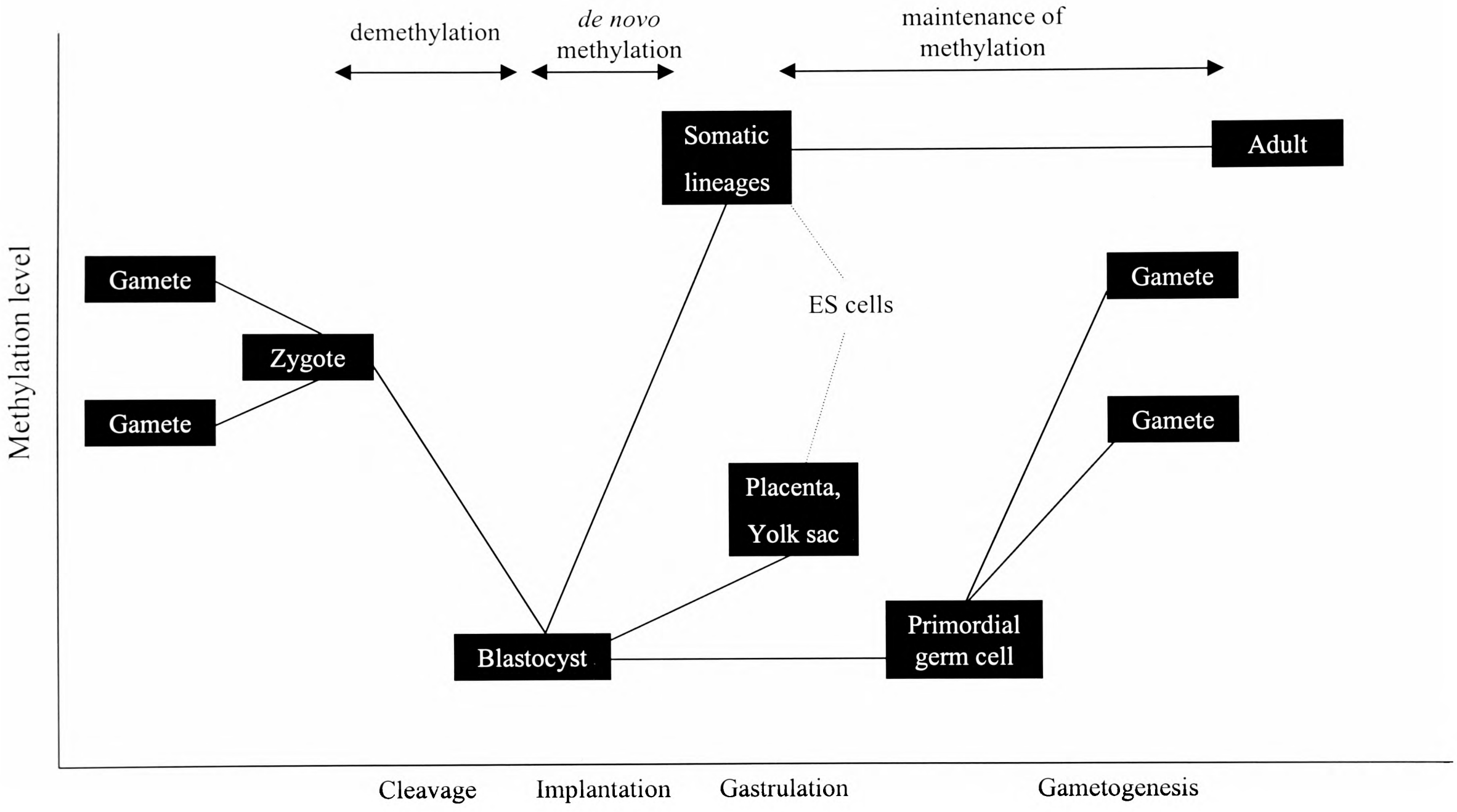


Figure 1.6: Dynamic methylation changes during murine development (after Jaenisch, R. 1997).

groups that believe methylation directly facilitates the control of tissue-specific gene expression (Beard et al., 1995; Li et al., 1993; Panning and Jaenisch, 1996; Yisraeli et al., 1988) and those that believe it does not but is primarily a mechanism designed to control expression of foreign transposons (Bestor and Tycko, 1996; Yoder et al., 1997).

Theoretically, methylation can repress transcription in a site specific or site non-specific manner (Kass et al., 1997). Site specific methylation can act directly to block the binding of transcription factors (Clark et al., 1997; Comb and Goodman, 1990; Gaston and Fried, 1995; Iguchi-Ariga and Schaffner, 1989; Radtke et al., 1996; Watt and Molloy, 1988) or indirectly by recruiting other proteins that block their access (Nan et al., 1998; Tate and Bird, 1993). These effects are essentially local in that a particular methylation profile at the promoter or regulatory element of a gene directs repression *in-cis*. The other possibility is that non-specific site methylation marks a region for repression so that the overall methylation density is critical for repression and not the precise methylation profile (Hsieh, 1994). It appears likely that both site specific and non-specific processes are responsible for methylation directed transcriptional repression.

The relationship between DNA methylation, gene expression and chromatin has been inferred for over twenty years, but until recently no direct causal relationship had been described to connect them. Three recent papers have now described this relationship in mechanistic detail revolutionising our understanding of the way in which methylation regulates gene expression (Eden et al., 1998; Jones et al., 1998; Nan et al., 1998). The mechanism is based on the protein MeCP2 which binds to methylated DNA in the genome in a sequence independent manner through a methyl-DNA binding domain (MBD) (Meehan et al., 1992). MeCP2 was thought to be involved in the control of gene expression through methylation as a

transcription repressor domain (TRD) had been identified, but it was not known how this mediated repression (Nan et al., 1997). The TRD has recently been shown to interact directly with the corepressor protein mSin3A and immunoprecipitation experiments with an anti-MeCP2 antibody precipitate mSin3A, HDAC1 and HDAC2 (histone deactelylases 1 and 2) (Nan et al., 1998). This putative repressor complex is remarkably similar to the multiprotein complexes described by Wolffe, et al. which achieve repression of gene expression through deacetylation of histones H3 and H4 in core nucleosomes (Wolffe, 1997).

Two further experiments formally demonstrated that methylation could repress gene expression. In the first, *in-vitro* methylated and unmethylated expression constructs were introduced into *Xenopus laevis* oocytes. Repression was only observed with the methylated construct and was relieved by treating the cells with the deacetylase inhibitor trichostatinA. This demonstrated that histone deacetylase activity was required for repression mediated through methylation. Immunoprecipitation of nuclear protein with anti-MeCP2 antibodies precipitated MeCP2, mSin3A and deacetylase activity establishing a direct link between methylation, the methyl DNA binding repressor complex and histone deacetylation (Jones et al., 1998).

In the second experiment cells were stably transfected with *in-vitro* methylated and unmethylated thymidine kinase (*tk*) expression constructs. The highly acetylated nucleosomal fraction was isolated from cell nuclei by co-immunoprecipitation with anti-acetylated histone antibodies and the presence of the *tk* gene determined by PCR. The *tk* construct was found preferentially in the highly acetylated fraction when isolated from the cells transfected with the unmethylated construct. Treatment of cells transfected with the methylated construct with

trichostatinA shifted the *tk* construct into the highly acetylated fraction and increased the transcription rate of the *tk* gene (Eden et al., 1998).

Together these data now formally demonstrate that methylation is capable of directing the repression of gene transcription through the recruitment of the methyl binding protein MeCP2 and its interaction with the evolutionarily conserved Sin3A/HDAC1/HDAC2 repressor multiprotein complex. The mammalian inactive X-chromosome is a striking example of this interaction in-vivo in which the chromosome is heavily methylated and strikingly devoid of hyper-acetylated histones (Jeppesen and Turner, 1993).

1.2.3 Methods of analysing genomic DNA methylation.

i) Sequence specific restriction enzyme analysis.

Certain restriction enzymes cleave their target DNA sequences at lower rates when they are methylated (McClelland et al., 1994) and have been particularly useful in the analysis of genomic DNA methylation.

Typically, genomic DNA is digested with a restriction enzyme flanking the region of interest then further digested with one or more methylation sensitive restriction enzymes. Restriction fragments are then fractionated on agarose or acrylamide gels, blotted onto nylon filters and hybridised to a probe spanning the region to detect all possible fragments (Saluz and Jost, 1993). The pattern of hybridised bands allows the elucidation of the methylation status of sensitive sites within the region covered by the probe.

This method of analysing methylation suffers from a number of technical problems and absolute limitations. One of the most restricting features is that only a limited number of

potentially methylated sites can be analysed due to the frequency of cleavage of restriction enzymes. Even the use of enzymes which cut more frequently than average and are methylation sensitive such as HpaII and HhaI can prove complicated as their sites tend to be clustered in CpG-rich areas so that cleavage fragments are difficult to resolve. A further limitation is that in order to achieve sensitive detection of small fragments from genomic DNA digests large quantities of DNA are required (2-10μg/lane) which greatly limits the potential sources of DNA. Any approach that relies on enzymatic cleavage of DNA is prone to incomplete digestion. In the case of methylation analysis this can be particularly problematic as partial digests are misinterpreted as partial or mosaic methylation and are difficult to diagnose as no positive controls for total digestion are available.

With the advent of higher resolution techniques (below) better methods for methylation analysis are available, but southern blotting remains an invaluable tool to corroborate this kind of data when suitable probes are available.

ii) Maxam and Gilbert sequencing

The first method to analyse genomic methylation at the nucleotide level was a by-product of the Maxam-Gilbert DNA sequencing procedure (Church and Gilbert, 1984). This sequencing technique relied on chemical base-specific cleavage to generate sequencing ladder fragments. The chemical hydrazine was used to cleave fragments at cytosine residues, but this was prevented when the cytosine was methylated so that a gap occurred in the sequencing ladder. A number of modifications were made to the original protocol to increase the ease of detection of methylated cytosines (Grigg and Clark, 1994), but it remained problematic to identify the gaps in sequencing autoradiographs. In addition methylation genomic sequencing

required large quantities of DNA which like blotting based approaches (section 2.4(i), above) significantly restricted its application.

iii) Methylation sensitive PCR

The advent of the polymerase chain reaction (PCR) revolutionised molecular biology and provided an opportunity to develop new sensitive methods to determine genomic methylation patterns. One of the earliest methods to exploit this used methylation sensitive restriction enzymes to assess the status of sites within a PCR target region. Genomic DNA would be completely digested with the methylation sensitive restriction enzyme, then PCR would be performed using the digest as template. If the sites within the target were methylated the restriction enzyme would be unable to cleave the sites and a product would be formed. Conversely if one or more of the sites were unmethylated then they would be cleaved by the enzyme and no product would be amplified by PCR. This technique has been used successfully on genomic DNA isolated from as little as 30 cells (Singer-Sam et al., 1990), but it still suffers from the limitations associated with restriction analysis. Incomplete cleavage by the restriction enzyme produces false positive amplification products which again cannot be controlled for. The choice of sites at which methylation can be analysed is limited by the frequency of cleavage sites within the region of interest. Nevertheless this technique is easy to perform, it is rapid and sensitive and often provides a good entry point for methylation analysis.

iv) Ligation mediated PCR (LMPCR)

One of the main limitations of using hydrazine sequencing to detect genomic methylation patterns was the large amount of genomic DNA needed. The advent of PCR led to the

development of a modified procedure for hydrazine sequencing called LMPCR (Pfeifer et al., 1989; Pfeifer et al., 1990) (see Figure 1.7). In this method chemical cleavage by hydrazine is performed then the DNA is purified, a gene specific primer (GSP) annealed and primer extension performed to the cleavage site using DNA polymerase to create a blunt end. A linker fragment is then ligated to this blunt end and the GSP and linker primer are used to amplify the fragment. Essentially this procedure amplifies the entire sequencing ladder so that small amounts of starting DNA can be used. Typically the amplification products are fractionated on a standard sequencing gel, blotted onto nylon filters and then hybridised to a radioactively labelled probe for the target.

The LMPCR technique still suffers from the fact that it is a negative display procedure, i.e. absence of a band denotes a methylated cytosine. The technique is not widely used, but there is potential to develop new LMPCR procedures using fluorescent labelled primers so that an entire sequencing run can be performed in one lane (exactly as for normal dye-primer fluorescent sequencing). The added benefit would be that the blotting and hybridisation steps would be removed which would save considerable time and obviate the need for using radioactivity. The greatly improved signal intensity of modern fluorescent dyes would also mean that the drop-out indicative of a methylated cytosine should be easily observed.

v) Bisulphite sequencing (Frommer genomic sequencing)

In excess of one hundred publications have used the bisulphite sequencing procedure developed by Marianne Frommer (Frommer et al., 1992) to analyse methylation making it the most widely used method of genomic methylation sequencing.

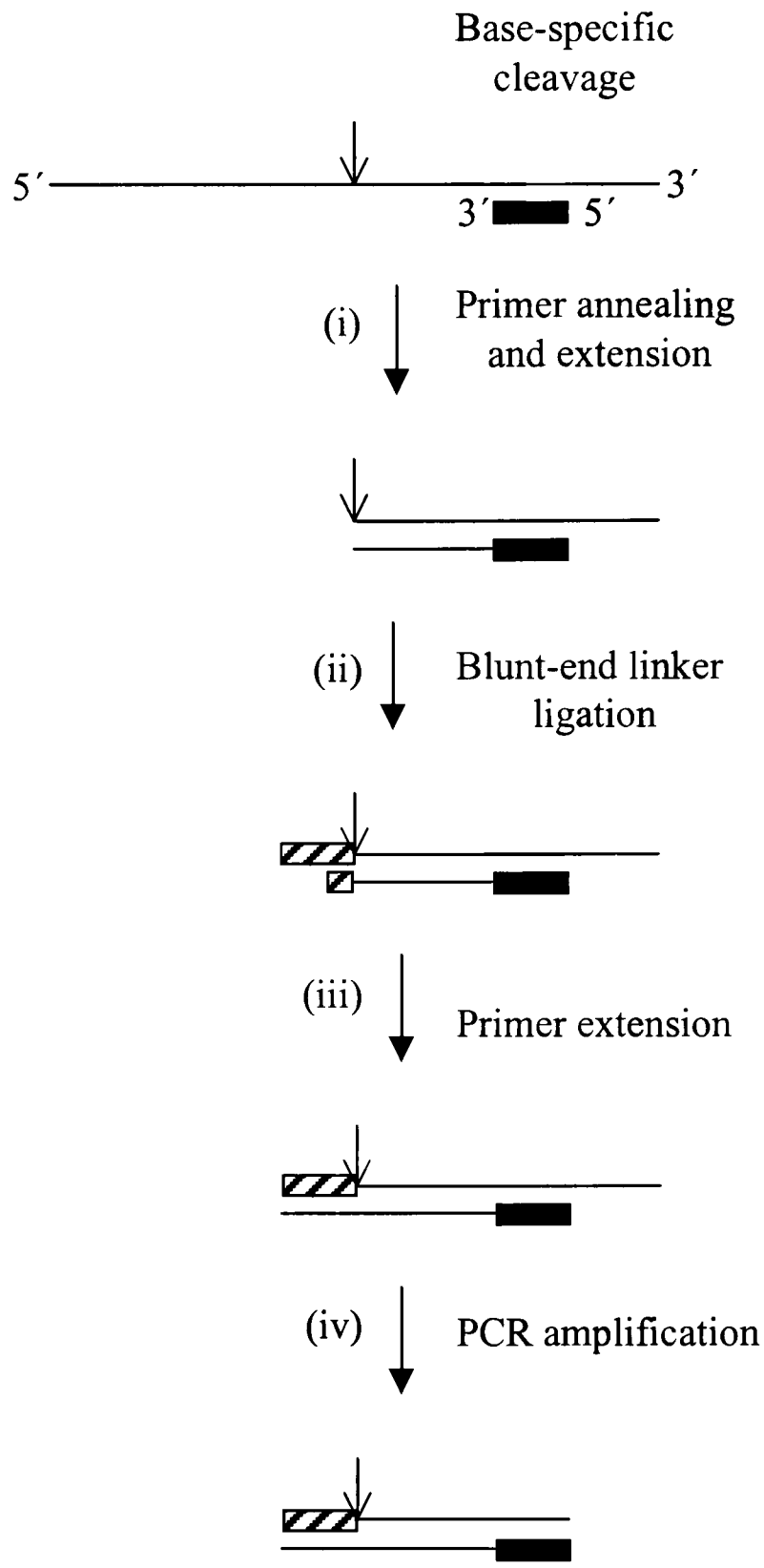


Figure 1.7: Ligation Mediated PCR (LMPCR) methylation sequencing.
LMPCR genomic methylation sequencing exploits the inability of hydrazine to cleave DNA at methylated cytosine residues. Sequencing is essentially performed exactly as described by Church and Gilbert (1984). Once the chemical cleavage has been performed a gene specific sequencing primer is annealed to the template and primer extension performed to the site of cleavage (i). A small linker fragment is ligated to the blunted end and again primer extension performed (iii) followed by PCR with linker and gene-specific primers (iv). Once the PCR amplification is complete the fragments are fractionated on a normal sequencing gel then blotted onto a nylon filter and hybridised to a labelled gene specific probe. The resulting sequencing ladder is identical to that for a normal chemical sequencing one, but with the exception that gaps appear in the cytosine cleavage lane where a methylated cytosine was present in the original template. Due to the PCR amplification of the sequencing ladder it is possible to obtain methylation profiles from small quantities of genomic DNA without the need for sub-cloning.

Under suitable conditions the bisulphite ion can covalently sulphonate the pyrimidine ring of cytosine and catalyse its oxidative deamination to uracil (see Figure 1.8). Cytosine residues that are methylated on the 5-carbon of the ring are not susceptible to bisulphite making the treatment an extremely efficient way of differentiating methylated cytosines from non-methylated ones. This chemical specificity was exploited in the derivation of the bisulphite sequencing procedure.

Typically, denatured genomic DNA is treated with sodium bisulphite and then subject to PCR amplification using gene-specific primers (see Figure 1.9). During the amplification step the uracil residues produced by bisulphite treatment of unmethylated cytosines are converted to thymines whereas methylated cytosines remain as cytosines. The bisulphite PCR products are cloned and then sequenced either individually or in pools (Paul and Clark, 1996) to build up a profile of the methylation status of the target region. Every cytosine band seen in the sequencing ladder represents a methylated cytosine in the original genomic DNA. This makes bisulphite sequencing a positive display procedure that assesses the status of every cytosine. In addition the PCR amplification step allows determination of methylation sequences from as little as 10ng of starting genomic DNA making it possible to assess methylation of DNA from archival material and very small tissue samples.

Primer design is critical to the success of bisulphite sequencing as the amplification step has the potential to generate a skewed representation of the template and consequently of the methylation profile (Clark, 1997; Voss et al., 1998). In order to overcome this potential, primers are designed to areas of the treated target that do not contain methylatable sites so that primer annealing will be identical between all possible template molecules (Woodcock, D. M. personal communication to the DMS). After bisulphite treatment the two-strands of

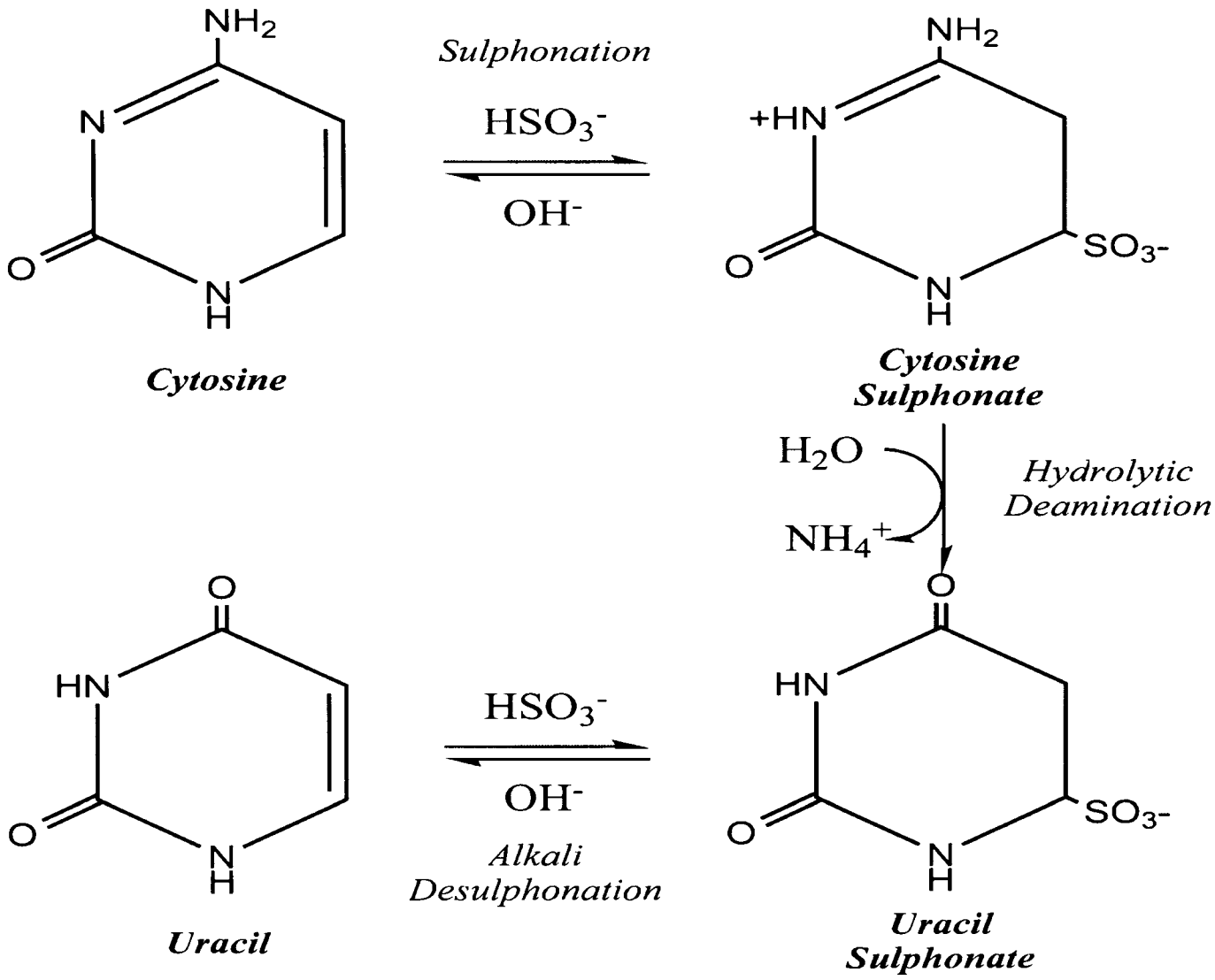


Figure 1.8: Chemistry of the bisulphite treatment of DNA.
The bisulphite treatment of DNA is efficiently achieved by carefully controlling the pH of the various steps of the process. In a typical experiment genomic DNA is vigorously denatured by alkali treatment to ensure that extremely GC-rich segments do not resist conversion. Following neutralisation the sulphonation step of the bisulphite conversion is carried out under acid conditions using one of a number of time/temperature regimes (see Materials and Methods 2.47). Spontaneous deamination occurs readily, but is driven through concomitant with the de-sulphonation step under alkaline conditions. The result is a DNA template that is greatly degraded due to the very stringent and damaging conditions that it has been subjected to. Yields from the bisulphite conversion can be very low, but improvements to the original protocol have greatly reduced damage during the treatment process. Figure 1.9 details the subsequent PCR amplification, sub-cloning and sequencing steps to acquire the genomic methylation sequence.

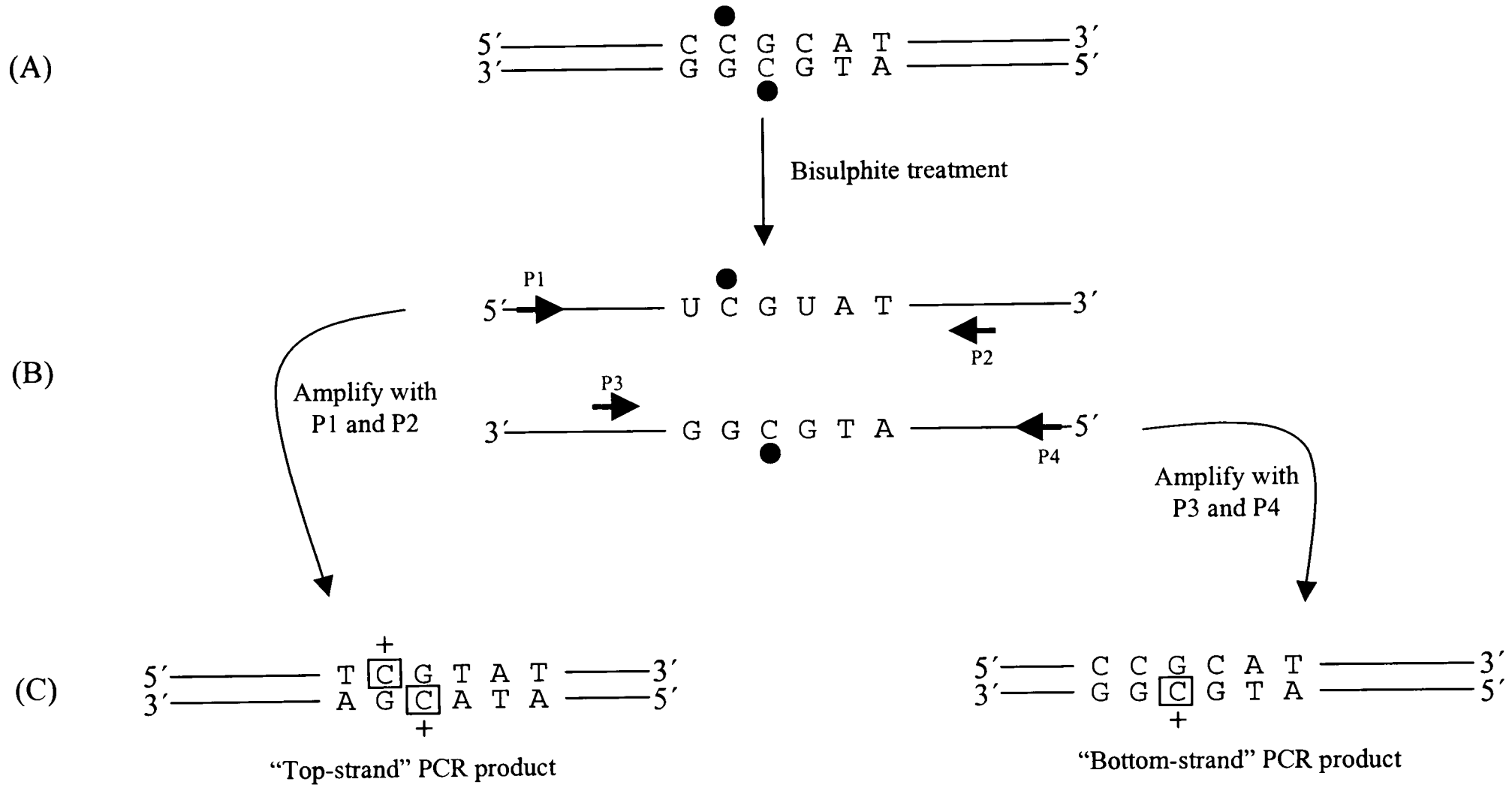


Figure 1.9: Bisulphite genomic methylation sequencing ("Frommer"-sequencing).
Genomic DNA (A) is treated with bisulphite as described in the text and figure1.8 then amplified with primers that are designed to specifically amplify bisulphite treated DNA, but not to differentiate between templates which differed by methylation status (see text for explanation). After bisulphite treatment the two-complementary strands of DNA become non-complementary (B) so that primers can be designed to characterise methylation on either strand thereby allowing the identification of asymmetric and hemi-methylated sites. After amplification products can be sequenced directly or sub-cloned and sequenced. Each cytosine residue present in the "C-lane" of the sequencing gel represents a methylated cytosine in the original genomic DNA top or bottom-strand when suitable sequencing primers are selected (C).

DNA become non-complementary (see Figure 1.9) so it is possible to design primers to either the top or bottom strand allowing the identification of asymmetric (Woodcock et al., 1997) and hemi-methylated (Millar et al., 1998) sites.

Bisulphite sequencing has to be performed with extreme care to ensure complete conversion of the original genomic DNA and unbiased amplification of the template. Several groups have published modified protocols to minimise these problems which were associated with the original method (Feil et al., 1994; McDonald and Kay, 1997; Olek et al., 1996; Paulin et al., 1998; Raizis et al., 1995).

Bisulphite sequencing has been used to quantitate methylation at sites in the genome by one of two methods (Paul and Clark, 1996; Xiong and Laird, 1997). In the first method the bisulphite PCR product is sequenced by fluorescent cycle-sequencing in which each dideoxy base is labelled with a different fluorescent dye (see Methods 2.42). Cytosine methylation is assessed at every site by the relative fluorescent signal intensity of the cytosine peak to the thymine. If no thymine signal is present then all molecules are deemed to have been methylated at the site, conversely no cytosine signal implies no methylation at the site. The second method exploits the generation of unique restriction sites by bisulphite treatment to distinguish molecules that were methylated from those that were unmethylated (for an example see section 3.2.3(v), Chapter 3). The bisulphite PCR product is purified and then digested with the discriminatory enzyme and the intensities of the fragments quantitated to determine the ratio of methylated to unmethylated fragments. Notably both of these methods implicitly rely upon unbiased PCR amplification of the bisulphite treated template so rigorous controls must be included to confirm that this is the case.

vi) Methylation specific PCR (MSP)

Methylation specific PCR is a modification of the bisulphite sequencing procedure in which bisulphite treated genomic DNA is amplified by different sets of PCR primers (Herman et al., 1996). Primers are designed to discriminate between methylated and unmethylated templates and between treated and untreated templates. This is achieved by designing primers that have one or more CpG sites at the 3′-end so that they can only amplify bisulphite treated DNA that was methylated at those CpG sites. Primers are then designed to the same target, but with one or more TpG sites at the 3′-end to amplify bisulphite treated DNA that was unmethylated at those sites. Applying similar logic any number of discriminatory primer sets can be designed to test the methylation status of a number of sites in a target without having to resort to expensive and time-consuming product sequencing. This procedure is best applied to regions that are expected to have very high or very low levels of methylation and has been applied particularly well to the analysis of both primary and cultured tumour samples (Cote et al., 1998; Gonzalgo et al., 1997; Herman et al., 1996; Quesnel et al., 1998; Wong et al., 1997).

vii) Immuno-detection

The simplest way to detect methylation is arguably the use of a monoclonal antibody raised to 5-methyl cytosine. Although the resolution of this technique is not high enough to detect sequence specific methylation, it has proved extremely informative in the characterisation of overall genomic methylation (Bensaada et al., 1998; deCapoa et al., 1996; KokaljVokac et al., 1998; Mizugaki et al., 1996; Rougier et al., 1998).

1.3 Outline of this study.

The work described in this thesis involves the analysis of cytosine methylation at the promoters of several X-linked genes to contribute towards the characterisation of the role of

methylation in the X-inactivation process. In addition several novel promoters of X-linked genes are isolated and compared to attempt to identify conserved components that may be involved in the regulation in the context of X-inactivation.

1.3.1 Methylation analysis of the promoter of the pyruvate dehydrogenase E1α (*PDHA1*) gene – Chapter 3.

In this section the promoter of the housekeeping gene *PDHA1* was characterised using the bisulphite genomic sequencing procedure, methylation sensitive PCR and by unique restriction sites generated as a result of bisulphite treatment. The intention was not only to establish the procedure in the laboratory for future use in the project, but also to resolve a debate in the field of X-inactivation. Prevailing opinion was that the CpG-islands of X-linked genes were completely methylated on the inactive X-chromosome and unmethylated on the active, based on data from the analysis of the *PGK-1* and *HPRT* gene promoters. By testing the methylation status of a third housekeeping gene with a CpG-island at the 5′-end this issue could be addressed at a different locus. The resulting methylation profiles were reconciled with published promoter analyses detailing functional transcription factor sites in the promoter. In addition *in-vitro* methylation experiments of promoter-reporter gene constructs were used to demonstrate that promoter methylation was capable of repressing transcription in tissue-culture cells.

1.3.2 Methylation analysis of the promoters of the human and murine Norrie disease protein (*NDP/Ndp*) genes – Chapter 4.

This section describes the characterisation of methylation at the promoters of both the human and murine *NDP/Ndp* genes. Methylation was analysed by bisulphite genomic sequencing in expressing and non-expressing cells and tissues and correlated to the prevailing expression

status of the gene. Comparative sequence and methylation profile analysis between human and murine sequences was undertaken to identify whether methylation profiles could be conserved at putative regulatory sites within the promoters. The work addresses the control of tissue-specific gene expression in the context of the X-inactivation process and demonstrates that *in-vitro* methylation is capable of repressing promoter activity in tissue culture cells.

1.3.3 Isolation of the human and murine *SMCX/Smcx* gene promoters – Chapter 5.

Genes which escape the X-inactivation process have frequently been found to lie closely between other genes which do not escape suggesting that local factors can override the inactivating signal. Escaping genes are usually found to have unmethylated promoters on both the active and inactive X-chromosomes consistent with the promoter being protected from the hyper-methylation that is characteristic of the inactive X-chromosome. In order to generate the resources to study the promoters of genes escaping X-inactivation the promoters of the human and murine *SMCX/Smcx* genes were isolated and sequenced. The section describes their isolation and comparative sequence analysis and demonstrates that the sequences are capable of driving reporter gene expression in tissue culture cells. *In-vitro* methylation of promoter-reporter gene constructs is shown to be capable of repressing this activity. The future analysis of methylation at these promoters is of particular importance as the *Smcx* gene has been shown to partially escape inactivation differentially by tissue after being initially inactivated early in development. These kinds of changes in expression profile are reminiscent of those associated with methylation changes during development.

Chapter 2 - Materials and Methods

2.1 General Solutions.

TAE (50X)

2M Tris-HCl (pH=7.5)
50mM EDTA (pH=8.0)
5.7% (v/v) glacial acetic acid

TBE (5X)

0.45M Tris-HCl (pH=7.5)
0.45M Orthoboric acid
10mM EDTA (pH=8.0)

TE buffer (pH=7.4, 7.6, 8.0)

10mM Tris-HCl (pH=7,4, 7,6, 8.0)
1mM EDTA (pH=8.0)

2.2 PCR buffers.

PCR buffer series I (10x) {named A-E} KCl based

0.5M KCl
0.1M Tris-HCl (pH=8.4)
0.1% (w/v) Gelatine
1.0% (v/v) Triton X-100
A-7.5, B-15, C-22.5, D-30, E-37.5mM $MgCl_2$

PCR buffer series II (10x) {named F-J} NaCl based

0.4M NaCl
0.1M Tris-HCl (pH=8.4)
0.1% (w/v) Gelatin
1.0% (v/v) Triton X-100
F-7.5, G-15, H-22.5, I-30, J-37.5mM $MgCl_2$

PCR buffer series III (10x) {named K-O) $(NH_4)_2SO_4$ based

0.1M $(NH_4)_2SO_4$
0.1M KCl
0.1M Tris-HCl (pH=8.4)
0.1% (w/v) Gelatin
1.0% (v/v) Triton X-100
K-7.5, L-15, M-22.5, N-30, O-37.5mM $MgCl_2$

2.3 Solutions for hybridisation.

DNA depurinating solution
0.4M NaOH
0.6M NaCl

DNA neutralising solution
1.5M NaCl
0.5M Tris-HCl (pH=7.5)

DNA post-neutralising solution
0.2M Tris-HCl (pH=7.5)
1xSSC

Church hybridisation buffer
0.5M Na_2HPO_4 (pH=7.5)
1mM EDTA (pH=8.0)
7% (w/v) SDS (Sodium Dodecyl Sulphate)
1% (w/v) BSA (Bovine Serum Albumin - filter sterilised 0.4μm)
Stored at 60°C

20X SSC
3M NaCl
0.3M Sodium citrate
Add drops conc. NaOH to pH=7.0

Membrane washing solutions
2xSSC/1% (w/v) SDS
1xSSC/1% (w/v) SDS
0.5xSSC/1% (w/v) SDS
0.1xSSC/1% (w/v) SDS

Membrane stripping solution
2mM Tris-HCl (pH=7.5)
1mM EDTA (pH=8.0)
0.1% (w/v) SDS

2.4 Solutions for DNA sequencing.

Sequencing gel mix
6% (w/v) Acrylamide/bisacrylamide (19:1)
8M Urea
1xTBE
0.1% (w/v) Ammonium persulphate
0.1% (v/v) TEMED

Sequencing gel fixative

5% (v/v) Glacial acetic acid
15% (v/v) Methanol

2.5 Loading Buffers.

Sequencing gel loading buffer (6x)

95% (v/v) Formamide
20mM EDTA (pH=8.0)
0.05% (w/v) Bromophenol blue
0.05% (w/v) Xylene cyanol FF

Agarose gel loading buffer I (6x)

30% (v/v) Glycerol
0.25% (w/v) Bromophenol blue
0.25% (w/v) Xylene cyanol FF

Agarose gel loading buffer II (6x)

30% (v/v) Glycerol
0.25% (w/v) Bromophenol blue

Agarose gel loading buffer III (6x)

30% Glycerol
0.25% (w/v) Xylene cyanol FF

2.6 DNA isolation reagents.

Tissue buffer

100mM NaCl
50mM Tris-HCl (pH=7.5)
1% (w/v) SDS
50mM EDTA (pH=8.0)

SET buffer

0.15M NaCl
50mM Tris-HCl (pH=7.8)
5mM EDTA (pH=8.0)

Re-distilled phenol

Equilibrated with 1M Tris-HCl (pH=8.0) until pH≈7.0
then with 0.1M Tris-HCl (pH=8.0)

Proteinase K stock

Proteinase K 30mg/ml in ddH_2O

Blood DNA isolation solution I

10mM Tris-HCl (pH=7.6)
10mM KCl
10mM $MgCl_2$

Blood DNA isolation solution II

10mM Tris-HCl (pH=7.6)
10mM KCl
10mM $MgCl_2$
0.5M NaCl
0.5% (w/v) SDS
2mM EDTA (pH=8.0)

Silica suspension

2g Silica (***Sigma***) size fractionated in 50ml ddH_2O

Bacterial resuspension solution

50mM Tris-HCl (pH=8.0)
10mM EDTA (pH=8.0)

Bacterial lysis solution

200mM NaOH
1% (w/v) SDS

Bacterial neutralisation solution

3M Potassium acetate
pH to 5.5 with glacial acetic acid

Silica wash solution

50mM NaCl
10mM Tris-HCl (pH=7.5)
2.5mM EDTA (pH=8.0)
50% (v/v) Ethanol

PAC/Cosmid DNA isolation solution P1

15mM Tris-HCl (pH=8.0)
10mM EDTA (pH=8.0)
100µg/ml RNaseA
filter sterilised store at 4°C

PAC/Cosmid DNA isolation solution P2

0.2M NaOH
1% (w/v) SDS

PAC/Cosmid DNA isolation solution P3

3M Potassium acetate (pH=5.5)
a/c store at 4°C

2.7 RNA isolation reagents.

Denaturing solution

4M guanidine thiocyanate
26mM Sodium citrate (pH=6.8)
0.125M β-mercapto ethanol
0.5% (w/v) N-lauryl sarcosine

PCIA (Phenol:Chloroform:Iso-amyl alcohol)

Ratio 125:25:1 (pH=4.7)

2.8 Bacterial Media and Reagents.

LB medium

1% (w/v) Bacto-tryptone
0.5% (w/v) Bacto-yeast-extract
0.5% (w/v) NaCl
1mM NaOH

LB-Miller medium

1% (w/v) Bacto-tryptone
0.5% (w/v) Bacto-yeast-extract
1.0% (w/v) NaCl
1mM NaOH

SOB medium

2% (w/v) Bacto-tryptone
0.5% (w/v) Bacto-yeast-extract
0.05% (w/v) NaCl
2.5mM KCl
Add drops NaOH to pH=7.0

Post autoclaving: -

Supplement $MgCl_2$ to 10mM

SOC medium

SOB medium
Supplement glucose to 20mM

Filter sterilised, stored at -20°C

Superbroth medium

3.2% (w/v) Bacto-tryptone
2% (w/v) Bacto-yeast-extract
0.5% (w/v) NaCl
5mM NaOH

TSS solution

1% (w/v) Bacto-tryptone
0.5% (w/v) Bacto-yeast-extract
1% (w/v) NaCl
10% (w/v) Polyethylene glycol (average M_w=8000)
25mM $MgCl_2$

Filter sterilised, stored at -20°C

Glycerol strain storage mix

65% (v/v) Glycerol
0.1M $MgSO_4$
25mM Tris-HCl (pH=8.0)

Kanamycin sulphate stock

30mg/ml in ddH_2O

Ampicillin stock

50mg/ml in ddH_2O

IPTG stock

0.1M IPTG in ddH_2O
Filter sterilised (0.2μm), store at 4°C

X-Gal stock

X-gal 50mg/ml in N,N'-dimethylformamide
Store at -20°C in the dark

LB Agar plates

1% (w/v) Bacto-tryptone
0.5% (w/v) Bacto-yeast-extract
0.5% (w/v) NaCl
1mM NaOH
1.5% (w/v) Agar
Antibiotics to 50μg/ml as required

2.9 Vectors.

i) pZErO™-2 (***Invitrogen***) – sequence available at http://www.invitrogen.com

The majority of routine cloning and all PCR product cloning was carried out with the pZErO-2 vector. This vector is one of a range of new vectors which use lethal genes to allow "zero background cloning". The vector contains a kanamycin resistance cassette and the lethal *E.coli* gene *ccd*B (Bernard et al., 1994) fused to the C-terminus of lacZα under the control of the P_{lac} promoter. Insertion of a DNA fragment into the multiple cloning site (MCS) of this vector leads to the disruption of the lethal fusion protein, allowing bacteria carrying these clones to grow. If no insert is present the lethal protein is expressed and the bacteria die. In the clones characterised during the course of this thesis 90-99% contained the correct insert.

As the expression of the fusion protein is essential for selection, the P_{lac} promoter must be constitutively active. The promoter can be induced in bacteria either with IPTG (which binds to the P_{lac} repressor protein $lacI^q$) or by using bacteria that do not carry the $lacI^q$ gene. The *E.coli* strain TOP10 (***Invitrogen***) does not have the F' episome which carries the $lacI^q$ gene, so pZErO-2 transformed into these bacteria constitutively expresses the LacZα-ccdB fusion protein.

The ability to use lethality to avoid a background of insert-less clones means that you no longer need to de-phosphorylate the 5' ends of vectors prior to cloning; nor do you need to use blue/white selection to identify clones with inserts. Overall this allows low frequency ligation events to be isolated when combined with high transformation efficiency. This greatly facilitates blunt-cloning of PCR products and this approach has been used widely in this project. Once cloned, inserts can be readily sequenced as the MCS of pZErO-2 is flanked by SP6 (ATTTAGGTGACACTATAG) and T7 (GTAATACGACTCACTATAGGGC) primer sites. The MCS contains restriction sites for the following enzymes amongst others. HindIII, SacI, BamHI, EcoRI, PstI, EcoRV, NotI, XhoI, XbaI and ApaI providing a wide choice for cloning.

ii) pBluescript® II KS (+) phagemid (***Stratagene***) – Genbank Accession X52327

This vector was also used for routine cloning. The vector is a high copy number phagemid of 2691bp containing a gene for ampicillin resistance. The multiple cloning site (MCS) interrupts the lacZ gene which is under the control of the inducible P_{lac} promoter allowing blue/white colour screening for insert carrying clones. The MCS is flanked by T7 and T3 (AATTAACCCTCACTAAAGGG) primer sites to facilitate insert sequencing. The MCS has restriction sites for the following enzymes, amongst others. ApaI, XhoI, SalI, ClaI, HindIII, EcoRV, EcoRI, PstI, SmaI, BamHI, XbaI, NotI, EagI and SacI giving a wide choice for cloning.

iii) pEGFP-1 Vector (***Clontech***) – Genbank Accession U55761

This mammalian expression vector encodes the EGFP variant (Cormack, 1996) and (Yang, 1996) for monitoring the activity of promoters cloned into its multiple cloning site. Sequences flanking the EGFP gene have been converted to a Kozak consensus translation

initiation sequence and the EGFP coding sequence has been codon optimised for mammalian cell expression. The vector contains an SV40 and f1 origin of replication, a mammalian polyadenylation signal and a neomycin resistance cassette. The MCS contains the following restriction sites amongst others. XhoI, SacI, HindIII, EcoRI, PstI, SalI, ApaI, SmaI and BamHI.

iv) mtEGFP Vector – fusion construct derived from pEGFP-N1 Vector (***Clontech***) that is targeted to the mitochondrion (gift of R.M. Brown).

2.10 Bacterial Strains.

TOP10F' (***Invitrogen***) : F' {*lac*I^q*Tn*10(Tet^R)} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ* ΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G

SCS110 (***Stratagene***) : F' {*tra*D36 *pro*AB *lac*I^{q}ZΔM15} *rps*L (Str^R) *thr leu end*A *thi*-1 *lac*Y *gal*K *gal*T *ara ton*A *tsx dam dcm sup*E44 Δ(*lac-pro*AB)

LE392 (***Stratagene***) : e14$^-$ (*Mcr*A$^-$) *hsd*R514 *sup*E44 *sup*F58 *lac*Y1or Δ(*lac*IZY)6 *gal*K2 *gal*T22 *met*B1 *trp*R55

DH5α (***Clontech***) : *deo*R *end*A1 *gyr*A96 *hsd*R17($r_k^- m_k^+$) *rec*A1 *sup*E44 *thi*-1 Δ(*lac*ZYA-*arg*FV169) ϕ80δ*lac*ZΔM15 F$^-$ λ$^-$

2.11 Libraries.

i) Mouse PAC library.

RPCI21 (***HGMP-RC***) constructed by Kazutoyo Osoegawa in Pieter de Jongs' laboratory at the Rosell Park Cancer Institute, Buffalo. The library is cloned in the vector pPAC4 which was constructed by Erkik Frengen. The library has approx. 128,899 clones with an average insert size of 147kb. The DNA source was a female 129/SvevTACfBr mouse spleen.

ii) Human X-only Cosmid library.

Chromosome X LL0XNC01 "U" (***HGMP-RC***) constructed at the LLNL. The genomic fragments were cloned into the vector Lawrist16. The library has approx. 24,000 clones representing a 5.3 fold coverage of the chromosome with an average insert size of 32-47kb.

iii) Human "CpG-Island" library.

CGI1-library (***HGMP-RC***) constructed by Sally Cross in Adrian Birds' laboratory at the ICMB, Edinburgh. The library was cloned in the vector pGEM-5Zf(-) and transformed into *E.coli* strain XL1-BLUE MRF' as described in (Cross et al., 1994). The library has approx. 6 fold coverage of the predicted 45,000 "CpG-Islands" in the human genome and has an average insert size of 760bp.

iv) Murine "CpG-Island" library.

Mouse CGI library (***HGMP-RC***) constructed by Sally Cross in Adrian Birds' laboratory at the ICMB, Edinburgh. The library was cloned in the vector pGEM-5Zf(-) and transformed into *E.Coli* strain SURE MRF' as described in (Cross et al., 1997). The library has approx. 5 fold coverage of the predicted 37,000 "CpG-Islands" in the mouse genome and has an average insert size of 582bp.

2.12 Molecular Weight ladders.

100bp ladder - DNA Molecular weight marker XIV (***Boehringer-Mannheim***) – with DNA fragments of sizes from 100-1500bp at 100bp intervals and an additional band at 2642bp.

500bp ladder - DNA Molecular weight marker XVII (***Boehringer Mannheim***) – with DNA fragments of sizes 0.5kb-5.0kb at 0.5kb intervals.

λ DNA/Hind III fragments (***Gibco-BRL***) – with DNA fragments of sizes 125, 564, 2027, 2322, 4361, 6557, 9416 and 23130bp.

1Kb DNA ladder (***Gibco-BRL***) – with DNA fragments of sizes 506, 517, 1018, 1636, 2036, 3054, 4072, 5090 and 6108bp continuing up to 12216bp at approximately 1kb intervals.

2.13 Tissue Culture Reagents and Cell lines.

Reagents by ***Sigma*** unless otherwise indicated.

Dulbecco's modified eagle modified nutrient mixture F12-HAM (DME/F-12 1:1 Mixture).

RPMI 1640 – HEPES modification.

Dulbecco's phosphate buffered saline (DPBS).

Hank's balanced salt solution (HBSS).

Fetal Bovine Serum (FBS/FCS) (F7524).

Penicillin-Streptomycin solution (P0781).

Trypsin (1x solution – 0.25% (w/v) in HBSS) (T4424).

EDTA (di-sodium salt - 2% (w/v)) (28-205-49) (***ICN***).

HEPES buffer (1M) (H0887).

L-Glutamine 200mM (G7513).

Cell freezing medium (DMSO) (C6164).

Trypsin working solution
0.025% (w/v) Trypsin
0.02% (w/v) EDTA
make up with PBS

DME/F-12 working solution
DME/F-12 stock
10% (w/v) FCS
20U/ml Penicillin
20μg/ml Streptomycin
2mM L-Glutamine
(and same for RPMI-1640)

Freezing mix
95% (v/v) FCS
5% (v/v) DMSO

Non-aqueous Fixative
3:1 (v/v) methanol/glacial acetic acid

Aqueous-Fixative
4% (w/v) Paraformaldehyde in PBS
filter sterilise 0.4μm store at -20°C

Aqueous Slide Mountant (Moviol 4-88) (***Harco***)
(see Methods 2.54)

2.14 Tissue culture cell lines.

i) CCL-2 (***ATCC***) (HeLa) - derived from human cervical adenomcarcinoma, grow as monolayer, polyploid.

ii) HTB-169 (***ATCC***) (Weri-Rb-1) - derived from human female retinoblastoma, grow as grape-like clusters, near diploid.

iii) CRL-2302 (***ATCC***) (ARPE-19) - derived from human male normal eye retinal pigment epithelium, grow as monolayer, diploid.

iv) 45504 (***ATCC***) (293) - derived from human embryonic kidney, grow as monolayer, unknown ploidy.

v) 661W (gift of ***A.Wright/Al-Ubaidi***) - derived from mouse retinal tissue (Alubaidi et al., 1992) and grow as grape-like monolayer, ploidy unknown.

2.15 Genomic DNA isolation from tissue culture cells.

1.) Wash the pellet with PBS buffer and then resuspend fully in 10ml SET buffer.

2.) Add Proteinase K to 1mg/ml.

3.) Add 0.5ml 10% (w/v) SDS and mix well by inversion.

4.) Incubate overnight at 37°C.

5.) Extract the protein by adding an equal volume of re-distilled phenol and shaking for 10-15 minutes or sufficiently to obtain a uniform emulsion.

7.) Centrifuge at 2000rpm for 5 minutes.

8.) Remove the upper aqueous layer.

9.) Extract again with 1 volume of chloroform/1% (v/v) *iso*-amyl alcohol.

10.) Remove upper aqueous phase.

11.) Add RNase to 50µg/ml from a heat treated stock and incubate at 37°C for 1-2 hours.

12.) Repeat 9) and 10) above.

13.) Add 2-2.5 volumes of cold absolute ethanol to the aqueous phase and mix gently by inversion.

14.) Spool the precipitating DNA onto a sterile sealed capillary tube and transfer to a clean 1.5ml microfuge tube.

14a.) (Optional). If there is little precipitate, centrifuge to collect the DNA and remove the supernatant.

15.) Wash the pellet by rinsing with 75% (w/v) ethanol.

16.) Remove all remaining liquid and allow to air dry briefly.

17.) Resuspend the DNA in an appropriate amount of TE buffer at 4°C overnight (typically about 0.5ml for a preparation from $1X10^7$ cells).

2.16 Genomic DNA isolation from peripheral blood leukocytes.

Protocol can be scaled down.

1.) Collect blood into a 15ml tube containing di-sodium EDTA and shake well.

2.) Make volume up to 10 ml with Blood DNA isolation solution I.

3.) Add 120µl Nonidet P40 (***BDH***) and mix well by inverting several times.

4.) Spin down the nuclear pellet by centrifuging at 2000rpm for 10 minutes.

5.) Pour off the supernatant without dislodging the pellet. (These pellets can be stored frozen at -80°C).

6.) Gently resuspend pellet well in 800µl Blood DNA isolation solution II and transfer to a 1.5ml microfuge tube.

7.) Add 400µl of distilled phenol (pH=8.0) and mix well.

8.) Centrifuge for 1 minute in a microfuge and transfer the upper (aqueous) phase to a clean 1.5ml microfuge tube.

9.) Add 200µl of phenol and 200µl of chloroform:iso-amyl alcohol (24:1) and mix well by inversion.

10.) Centrifuge for 1 minute then transfer the upper phase to a clean 1.5ml microfuge tube.

11.) Add 700µl of chloroform:iso-amyl alcohol (24:1) and extract as above.

12.) Transfer the upper phase to a 1.5ml microfuge tube add 2 volumes of ice cold absolute ethanol and mix to precipitate the DNA.

13.) Spool out the DNA with a sealed capillary tube into a microfuge tube containing 1ml 70% (v/v) ethanol and mix well.

14.) Centrifuge for 5 minutes, remove excess ethanol and dry pellet in a vacuum dessicator. Do not over dry.

15.) Resuspend the DNA in a suitable volume of TE or 10mM Tris-HCl (pH=7.5) at 65°C.

16.) The DNA can be stored at 4°C.

2.17 Genomic DNA isolation from tissues.

1.) Excise the desired quantity of fresh or frozen tissue (typically 100mg or less) and mince with a scalpel blade or fine scissors.

2.) Transfer the minced tissue to a medium sized dounce homogeniser containing 300µl of tissue digestion buffer and homogenise firmly until the solution is fully homogenised.

3.) Transfer the solution to a 1.5ml microfuge tube and add Proteinase K to a final concentration of 100µg/ml.

4.) Incubate the digest at 50°C for 2 hours then at 37°C overnight.

5.) After digestion add an equal volume (300µl) of 5M LiCl and mix thoroughly by inversion for 1 minute.

6.) Add 2 volumes (600µl) of chloroform, seal the tube with parafilm and then place on a rotating wheel for 30 minutes.

7.) Centrifuge the sample for 15 minutes then carefully remove the supernatant to a clean 1.5ml microfuge tube.

8.) Add exactly 2 volumes of room temperature absolute ethanol and invert until DNA has precipitated.

9.) Recover the DNA by centrifugation then wash pellet with 70% (v/v) ethanol and centrifuge again.

10.) Remove excess ethanol and allow the sample to air dry for at least 10 minutes.

11.) Resuspend the sample in an appropriate volume of TE buffer or 10mM Tris-HCl (pH=7.5) (typically 200-400µl) and allow to dissolve overnight at 4°C.

2.18 RNA isolation from tissue culture cells.

RNA was isolated from tissue culture cells using the High Pure RNA Isolation Kit (***Boehringer Mannheim***) according to the manufacturer's instructions.

1.) Resuspend cells (typically $1x10^6$ cells) in 200µl PBS.

2.) Add 400µl lysis/binding buffer and mix well by inversion.

3.) Transfer the lysate into the upper reservoir of an assembled purification column

4.) Centrifuge in a microfuge for 15 seconds then discard the eluent.

5.) Pipette 90µl DNase I incubation buffer into a 0.5ml microfuge tube and add 10µl DNase I. Mix and transfer to the upper reservoir of the assembled column and incubate at room temperature for 15 minutes.

6.) Add 500µl wash buffer I and centrifuge for 15 seconds.

7.) Discard the eluent and then add 500µl wash buffer II and centrifuge for 15 seconds.

8.) Discard the eluent and then add 200µl wash buffer II and centrifuge for 2 minutes.

9.) Place the upper reservoir into a 1.5ml microfuge tube and add 50-100µl of elution buffer to the upper reservoir.

10.) Centrifuge for 1 minute. Store the eluted RNA at -80°C or use immediately.

2.19 RNA isolation from tissues.

1.) Allow PCIA mix to come to room temperature for at least 15 minutes and check aqueous phase is separate from the organic phase. Chill the denaturing solution on ice for at least 5 minutes.

2.) Aliquot 300µl denaturing solution into a chilled sterile dounce homogeniser.

3.) Handling quickly, excise the desired quantity of fresh or frozen tissue (typically 25mg or more) and mince with a sterile blade or fine scissors. Transfer the minced tissue to the dounce homogeniser containing denaturing solution.

4.) Homogenise the tissue to completion then transfer the solution to a 1.5ml microfuge tube.

5.) Add 30µl of 2M Sodium acetate and mix by inverting several times.

6.) Add 300µl PCIA and mix by inversion 3-4 times, then by shaking vigorously for 10 seconds.

7.) Chill on ice for 15 minutes.

8.) Centrifuge in a microfuge for 10 minutes at 4°C.

9.) Remove the top (aqueous) phase to a clean 1.5ml microfuge tube taking care to avoid the interface.

10.) Add 1 volume of *iso*-propanol and incubate at -20°C for 30 minutes.

11.) Centrifuge for 10 minutes at 4°C.

12.) Wash the pellet with 1ml 75% (v/v) ethanol breaking up the pellet to ensure adequate washing.

13.) Centrifuge for 10 minutes at 4°C.

14.) Vacuum desiccate for 5 minutes.

15.) Resuspend pellet in a suitable volume of nuclease free water (typically approx. 50µl).

2.20 Isolation of PAC DNA.

Suitable for PAC, Cosmid or Plasmid DNA preparation.

1.) Inoculate a single bacterial colony into 2ml LB medium supplemented with Kanamycin to 25µg/ml and allow to grow overnight (<16 hours) at 225-300rpm and 37°C.

2.) Centrifuge at 3000rpm for 10 minutes.

3.) Resuspend the pellet in 300µl of PAC/Cosmid DNA isolation solution P1, add 300µl PAC/Cosmid DNA isolation solution P2 and mix gently by inverting a few times. Leave at room temperature for 5 minutes.

4.) Add 300µl of PAC/Cosmid DNA isolation solution P3 shaking gently during addition. Place the tubes on ice for at least 5 minutes.

5.) Centrifuge at 10,000rpm for 10 minutes at 4°C.

6.) Place the tubes on ice then transfer the supernatant to a 1.5ml microfuge tube containing 0.8ml of ice cold *iso*-propanol and mix.

7.) Place the tubes on ice for at least 5 minutes (can be left o/n at -20°C).

8.) Centrifuge in a microfuge at 4°C for 15 minutes.

9.) Wash pellet at least once in 0.5ml 75% (v/v) ethanol recovering pellet by centrifuging for 5 minutes at 4°C.

10.) When washing completely remove excess ethanol and allow to air-dry until the pellet has turned from white to translucent.

11.) Gently resuspend in 40µl 10mM Tris-HCl (pH=7.5) or TE buffer, by flicking the bottom of the tube.

12.) Store at -20°C.

2.21 Plasmid Maxi DNA Preparation.

Maxi preparations of plasmid DNA were achieved by using the "Hybaid Recovery™" Quick-flow Maxi Kit (***Hybaid***) according to the manufacturer's instructions.

1.) Inoculate a single bacterial colony into approximately 5ml LB medium supplemented to 50µg/ml with Kanamycin or Ampicllin as appropriate. Culture overnight at 225-300rpm at 37°C.

2.) Inoculate the overnight culture into 100ml LB medium supplemented to 50µg/ml with Kanamycin or Ampicillin as appropriate. Culture overnight at 225-300rpm at 37°C. Final absorption at 600nm should be between 1-1.5 equating to roughly $1x10^9$ cells/ml.

3.) Harvest the bacterial cells by centrifugation at 6000g for 15 minutes at 4°C.

4.) Whilst harvesting the cells equilibrate the column by adding 30ml equilibration solution.

5.) Remove the supernatant from the bacterial harvest ensuring any "run-back" is also removed.

6.) Resuspend the pellet in 10ml resuspension solution until homogenous.

7.) Add 10ml of cell lysis solution and mix by gentle inversion 5 times. Do not vortex. Incubate at room temperature for 5 minutes.

8.) Add 11ml neutralisation solution and mix immediately by inverting the tube 5 times. Do not vortex.

9.) Centrifuge at ≥ 15000g at 20°C for 10 minutes.

10.) Transfer the supernatant promptly to the equilibrated column being careful to avoid any pellet material.

11.) Allow the lysate to run through by gravity.

12.) Apply 60ml of wash solution to the column and allow to run through by gravity.

13.) Elute the plasmid DNA by adding 15ml of elution solution to the column and allowing to run through by gravity.

14.) Add 0.7 volumes of *iso*-propanol to the eluent (approx. 10.5ml) and mix to precipitate the plasmid DNA.

15.) Centrifuge at ≥ 15000g at 4°C for 30 minutes.

16.) Wash the plasmid DNA with 70% (v/v) ethanol and centrifuge as in 15.) (Above)

17.) Remove excess ethanol and allow to air-dry at room temperature for 10 minutes.

18.) Redissolve the DNA in an appropriate volume of 10mM Tris-HCl (pH=7.5) or TE buffer. Typically approx. 300-500µl.

19.) Store the DNA at -20°C.

2.22 Plasmid DNA miniprep.

Standard plasmid mini-preparations were carried out using the Qiaprep Spin Miniprep kit (***Qiagen***) according to the manufacturer's instructions.

1.) Inoculate a single colony or a scrape from a frozen bacterial stock into 5ml LB medium supplemented to 50µg/ml with Kanamycin or Ampicillin as appropriate.

2.) Culture overnight at 225-300rpm at 37°C.

3.) Transfer 1.5ml of overnight culture to a microfuge tube and harvest the cells by centrifuging for 1 minute.

4.) Aspirate all excess supernatant and resuspend pellet gently in 300µl P1 buffer.

5.) Add 250µl P2 buffer and mix by inversion several times. Incubate at room temperature for 5 minutes.

6.) Add 350µl P3 buffer and mix immediately by inversion.

7.) Centrifuge for 10 minutes at room temperature.

8.) Apply supernatant to a spin column and centrifuge for 30-60 seconds. Remove eluent.

9.) (Optional). Apply 0.5ml of PB buffer and centrifuge for 30-60 seconds. Remove eluent.

10.) Apply 0.75ml PE buffer and centrifuge for 30-60 seconds.

11.) Remove eluent and centrifuge again for 1 minute to remove any excess wash buffer.

12.) Add 30-50µl elution buffer to the column. Incubate at room temperature for 1 minute then centrifuge for 1 minute to elute the Plasmid DNA.

13.) Store at -20°C.

2.23 Rapid Miniprep of plasmid DNA.

This method provides a cheap alternative to commercial minipreps and is faster and less toxic than other alkali-lysis approaches.

2.23.1 Silica preparation.

1.) Prepare the silica (***Sigma*** S5631) by suspending two grams in 15ml of ddH_2O and allow it to settle for 45 minutes or until most of the silica forms a loose pellet.

2.) Aspirate the supernatant which contains very fine silica particles and resuspend the remaining silica in 15ml ddH_2O. Allow to settle again and repeat 1.) and 2.) once more.

3.) Finally resuspend the silica pellet in 50ml ddH_2O. Store at room temperature.

2.23.2 Miniprep protocol.

1.) Pellet 1.5-3.0ml overnight bacterial culture in a microfuge for 10 seconds.

2.) Resuspend the pellet in 200µl resuspension solution.

3.) Add 200µl lysis solution and mix by inversion until the solution has cleared.

4.) Add 200µl neutralisation solution and mix by inversion.

5.) Centrifuge in a microfuge for 3-5 minutes at room temperature.

6.) Aspirate the supernatant into a clean 1.5ml microfuge tube containing 200µl of the stock silica solution. Mix by inverting several times.

7.) Centrifuge for 10 seconds and remove the supernatant.

8.) Wash the pellet by adding 0.5ml wash solution and vortexing briefly.

9.) Centrifuge for 10 seconds and then aspirate supernatant.

10.) Centrifuge again for 10 seconds and remove any remaining supernatant.

11.) Allow to dry for 10 minutes at room temperature.

12.) Resuspend the silica pellet in 35µl H_2O or 10mM Tris-HCl (pH=7.5) and heat to 55°C for 2 minutes.

13.) Centrifuge for 30 seconds then collect 30µl of supernatant to a fresh tube.

14.) Store plasmid DNA preparation at -20°C.

2.24 Restriction enzyme digests.

Restriction endonuclease digestion was carried out with purified recombinant enzymes according to conditions recommended by the manufacturers. When more than one enzyme was used in a digest the optimal buffer was used as described by New England Biolabs (***NEB***). If there was no suitable buffer for combined digestion the digests were carried out in series.

Typically, genomic DNA was digested with three-fold excess of enzyme at the recommended temperature overnight. Plasmid and PCR products were typically digested with 1-10 units of enzyme in a 20µl reaction at 37°C for 30 minutes to 1 hour.

Digests were frequently de-salted by ethanol precipitation prior to subsequent manipulation, especially when the DNA was to be used in gels for Southern blotting or in ligations.

2.25 Agarose gel electrophoresis.

DNA fragments were separated in agarose gels of 0.6-2% (w/v) using agarose for conventional electrophoresis (***Sigma***). Gels were prepared in 1xTBE buffer and run in 1xTBE buffer at 3-8V/cm. Fragments smaller than ≈300bp were separated in gels of 2-3% (w/v) NusieveGTG agarose (***FMC Bioproducts***). Genomic DNA digests and digests of PACs and Cosmids were run in 0.8% (w/v) gels at a lower voltage gradient of 1.5-2V/cm overnight.

Gels were typically made containing ethidium bromide at a final concentration of 0.1µg/ml from a stock at 10mg/ml. Gels run overnight were stained after electrophoresis in a bath of 1xTBE containing ethidium bromide at 0.5µg/ml for 15-20 minutes with shaking. DNA was loaded with one of three glycerol based loading buffers containing either one or both of the dyes xylene cyanolFF and bromophenol blue (see Materials 2.5).

Gels were visualised on an UV trans-illuminator and images were captured with a Polaroid land camera using Polaroid665 film, or by an IS-1000 digital imaging system (***Alpha Innotech Corporation***) as a TIF computer image file or as a print-out in thermal image format.

Quantitation and sizing of DNA fragments was achieved by comparison to a lane containing a reference molecular weight ladder. This was achieved either by eye or by using the IS-1000 imaging system, which can analyse images of the gel with in-built software. The software estimated the approximate concentration of bands in the lanes of gels by calculating the area under the intensity curve for each band.

2.26 Gel extraction of DNA.

DNA bands were extracted from agarose gels using the Qiaquick Gel Extraction kit (***Qiagen***) according to the manufacturer's instructions.

1.) Excise the DNA fragment from the agarose gel minimising the amount of gel.
2.) Weigh the gel slice in a 1.5ml microfuge tube and add 3 volumes of QG buffer (assuming 100mg gel ≈ 100µl).
3.) Incubate at 50°C for 10 minutes or until the gel slice has completely dissolved. Mix every few minutes to help dissolution.
4.) Add 1 volume of *iso*-propanol and mix by inversion.
5.) Apply the mix to the column, which can take about 800µl. Centrifuge for 60 seconds. Remove eluent.
6.) Continue applying the solution to the column and centrifuging until it has all been applied.
7.) Add 0.5ml QG buffer to the column and centrifuge for 60 seconds. Remove eluent.
8.) Add 0.75ml of PE buffer to the column and centrifuge for 60 seconds.
9.) Remove the eluent and centrifuge again for 1 minute to remove any remaining PE buffer.
10.) Apply 20-50µl elution buffer to the column and incubate for 60 seconds at room temperature.
11.) Centrifuge for 60 seconds to elute the DNA. Store at –20°C.

2.27 General DNA Clean-up.

De-salting of DNA and removal of proteins was carried out by using the Wizard DNA Clean-Up System (***Promega***) according to the manufacturer's instructions.

1.) Make the sample volume up to 50-500µl with sterile water.
2.) Vigorously shake up the resin to break up any precipitate. Alternatively incubate the resin at 25-37°C for 10 minutes. Allow cooling to < 30°C before use.
3.) Add 1ml of resin to the sample and mix by inversion.
4.) Place a column onto the end of a syringe barrel with the plunger removed. Aliquot the resin mix into the syringe barrel and slowly push the slurry through the column.
5.) Detach the column and remove the plunger from the syringe barrel. Re-attach the column and add 2ml 80% (v/v) *iso*-propanol to the barrel.
6.) Push the *iso*-propanol gently through the column.
7.) Remove the column and insert it into a clean 1.5ml microfuge tube. Centrifuge for 2 minutes to dry the resin.
8.) Transfer the column to a new microfuge tube and apply 20-100µl pre-warmed (65-70°C) 10mM Tris-HCl (pH=7.5) or TE buffer. Incubate for 1 minute at room temperature.
9.) Centrifuge for 20 seconds to elute the DNA. Store at -20°C.

(N.B. Large DNA fragments >20kb should be eluted with buffer at 80°C).

2.28 Ethanol precipitation of DNA.

DNA was usually precipitated using sodium acetate, but both ammonium acetate and lithium chloride were also used where excess carry through of salt would have impeded subsequent manipulations.

1.) Add $1/10^{th}$ volume of 3M Sodium acetate (pH=5.2)
Or 1a.) Add 1.5 volumes of 5M Ammonium acetate (pH=7.0)
Or 1b.) Add $1/10^{th}$ volume of 4M LiCl
2.) Add 2 volumes of cold absolute ethanol and mix by inversion.
3.) Incubate on ice or at -20°C for 20 minutes to overnight.
4.) Centrifuge for 20 minutes in a microfuge.
5.) Remove the supernatant and wash the pellet briefly with 250-500µl of 75% (v/v) ethanol.

6.) Centrifuge for 5-10 minutes in a microfuge.

7.) Remove supernatant and allow to air dry.

8.) Resuspend the pellet in a suitable volume of either water, 10mM Tris-HCl (pH=7.5) or TE buffer.

2.29 Southern blotting from agarose gels.

DNA was transferred from agarose gels onto nylon membranes by southern blotting as described by (Southern, 1975). GeneScreenPlus® (***DuPont***) or HybondN$^+$ (***Amersham***) nylon membranes were used in all Southern transfers. One of two methods was used; protocol 2.29.1 for genomic DNA which preserves the integrity of the DNA better than protocol 2.29.2 which was used for rapid transfer of DNA from gels with higher copy number targets.

2.29.1 Blotting by salt transfer

1.) Cut a section of nylon membrane to the size of the gel from which the DNA is to be transferred.

2.) Pre-wet the membrane with distilled water for a few seconds.

3.) Equilibrate the membrane in 10xSSC for 15 minutes.

4.) Agitate the gel in 0.25M HCl for 10 minutes.

5.) Rinse the gel briefly with distilled water and then soak in 0.4M NaOH/0.6M NaCl for 30 minutes.

6.) Neutralise the gel by agitating in 1.5M NaCl/0.5M Tris-HCl (pH=7.5) for 30 minutes.

7.) Set up blotting apparatus with two sheets of 3MM Whatman paper as wicks dipping into pool of 10xSSC.

8.) Place the pre-treated gel onto the wick-covered platform avoiding bubbles then place the pre-treated membrane on top followed by two sheets of 3MM Whatman cut to the same size as the membrane.

9.) Smooth the assembly out by rolling a pipette across the surface of the top piece of Whatman paper. Pile additional filter paper or paper towels on top to act as a sink for the transfer solution. Secure with a light weight of approximately 0.5kg.

10.) Allow the transfer to continue for as little time as necessary. Typically transfer is performed overnight although a few hours can be sufficient for transfer of high copy number targets.

11.) When transfer is complete, remove the membrane from the apparatus and denature the attached DNA by treating with 0.4M NaOH for 1 minute.

12.) Neutralise the membrane in 0.2M Tris-HCl (pH=7.5)/1xSSC for 1 minute.

13.) The DNA is now ready for hybridisation although it can be additionally fixed to the membrane by cross-linking either by baking at 80°C for 1-2 hours or by brief exposure to UV light (1-2 minutes at 254nm).

14.) Membranes can be stored wet or dry once washed in 2xSSC and stored at either room temperature or -20°C.

2.29.2 Rapid alkaline transfer

1.) Cut a section of nylon membrane to the size of the gel from which the DNA is to be transferred.

2.) Pre-wet the membrane with distilled water for a few seconds.

3.) Equilibrate the membrane in 0.4M NaOH for 10-15 minutes.

4.) Agitate the gel in 0.25M HCl for 10 minutes.

5.) Denature the DNA by soaking the gel in 0.4M NaOH for 30 minutes.

6.) Set up blotting apparatus as described above, but replacing 10xSSC with 0.4M NaOH.

7.) After transfer wash the membrane with excess 2xSSC for 1-2 minutes. The filter can now be stored wet or dry and stored at either room temperature or -20°C.

2.30 Multiprime labelling of DNA probes.

DNA probes were labelled using the DNA multiprime labelling kit (***Amersham***) according to the manufacturers instructions. Typically the reaction volume was 25µl.

1.) Add 25ng of probe to a 1.5ml microfuge tube and make the volume up to 14.5µl with distilled water.

2.) Denature the probe in a boiling water bath for 3 minutes and then quench on ice.

3.) Add the following components in order,

5.0µl 5x Reaction buffer

2.5µl 10x Random hexamers

1.0µl Klenow DNA Polymerase

2.0µl α^{32}P-dCTP (***Amersham)*** (10µCi/µl)

Incubate at 37°C for 30 minutes – overnight

4.) The probe is now ready to be used in hybridisation or it can be cleaned up prior to use by spin dialysis. Labelled probe can be stored for short periods at -20°C.

2.31 Spin dialysis – Sephadex G-50.

Labelled probes can be purified from unincorporated dNTPs and other contaminants by spin dialysis through a Sephadex G-50 spin column.

1.) Make a hole in a 1.5ml microfuge tube and plug the bottom by inserting a small wad of glass wool.
2.) Apply a slurry of Sephadex G-50 in TE buffer (pH=8.0) to fill the tube and insert it into the bottom of a 1.5ml cryotube. Centrifuge for 2 minutes to make the column.
3.) Replace the bottom cryotube with a new one.
4.) Make the labelled probe up to 75µl with distilled water and apply it to the column. Centrifuge for 2 minutes.
5.) Add 100µl of TE buffer (pH=8.0) and centrifuge again for 2 minutes.
6.) Relative incorporation of label can be estimated by measuring the radioactivity of the column compared to the eluent with a Geiger counter at a fixed distance.
7.) The eluent is ready for use in hybridisation or can be stored at -20°C.

2.32 Hybridisation.

1.) Pre-soak any membrane(s) and meshes to be used in 2xSSC for a few seconds.
2.) Alternately layer mesh and membrane up to a maximum of about 6 per tube.
3.) Using a long glass pipette, or comparable tool, roll the membranes and then slide them off into a hybridisation bottle.
4.) Rotate the bottle to allow the membrane(s) to unroll evenly against the surface of the bottle without air bubbles.
5.) Pour off any excess 2xSSC and then introduce 20ml of Church hybridisation buffer. Make sure the bottle is well sealed then pre-hybridise the membrane in a hybridisation oven at 65°C for at least 30 minutes with rotation.
6.) Once pre-hybridisation is complete denature the labelled probe solution at 100°C for 5 minutes then quench on ice for at least 3 minutes.

7.) Mix the denatured probe briefly with about 0.5ml of Church hybridisation buffer and introduce into the hybridisation bottle containing the rolled membrane(s).

8.) Hybridise for 3 hours to overnight at 65°C with rotation.

2.33 Membrane washing, autoradiography and stripping.

1.) Once hybridisation is complete rinse the membrane(s) twice with 2xSSC at room temperature.

2.) Add 100ml of 2xSSC/1% (w/v) SDS at 65°C and incubate with rotation in the hybridisation oven at 65°C for 15 minutes.

3.) Remove the membrane(s) from the bottle and place flat in a washing box. Assess the effect of the first wash by estimating the signal on the membrane(s) with a Geiger counter. If the background membrane count is 5-10cps washing is complete if not, subsequent stringency washes are required.

4.) Any additional stringency washes are carried out with 100ml of wash buffer at 65°C for 20 minutes in the washing box. At the end of each wash scan the membrane(s) with a Geiger counter and stop washing when the background counts on the membrane(s) are down to 5-10cps. The typical stringency wash series is,

2xSSC/1% (w/v) SDS	1st wash
1xSSC/1% (w/v) SDS	2nd wash
0.5xSSC/1% (w/v) SDS	3rd wash
0.1xSSC/1% (w/v) SDS	4th wash

5.) Once washing is complete rinse the membrane(s) briefly in 2xSSC, wrap in cling film and then expose to XOMAT-AR film (***Kodak***) in a cassette with an intensifying screen at -70°C.

6.) Leave the film to expose for 30 minutes – 1 week depending on signal intensity and then develop manually to desired intensity and contrast.

7.) If the membrane(s) need to be re-probed strip them of labelled probe by one of three methods,

0.1xSSC/1% (w/v) SDS at 65°C for >2 hours (gentle)

0.1xSSC/1% (w/v) SDS at 100°C for 10-30m (harsh)

TE buffer (pH=7.5)/ 1% (w/v) SDS at 100°C for 10-30 minutes (harsh)

8.) Rinse the stripped membrane(s) briefly in 2xSSC and re-use or store wrapped in cling-film at -20°C.

2.34 Preparation of chemically competent cells.

This protocol makes enough for 50 transformations and can be scaled up or down as desired.

1.) Inoculate a single bacterial colony into approximately 5ml LB medium in sterile conditions. Culture overnight at 225-300rpm at 37°C.

2.) Inoculate 50ml of LB medium with 0.5ml of 1M $MgCl_2$ and 0.5ml of 20% (w/v) glucose and with 500µl of the overnight culture.

3.) Incubate with shaking at 37°C for 2-5 hours until the absorbance at 600nm = (0.3-0.4).

4.) Place the culture on ice for 10-15 minutes and then spin for 15 minutes at 2400rpm at 4°C.

5.) Resuspend the pellet in 4.5ml of TSS solution and add 350µl DMSO.

6.) Aliquot 100µl of the mix into 0.5ml microfuge tubes and freeze immediately in dry ice. Store at -70°C.

2.35 Chemical transformation of bacterial cells.

1.) Thaw a 100µl aliquot of chemical competent cells on ice.

2.) Directly add 10-100ng of the DNA to be transformed to the thawed cells, mix by tapping gently, and then replace on ice for 30 minutes.

3.) Heat shock the mix at 42°C for 90 seconds then immediately add 400µl of SOC medium and mix by tapping.

4.) Recover the bacteria by incubating the mix at 37°C with shaking for 1 hour.

5.) Plate different dilutions of the transformed bacteria onto suitable selective LB agar plates.

6.) Incubate overnight at 37°C.

2.36 Preparation of electro-competent cells.

This protocol makes enough for 40-60 transformations and can be scaled up or down as desired.

1.) Inoculate a single bacterial colony into approximately 5ml LB medium in sterile conditions. Culture overnight at 225-300rpm at 37°C.

2.) Inoculate 5ml of overnight culture into 400ml superbroth medium and incubate with shaking at 37°C until the absorbance at 600nm = (0.5).

3.) Place the culture on ice for 10-30 minutes then centrifuge at 4000g for 15 minutes at 4°C.

4.) Remove the supernatant and resuspend the pellet in 400ml of ice cold ultrapure water. Centrifuge at 4000g for 15 minutes at 4°C.

5.) Resuspend the pellet in 200ml of ice cold ultrapure water. Centrifuge at 4000g for 15 minutes at 4°C.

6.) Resuspend the pellet in 20ml of ice cold 10% (w/v) glycerol. Centrifuge at 4000g for 15 minutes at 4°C.

7.) Resuspend the pellet in 2-3ml of ice cold 10% (w/v) glycerol then keep on ice.

8.) Make 55µl aliquots of the electro-competent cells into 0.5ml microfuge tubes and freeze immediately on powdered dry ice. Store at -70°C.

2.37 Electroporation of the bacterium *Escherichia coli.*

The conditions for capacitor discharge have been empirically calibrated specifically for the bacterial strains used in this thesis and optimal conditions for other strains will vary from these.

1.) Thaw an aliquot of electro-competent cells on ice and chill an electroporation cuvette on ice prior to use.

2.) Add 10-100ng of the DNA to be transformed directly to the aliquot of electro-competent cells. Tap and incubate on ice for 1 minute.

3.) Transfer the mix to the chilled electroporation cuvette (EC2 cuvette ***Equibio-Flowgen***) and dry the outside thoroughly.

4.) Introduce the cuvette into the electroporator (***Equibio-EasyJect+***) and run the electroporation program with the following settings,

2500V, 25µF, 201Ω with a 5ms pulse time.

5.) Recover the cells immediately by adding 400µl of SOC medium.

6.) Incubate the cells at 37°C for 1 hour with shaking.

7.) Plate different amounts of the culture out on suitable selective LB agar plates.

8.) Incubate the plates overnight at 37°C.

2.38 Bacterial frozen stocks.

1.) Add either 0.5ml of a saturated overnight culture or 1.0ml of a mid-log phase culture to 0.5ml of glycerol strain storage mix.

2.) Mix carefully by pipetting and freeze immediately on powdered dry ice. Store long term at -70°C.

2.39 Cloning of DNA fragments into plasmid vectors.

Cloning of DNA sequences was achieved by ligating DNA fragments into suitably prepared plasmid vectors and the subsequent transformation of these ligation products into recipient strains of *E.coli.* The vectors pZErO™-2 (***Invitrogen***), pBluescript® II KS (+) (***Stratagene***) and pEGFP-1 (***Clontech***) were used in this project (see Materials 2.9).

2.39.1 De-phosphorylation of 5' ends.

When a non "zero background" vector is used in cloning it is essential to try to minimise the number of transformants that have no insert. This is negligible when "sticky" cloning is carried out with two different enzymes if they produce non-complementary cohesive ends on the vector. However, when cloning is attempted with the vector cut with only one enzyme, or with enzymes that create compatible cohesive ends, the number of empty clones can be large. This can be dramatically reduced by removing the 5' phosphate group from the cleaved ends of the vector as this prevents ligation between the two vector ends. The ligation of insert into this vector is not prohibited as the insert will still have 5' phosphate groups that can form covalent linkages on opposite strands of the cleaved vector ends. The phosphate groups are removed by the activity of Calf Intestinal Alkaline Phosphatase (CIAP).

1.) Add to the digested vector in a 0.5ml microfuge tube,

10.0µl 10x CIAP buffer

1.0-2.0µl CIAP (0.01U/pmol of vector ends)

and water to 100µl

pmol ends of vector = 3.04 x mass of DNA (µg)/size of vector (kb).

2.) Incubate,

At 37°C for 30 minutes for 5' protruding ends then add another 0.01U/pmol ends and incubate for a further 30 minutes at 37°C.

At 37°C for 15 minutes for 5' recessed or blunt ends then at 56°C for 15 minutes. Add another 0.01U/pmol ends and repeat the incubations again.

3.) Add 2.0µl of 0.5M EDTA and heat at 65°C for 10 minutes to stop the reaction.

4.) Clean up the DNA prior to use in a ligation reaction with a Wizard™ DNA column.

2.39.2 Polishing of PCR products.

Taq polymerase adds additional 3' A residues to the majority of PCR products (so called "TdT activity") which prevents their cloning into blunt cut vectors (Hu, 1993) and (Costa et al., 1994). The 3'-5' exonuclease activity of *PfU* polymerase (***Stratagene***) can be used to remove these additional residues and generate blunt products suitable for cloning ("polishing") (Costa and Weiner, 1994).

1.) Add 10-500ng of *Taq* generated PCR product into a 0.5ml microfuge tube in a volume of not more than 4.1µl. Make up to 4.1µl with ultrapure water if necessary.

2.) Add 0.3µl of 5mM dNTPs (250µM each) and 0.6µl *PfU* 10x reaction buffer.

3.) Place on ice and add 1.0µl *PfU* polymerase (2.5U/µl). Tap to mix and recover briefly in a microfuge. If the thermal cycler to be used doesn't have a 'hot-lid' to prevent evaporation of the reaction then overlay with a few drops of mineral oil.

4.) Incubate at 72°C for 30 minutes.

5.) The product can now be used directly in a ligation reaction with no clean up required. It may be stored for no more than a few days at -20°C.

2.39.3 Ligation reactions.

2.39.3(i) Blunt cloning of polished PCR products.

1.) Set up in a 0.5ml microfuge tube,

500ng pZErO-2 DNA

2.0µl 10x EcoRV digestion buffer

0.5µl EcoRV (10U/µl)

water to 20µl

incubate at 37°C for not longer than 30 minutes.

2.) Run the digest out on a 0.8% (w/v) agarose gel in 1xTBE then extract the band from a gel slice using the Qiaquick Gel Extraction kit (***Qiagen***). Elute the DNA in 50µl of elution buffer.

3.) Quantitate the purified vector and polished PCR product on a 1% (w/v) agarose gel in 1xTBE by comparing the band intensity to a known amount of 500bp Molecular weight marker XVI (***Boehringer-Mannheim***).

4.) In a 0.5ml microfuge tube set up on ice,

1.0µl 10x ligase reaction buffer
15ng EcoRV cut pZErO-2 vector
25ng Polished PCR product
0.67µl T4 DNA ligase (3U/µl) (***Promega***)
ultrapure water to 10µl.

The quantity of polished PCR product to include in the ligation is dependent on the ratio of insert : vector desired and the relative sizes of the vector and insert. This is calculated according to the equation,

$$\text{(amount of insert needed i} = \frac{\text{(ratio of insert : vect (size of insert in l (amount of vector in ng)}}{\text{(size in bp of}}$$

5.) Mix the ligation by tapping gently then incubate at 14°C overnight.
6.) Make volume up to 30µl with ultrapure water, add 3.4µl 3M sodium acetate (pH=5.3) and 83.4µl of ice cold absolute ethanol. Place on ice for at least 30 minutes.
7.) Centrifuge in a microfuge for 30 minutes at 4°C. Remove supernatant and add 150µl of 75% (v/v) ethanol.
8.) Centrifuge in a microfuge for 10 minutes at 4°C. Remove supernatant and air-dry for a few minutes. This ligation reaction can now be stored at -20°C until transformation.
9.) Prior to transformation, resuspend the pellet in 10µl of ultrapure water.

2.39.3(ii) Sticky cloning of DNA fragments.
1.) Digest 500ng of vector with the desired restriction enzyme(s) using the minimum incubation time recommended for total digestion. Treat with CIAP if not cloning into a "zero background" vector AND both sticky vector ends are complimentary to each other.
2.) Prepare the insert so that it has complimentary cohesive ends to the prepared vector.
3.) Gel purify both vector and insert and quantitate an aliquot of both next to a standard molecular weight marker.

4.) Set up in a 0.5ml microfuge tube on ice,

10ng of cut vector

enough insert for a 2:1 insert : vector ratio

1.0µl 10x ligation buffer

0.67µl T4 DNA ligase (3U/µl) (***Promega***)

water to 10µl.

5.) Incubate at 4°C overnight.

6.) Make volume up to 30µl with ultrapure water, add 3.4µl of 3M sodium acetate (pH=5.3) and 83.4µl of ice cold absolute ethanol. Place on ice for at least 30 minutes.

7.) Centrifuge in a microfuge for 30 minutes at 4°C. Remove supernatant and add 150µl of 75% (v/v) ethanol.

8.) Centrifuge in a microfuge for 10 minutes at 4°C. Remove supernatant and air-dry for a few minutes. This ligation reaction can now be stored at -20°C until transformation.

9.) Prior to transformation, resuspend the pellet in 10µl of ultrapure water.

2.40 PCR screening of bacterial clones.

Bacterial clones generated from transformed ligation reactions can be screened by PCR to see whether the vector they carry contains the correct insert. This is achieved either by using primers flanking the vectors multiple cloning site or by using primers specific for the intended insert.

1.) For each colony to be screened take a single isolated bacterial colony onto a pick. Streak lightly onto a fresh selective LB agar plate then vigorously shake the remainder of the colony from the pick into a 0.5ml microfuge tube containing 100µl of ultrapure water.

2.) Place the tube(s) into a boiling water bath and boil for 10 minutes to release the DNA.

3.) Centrifuge the tubes in a microfuge for 1 minute.

4.) Store the tubes at 4°C until ready to use.

5.) Make up a master mix for the number of colonies to be screened,

1x Reaction Mix,

2.5µl 10x PCR reaction buffer

0.625µl of forward primer (10µM)

0.625µl of reverse primer (10µM)

1.0µl of dNTPs (5mM)

0.04µl of *Taq* polymerase (5U/µl)

water to 15µl.

6.) Aliquot 15µl of master mix into a 0.5ml microfuge tube on ice.

7.) Aliquot 10µl of colony lysate supernatant into the tube. Mix by tapping gently and overlay with mineral oil if necessary

8.) Transfer to the thermal cycler and run the following program,

96°C for 2 minutes

96°C for 30 seconds }

50°C for 15 seconds } for 30 cycles

72°C for (30seconds/kb of product size) }

hold at 4°C.

9.) Run out the products on a 1% (w/v) agarose gel in 1xTBE.

2.41 DNA sequencing.

Clones and PCR products were sequenced by the dideoxy chain termination technique (Sanger et al., 1977) using the Sequenase version 2.0 DNA Sequencing kit (***USB***) (Tabor and Richardson, 1989) or the ABI BigDye™ dye terminator fluorescent cycle sequencing kit (***Perkin Elmer – Applied Biosystems***). Clones were typically sequenced with the vector multiple cloning site flanking primers SP6 (ATTTAGGTGACACTATAG) and T7 (GTAATACGACTCACTATAGGGC). PCR products were sequenced with the PCR primer(s) or internal primer(s) as appropriate.

2.41.1 Double stranded sequencing of plasmid DNA ("Manual" method).

2.41.1(i) Annealing.

1.) Denature isolated plasmid DNA by adding 1 volume of 0.4M NaOH/4mM EDTA to 3-5μg of DNA.

2.) Incubate at 37°C for 30 minutes.

3.) Add 1/10th volume of 3M sodium acetate (pH=5.2) and 2.5 volumes of ice cold ethanol. Mix and place at -70°C for 15 minutes.

4.) Centrifuge in a microfuge for 20 minutes.

5.) Remove the supernatant and add 250μl of ice cold 75% (v/v) ethanol, mix by inversion and then centrifuge in a microfuge for 5 minutes.

6.) Remove the supernatant and air dry the pellet.

7.) Resuspend the pellet in a mix of 2.0μl 5x reaction buffer, 1.0μl of 10μM primer and 7.0μl water. Anneal the primer to the denatured template by cooling from 65°C to 37°C over a 30 minute time period.

2.41.1(ii) Labelling.

1.) Dilute 5x dGTP labelling mix 1:4 in water.

2.) Dilute Sequenase 2.0 enzyme 1:7 in enzyme dilution buffer. Place on ice.

3.) Aliquot 2.5μl each of ddGTP, ddATP, ddTTP and ddCTP (80μM of each dNTP and 8μM of the ddNTP in 50mM NaCl) into four separate 0.5ml microfuge tubes labelled 'G', 'A', 'T' and 'C'. Pre-warm these tubes at 37°C.

4.) Add 1.0μl of 0.1M DTT, 0.5μl α^{35}S-dATP (***Amersham***) (10μCi/μl), 2.0μl diluted labelling mix and 2.0μl diluted sequenase to the annealed template. Mix by tapping gently, recover by spinning briefly, then incubate at room temperature for 3 minutes.

2.41.1(iii) Termination.

1.) Add 3.5μl of labelling reaction to the pre-warmed tubes containing the ddNTPs. Mix by tapping gently, recover by spinning briefly, then incubate at 37°C for 3 minutes.

2.) Add 4.0μl of stop solution (20mM EDTA, 95% (v/v) formamide, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanolFF). Mix by tapping gently, recover by spinning briefly, and then place on ice until ready to load. Samples can be stored frozen at -20°C.

2.42 Fluorescent cycle sequencing of DNA (Automatic sequencing).

Due to the high signal intensity provided by the ABI BigDye™ dyes (***Perkin Elmer – Applied Biosystems***) compared to previously used versions, the sequencing reactions were scaled down either by ½ or ¼. The protocol below represents a scaling down of ½. Plasmid DNA was prepared using the Qiaprep Spin Miniprep kit (***Qiagen***) (see Methods 2.22) according to the manufacturer's instructions. PCR products were cleaned up prior to sequencing by using the Qiaquick Gel Extraction kit (***Qiagen***) (see Methods 2.26) according to the manufacturer's instructions.

2.42.1 Cycle sequencing of DNA.

1.) Add 500ng of plasmid template or 250ng of a PCR product to a 0.5ml microfuge tube in a volume not greater then 5.0µl.

2.) Add 1.0µl of 10µM primer, 4.0µl of ready reaction mix and water to a final volume of 10µl.

3.) Vortex the mix and spin briefly in a microfuge to recover.

4.) Place the tube in a thermal cycler and run the following program;

96°C for 1 minute

96°C for 30 seconds }

50°C for 15 seconds } for 25 cycles

60°C for 1 minute }

then hold at 4°C

The annealing temperature can be varied to optimise for a particular primer. The 50°C annealing temperature works well for the SP6 and T7 primers flanking most vectors multiple cloning sites.

5.) Once cycling is complete add 2.0µl of 3M sodium acetate/EDTA (pH=4.6) and 50µl of ice cold absolute ethanol. Vortex the tube and place on ice for 10 minutes.

6.) Centrifuge the tube in a microfuge for 20-30 minutes.

7.) Remove all supernatant and rinse the pellet briefly in 250µl of 75% (v/v) ethanol.

8.) Vortex briefly and centrifuge in a microfuge for 10 minutes.

9.) Remove all of the supernatant and air-dry the pellet briefly.

10.) The pellet can be stored short term at -20°C prior to running on a sequencing gel.

2.43 Polyacrylamide gel electrophoresis of DNA sequencing reactions.

Manual DNA sequencing gels were run on a LKB 2010 Macrophor sequencing apparatus in the laboratory. Automatic sequencing gels were run by the Department of Biochemistry sequencing facility on a ABI-377 DNA Sequencer (***Perkin-Elmer Applied Biosystems***). Sequences were received from the facility by e-mail containing a text sequence interpretation of the fluorescent trace and an image file of the actual trace to allow visual resolution of unclear bases.

2.43.1 Manual sequence electrophoresis.

1.) Using EASI-GEL 40% stock acrylamide/*bis*-acrylamide (19:1) (***Anachem***) make a gel solution of 6% (w/v) acrylamide/*bis*-acrylamide (19:1), 7M Urea with 1xTBE buffer.

2.) Filter the solution and store in the dark at 4°C.

3.) Treat the notched sequencing plate with "Bind silane" (***LKB***) and the other with "Repelcote" (***BDH***) to allow the gel to bind only to the notched plate.

4.) Assemble the gel pouring apparatus by clamping 0.2mm spacers onto the bottom plate and aligning the notched plate on top.

4.) To catalyse gel polymerisation add 0.1% (v/v) TEMED (***Sigma***) and 1% (v/v) of 10% (w/v) ammonium persulphate to 50ml of 6% stock acrylamide and mix.

5.) Before allowing the acrylamide to set, pour quickly between the two plates on the apparatus, moving the notched plate down until it is aligned flush to the bottom plate.

6.) Insert the loading comb and then clamp the two plates together, fill the top and bottom of the apparatus with extra gel solution and leave to set for at least 1 hour.

7.) Mount the gel onto the sequencing apparatus and fill the top and bottom buffer wells with 1xTBE. Remove the comb and flush out the wells with 1xTBE.

8.) Attach the back plate and allow the gel to pre-run for 60 minutes at 2kV (≈ 38V/cm) in 1xTBE buffer.

9.) Switch off the power and once again flush out the wells with 1xTBE.

10.) Denature the samples to be run by incubating at 94°C for 2 minutes and then quench on ice.

11.) Using a fine horse-hair loading syringe or SeqPro™ loading tip (***Biorad***) load 2.0µl of sample into each lane, G, A, T and C.

12.) Run the gel at 2kV until the bromophenol blue line is at the bottom of the gel for a "short" run. Allow the gel to run until the xylene cyanolFF line is at the bottom of the gel for a "long" run.

13.) Once electrophoresis is complete remove the plates from the apparatus and separate them. The notched plate should retain the gel.

14.) Fix the gel in sequencing gel fixative for 30 minutes then dry it at 65°C for at least 2 hours.

15.) Expose the gel to Hyperfilm™- βmax (***Amersham***) at room temperature overnight.

16.) Develop the film manually to the desired intensity and contrast.

2.43.2 Automatic sequencing electrophoresis.

Automatic sequence reactions were run on acrylamide gels in a ABI-377 gel sequencer (***Perkin-Elmer Applied Biosystems***) according to the manufacturers instructions by the Department of Biochemistry "Core DNA Sequencing facility". The system uses four different dyes conjugated to the dideoxynucleotides each having different emission spectra when excited by the fixed wavelength laser of the ABI sequencer. This means that the normal four termination reactions of dideoxy sequencing can be carried out in the same tube and run in the same lane of the gel. The sequencer is able to separate out the signals from the different dyes and reconstruct the sequence data.

2.44 Polymerase Chain Reaction (PCR).

PCR is a technique for amplifying sections of DNA from very small quantities of starting material. Two short oligo-nucleotides (typically 16-24bp in length) are designed to be complimentary to opposite strands of the template and to delimit the region to be amplified. Repeated cycles of denaturation, primer annealing and DNA polymerase extension allow the amplification of the section of DNA between the two primer sites.

The convenient use of this procedure was made possible by the cloning of a thermostable DNA polymerase I from *Thermus aquaticus* (Chien et al., 1976). This facilitated successful PCR by repeated temperature cycling without destruction of the polymerase (Saiki et al., 1988).

2.44.1 Oligo-nucleotide (primer) design.

Most primer design was achieved by using the Primer3 program (ver1. ©1998 Whitehead Institute for Biomedical research http://www-genome.wi.mit.edu/cgi-bin/primer). Typically the user inputs the DNA sequence from which the primers are to be selected along with the target region, the preferred T_m of the primers and the product size range.

The program then suggests a list of potential primer pairs which have been screened to avoid repetitive material, high primer self- and inter-complementarity and high CG content amongst others. In situations where high CG content products were to be amplified, or PCR was on a bisulphite treated template, special primer design was undertaken and this is described separately in the relevant chapters. Primers were purchased from ***Genosys Biotechnologies (Europe) Ltd***. and arrived as dried pellets.

2.44.2 Primer stocks and storage.

Primers were resuspended in 200µl of ultrapure water, working stocks of 10µM were made from these and stored at -20°C short term. Long term storage of stocks was at -70°C or as dried pellets at -20°C.

2.44.3 Conditions for general PCR amplification.

A systematic method of PCR optimisation was adopted during this project. New primers were used in standardised PCR reactions using a series of 15 different PCR reaction buffers (see Materials 2.2). Typically 15 PCR reactions were set up as follows,

On ice,

1.) Label 15 sterile 0.5ml microfuge tubes A→O.

2.) Aliquot 2.5µl of 10x PCR reaction buffer into the relevant reaction tubes.

3.) Set up enough master mix for the PCRs,

1x master mix

1.25µl of forward primer (10µM)

1.25µl of reverse primer (10µM)

1.0µl of dNTPs (5mM)

0.2µl *Taq* polymerase (5U/µl)

genomic DNA to 5ng/µl

PCR water to 22.5µl

4.) Aliquot 22.5µl of master mix into each of the 15 microfuge tubes. Overlay with mineral oil if necessary.

5.) Place in a thermal cycler and run the following program,

96°C for 3 minutes

96°C for 1 minute }

Primer (T_m – 1.5°C) for 30 seconds } for 35 cycles

72°C for (30 seconds per kb product size) }

then hold at 4°C.

6.) Run the PCR products out on a 1% (w/v) agarose gel.

If there were no products using this approach the experiment was repeated with variations of the cycling procedure especially the annealing temperature. In addition various other components were added to the PCR reactions such as DMSO at 1-10% (v/v) and acetamide at 1-10% (w/v). In some cases it was necessary to re-design new primers or to use semi-nested or nested primer approaches.

If products were present in at least some of the different buffer lanes then further optimisation was performed on the lanes with the most promising results. One of the most effective ways of eliminating additional bands in these products is to repeat the PCR(s) with higher annealing temperatures and/or with a "hot-start". More complex band elimination can be effectively achieved by varying the primer : template ratio or by "touch-down" PCR. Touch-down PCR involves modifying the cycling program to start with early cycles having very high annealing temperatures and subsequent cycles having progressively lower annealing temperatures. This approach ensures very high primer stringency early in the PCR reaction where "mis-priming" is most prevalent.

Once a strong specific product is achieved, the PCR can usually be scaled up or down without any further modification being necessary.

2.45 First strand cDNA synthesis.

cDNA was made with the First-strand cDNA synthesis kit (***Pharmacia Biotech***).

1.) Aliquot 1-5µg of total RNA into 0.5ml microfuge tube and make up to 8µl with DEPC treated water.
2.) Incubate at 65°C for 10 minutes then chill on ice.
3.) Add 5.0µl of Bulk First-Strand cDNA reaction mix, 1.0µl DTT solution and 1.0µl of poly(dT) to a fresh 0.5ml microfuge tube.
4.) Add the denatured RNA mix and pipette to mix. Incubate at 37°C for 1 hour.
5.) The cDNA is now ready for use in an RT-PCR reaction or for long term storage at -70°C.

2.46 RT-PCR reactions.

The RT-PCR in this project has been by a two-step approach whereby first strand cDNA is made prior to the PCR reaction. As such, PCR on cDNA is generally less difficult to optimise than PCR on genomic DNA as the template is far less complex. It is generally sufficient to use a manufacturers recommended 10x PCR buffer and standard conditions as a starting point for optimisation. Typically,

On ice,
1.) Set up in a 0.5ml microfuge tube,

2.5µl of 10x PCR reaction buffer
1.25µl of forward primer (10µM)
1.25µl of reverse primer (10µM)
1.0µl of dNTPs (5mM)
0.2µl *Taq* polymerase (5U/µl)
2.0µl of first strand cDNA
16.8µl of PCR water

2.) Overlay with mineral oil if necessary. Place the tube in a thermal cycler and run the following program,

96°C for 3 minutes	
96°C for 1 minute	}
Primer (T_m-1.5°C) for 30 seconds	} cycle 35 times
72°C for (30 seconds per kb product size)	}
hold at 4°C.	

3.) Run the products out on a 1% (w/v) agarose gel. Optimise if necessary as suggested in 2.30 Polymerase Chain Reaction (PCR) (above).

2.47 Frommer bisulphite sequencing.

This procedure was described by (Clark et al., 1994); modifications were provided, amongst others, by (Feil et al., 1994) and confidentially by David Woodcock to the DNA Methylation Society (DMS).

Three protocols exist for successful bisulphite treatment of DNA,

2.47.1 Elevated temperature incubation.

1.) (Optional). Pre-digest the DNA to be used with one or more restriction enzymes flanking the target region. Or shear genomic DNA by passing through a fine gauge needle.

2.) Aliquot 500ng-5µg of DNA into a 1.5ml microfuge tube in a volume of 20-50µl, add $1/10^{th}$ volume of 3M NaOH and incubate at 37°C for 15 minutes.

3.) Add sodium bisulphite to a final concentration of 3.1M and hydroquinone to a final concentration of 0.5mM in a final total volume of not more than 500µl.

4.) Mix the sample gently without aerating, overlay with mineral oil, and incubate at 55°C for 16 hours in the dark.

5.) De-salt the DNA either by,

Dialysis at 4°C

(i) against 0.5mM hydroquinone/5mM sodium acetate (pH=5.3) for 1 hour (twice)

(ii) against 5mM sodium acetate (pH=5.3) for 1hour (twice)

(iii) against ultrapure water for 1 hour

(iv) against ultrapure water overnight

Or by clean up with a Wizard™ DNA column.

6.) Make the volume of de-salted DNA to 100-200µl. Add 1/10th volume of 3M NaOH and incubate at 37°C for 15 minutes.

7.) Add 1/10th volume of 3M sodium acetate (pH=5.3) or 1.5 volumes of 5M ammonium acetate (pH=7.5) and 2 volumes of ice cold absolute ethanol. Precipitate at -20°C for 30 minutes to overnight.

8.) Centrifuge in a microfuge for 30 minutes at 4°C. Remove the supernatant and add 250µl of ice cold 75% (v/v) ethanol and invert a few times to wash the pellet.

9.) Centrifuge in a microfuge for 15 minutes at 4°C. Remove the supernatant and vacuum dry for a few minutes.

10.) Resuspend in 20µl of either ultrapure water or 10mM Tris-HCl (pH=7.5). Store at -20°C.

2.47.2 Low temperature incubation.

1.) The DNA is treated as in 1.) and 2.) above.

2.) Add sodium bisulphite to a final concentration of 3.1M and hydroquinone to a final concentration of 1mM. Make to pH=5.0 with drops of 10M NaOH in a final total volume of not more than 500µl.

3.) Overlay with mineral oil and incubate in the dark at 4°C overnight.

4.) The DNA is cleaned up as in 4.) above.

5.) Add 1/10th volume of 2M NaOH, mix and incubate at 37°C for 20 minutes.

6.) Precipitate, wash and resuspend the DNA as in 7.) to 10.) above.

2.47.3 Thermal cycling modification.

1.) De-gas 100ml of ultrapure water either by boiling for 10 minutes or by vacuum until bubbles stop coming out of the water.

2.) Make stock solutions of 2M sodium bisulphite, 10mM hydroquinone and 10M NaOH with de-gassed water.

3.) Make the sodium bisulphite to pH=5.0 with drops of 10M NaOH.

4.) Aliquot 500ng-5µg of DNA into a 1.5ml microfuge tube and make the volume up to 50µl with de-gassed water.

5.) Add 1.5µl of 10M NaOH, tap to mix and incubate at 37°C for 15 minutes.

6.) Immediately add 2.5µl of 10mM hydroquinone and 450µl of 2M sodium bisulphite. Mix without aeration and aliquot 100µl into each of five 0.5ml microfuge tubes.

6.) Overlay with a few drops of mineral oil, place in a thermal cycler and run the following program,

95°C for 3 minutes }
50-55°C for 57 minutes } 5 times.

7.) Recombine the aqueous phases from the five tubes and de-salt the DNA by using a Wizard™ DNA column. Elute the DNA with 100µl of 10mM Tris-HCl (pH=7.5) at 55°C.

8.) Add 3.0µl of 10M NaOH to the eluted DNA. Mix and incubate at 37°C for 15 minutes.

9.) Add 150µl of 5M ammonium acetate (pH=7.0), mix by tapping and add 500µl of ice cold absolute ethanol.

10.) Precipitate at -20°C for 30 minutes to overnight.

11.) Centrifuge in a microfuge for 20 minutes at 4°C. Remove the supernatant and add 0.5ml of 75% (v/v) ethanol. Invert several times to wash the pellet.

12.) Centrifuge in a microfuge for 5 minutes at 4°C. Remove all supernatant and vacuum dry for a few minutes.

13.) Resuspend the treated DNA in 20µl of either 10mM Tris-HCl (pH=7.5) or TE buffer (pH=7.5). Store at -20°C.

2.48 Artificial methylation of promoter-constructs.

Promoter constructs were methylated using *Sss* I methylase (***NEB***) which methylated all CpG di-nucleotides under optimal conditions.

1.) The following reaction was set up in a 0.5ml microfuge tube

2.0µl 10xNEB2 reaction buffer

5.0µl *Sss* I methylase (2U/µl)

5.0µl SAM (S-adenosyl methionine) 1mM

5µg of vector DNA and water to 20.0µl

2.) Mix the reaction components by tapping the tube and collect by brief centrifugation.

3.) Incubate the reaction overnight at 37°C.

4.) Clean up the vector DNA post-methylation using the Wizard DNA Clean-Up system (***Promega***) (see Methods 2.27).

2.49 Tissue Culture Methods.

2.49.1 General tissue culture.

Cells were cultured in either RPMI-1640 or DME/F-12 working solutions containing Penicillin, Streptomycin, Glutamine and FCS (see Materials 2.13). Cells were incubated at 37°C in an atmosphere of 5% CO_2. All manipulations were carried out with either sterilised glassware or disposable sterile plastic-ware.

2.49.2 Trypsinisation of cells (For small or medium sized flasks).

1.) Pour off the medium from the cells. Add a few ml of PBS buffer rinse gently over the surface of the cells and pour off the PBS.

2.) Aspirate off any remaining PBS with a pasteur pipette.

3.) Add approx. 1ml of Trypsin solution and swirl gently over the surface of the cells.

4.) Quickly aspirate off the Trypsin solution with a pasteur pipette.

5.) Add approx. 1ml of Trypsin solution swirl gently over the surface of the cells and incubate at 37°C until the cells start to come off the culture flask when gently tapped.

6.) Add 2ml of culture medium to the flask and gently use a pasteur pipette to break up the cell clumps and wash the surface of the flask to maximise cell yield.

7.) Transfer all of the liquid into a 15ml centrifuge tube containing 8ml of culture medium. Mix gently by pipetting.

8.) The cells are now ready to be split into fresh medium containing flasks or frozen.

2.49.3 Making frozen stocks of cells.

1.) Take cells from step 8.) of the trypsinisation protocol and centrifuge at 1000rpm for 5 minutes.

2.) Aspirate medium and then tap the bottom of the tube to loosen the pellet. Resuspend the pellet gently in 8ml of PBS.

3.) Centrifuge the tube again at 1000rpm for 5 minutes. Aspirate the PBS taking off as much as possible. Tap the bottom of the tube to loosen the pellet.

4.) Resuspend the pellet gently in 1-2ml of freezing mix to a concentration of $3\text{-}5x10^6$ cells/ml.

5.) Transfer to cryotubes in 1ml aliquots and place at -70°C for a few hours.

6.) Transfer the tubes to a liquid nitrogen tank for long term storage.

2.50 Electroporation of tissue culture cells.

Tissue culture cells can be transfected with exogenous DNA by electroporation (Andreason and Evans, 1988).

1.) Take cells from Trypsinisation and centrifuge at 1000rpm for 5 minutes.

2.) Resuspend in a suitable volume of PBS and count an aliquot on a hemeocytometer.

3.) Centrifuge the cells at 1000rpm for 5 minutes and resuspend at a concentration of $2.5x10^7$ cells/ml.

4.) Add the DNA to be transfected (typically 1-5µg) to a 1.5ml microfuge tube on ice.

5.) Add 0.8ml of resuspended cells ($2.5\text{-}4x10^6$ cells) to the tube containing the DNA and flick several times to mix. Incubate on ice for 5 minutes.

6.) Transfer the cells to a 4mm gap electroporation cuvette (***Flowgen***) on ice.

7.) Dry the outside of the cuvette thoroughly then transfer to the electroporation machine (***Equibio EasyJect+***).

8.) Electroporate the cells at 250V and 1500µF (HeLa and ARPE-19).

9.) Immediately transfer to a 15ml centrifuge tube containing 8ml of medium.

10.) Split the recovering cells between wells of an 8 multi-well plate containing sterile cover-slips.

11.) Incubate at 37°C in an atmosphere of 5% CO_2 for 24-72 hours prior to fixation, mounting and visualisation.

2.51 Lipid based transfection with FuGene™6 transfection reagent.

An alternative to electroporation for the transfection of tissue culture cells is to use lipid reagents that form DNA/lipid conjugates (Felgner et al., 1993). FuGene™6 (***Boehringer-Mannheim***) contains a novel blend of lipids that do not form liposomes and which have a very low level of cytotoxicity. The reagent can be used in the presence of serum and yields a high level of transfection efficiency over a broad range of DNA concentrations. This makes FuGene™6 a very reliable and efficient reagent for transfection and it has been used successfully with every cell line in this project.

1.) On the day prior to transfection, plate out $1\text{-}3x10^5$ cells in a 35mm dish containing a sterile cover-slip (approx. 50-80% confluent).

2.) Aliquot 100µl of serum free medium into a sterile 0.5ml microfuge tube.

3.) Add 3.0-6.0µl of FuGene™6 to the serum free medium ensuring that the pipette does not touch the plastic side of the tube.

4.) Incubate at room temperature for 5 minutes.

5.) Aliquot the DNA to be transfected into a separate sterile 0.5ml microfuge tube.

6.) Add the FuGene™6/medium mix directly to the DNA and tap to mix. Incubate at room temperature for 15 minutes.

7.) Add the FuGene™6/medium/DNA mix directly to the 35mm dish and swirl to mix.

8.) Incubate the 35mm dish at 37°C in an atmosphere of 5% CO_2 for 24-48 hours.

9.) Fix the cells and visualise.

2.52 Making 4% Paraformaldehyde (PFA) fixative.

1.) Heat 30ml of water to 60°C and add 2g of paraformaldedyde and stir.

2.) Cover the container, transfer to a fume hood and place on a hot-plate at 60°C.

3.) Add 1drop of 2M NaOH with a pasteur pipette. Do not overheat the solution.

4.) Add 16.7ml of 3xPBS buffer and add HCl to pH=7.2.

5.) Make volume up to 50ml with water and filter through a 0.4µm filter.

6.) Cool to room temperature.

7.) Store short term at 4°C or -20°C long term.

2.53 Cold fixation of adherent cells.

1.) Aspirate culture medium from well using a pasteur pipette.

2.) Wash the cover-slip twice with ice cold PBS buffer.

3.) Aspirate all PBS buffer and then add enough drops of 4% (w/v) paraformaldehyde to cover the cover-slip completely.

4.) Incubate on an ice bath for 30 minutes.

5.) Aspirate all paraformaldehyde and wash the fixed cover-slip twice with room temperature PBS buffer.

6.) Aspirate all PBS buffer and mount the cover-slip with an aqueous mountant. Seal with rubber cement. Store short term at 4°C.

6a.) (Optional). Add propidium iodide to the aqueous mountant to a final concentration of 2µg/ml as a nuclear counter-stain.

2.54 Moviol aqueous mountant preparation.

1.) Weigh 6g of glycerol and 2.4g of Moviol 4-88 into a 50ml plastic conical centrifuge tube. Add 6ml of ultrapure water, mix and leave at room temperature for 2 hours.

2.) Add 12ml of 0.2M Tris-HCl (pH=8.5) and incubate in a water bath at 50°C, with occasional stirring, for 10 minutes or until everything has dissolved.

3.) Clarify the solution by centrifugation at 5000g for 15 minutes.

4.) Aliquot the supernatant into glass vials or a glass bottle and store at 4°C.

5.) Prior to use add counter-stain to the mountant if required.

2.55 Visualisation of slides.

Slides were viewed with a Zeiss Axioszop photo-microscope (***Zeiss***) under UV illumination. The microscope was fitted with FITC and Texas-Red filters, a CCD camera and an Apple Macintosh computer running the IPLab and IPLabExtensions image capture and analysis software package. All images were taken at the same exposure and magnification (as indicated next to figures) to allow their direct comparison. Images were captured and stored as high-resolution TIF compressed image files.

2.56 Computer programs.

Unless otherwise indicated, programs were used as part of the GCG-Package at the Human Genome Mapping Project – Resource Centre (HGMP-RC) under licence from the Genetics Computer Group (GCG) {Wisconsin Package Version 9.0, Madison, Wisconsin}. The HGMP-RC can be accessed on the web at the following URL http://www.hgmp.mrc.ac.uk.

2.56.1 Seqed.

Seqed is a sequence manipulation package for the editing and splicing of DNA sequences. It was used extensively in this project for the analysis of sequence data, the design of plasmid constructs and DNA probes. The program saves sequence files in the universal GCG file format which are used as input files for other packages in the GCG suite.

2.56.2 Gap.

Gap is a commonly used package for the alignment of two sequences. It considers all possible alignments of the two sequences and assigns an alignment as being optimal under pre-defined

constraint parameters (Feng and Doolittle, 1987). The program runs under a standard set of such parameters which can be modified as required. In this project the standard parameters have been used for all alignments so that the quality of alignments is directly comparable between experiments.

2.56.3 PileUp.

PileUp is a package to align multiple sequences using a simplification of the technique of Feng and Doolittle (Feng and Doolittle, 1987) similar to the method described by Higgins and Sharp (Higgins and Sharp, 1989). The package was used to align groups of bisulphite sequences from multiple independent clones.

2.56.4 Compare and DotPlot.

Graphical displays of similarity can be particularly useful for discerning patterns that are not otherwise easily observed. The "DotPlot" has been widely used as a method of showing similarity between both nucleic acid and protein sequences (Maizel and Lenk, 1981). The programs analyse the similarity between two sequences by calculating the number of matching residues between two 'windows' that are moved along the sequences. If the number of matches between the 'windows' satisfy a 'stringency' criteria a dot is plotted for that 'window' pair. When the 'windows' across one sequence are plotted against the 'windows' for the other sequence a 2-D graph is generated that contains every possible 'window' comparison. Any lines that appear on the graph represent regions of similarity.

2.56.5 CpG Plot.

CpG plots graphically represent the Cytosine and Guanine content of a section of DNA sequence. The sequence is analysed in windows of a specified size which are moved along the whole sequence at intervals of a specified size. At each window the absolute (C+G)% content and relative CpG di-nucleotide abundance (to that found on average in the genome for a window of the same size) are calculated. The relative CpG di-nucleotide abundance and (C+G)% content of the sequence are then plotted against the nucleotide position.

In order to identify sections of DNA that have uncommonly high (C+G)% and CpG content (so called "CpG islands") this method was applied to numerous sequences and an arbitrary definition of a "CpG island" was established by Gardiner-Garden and Frommer (Gardiner-

Garden and Frommer, 1987). The definition stipulates that a "CpG island" is a section of DNA with a (C+G)%≥ 50% and (CpG relative abundance)≥ 0.6 for an uninterrupted stretch of 200bp. The program CpG Plot identifies any such regions in an input sequence if one exists. In this project the raw data from the CpG Plot program was extracted and subjected to a more comprehensive analysis using Microsoft® Excel97.

2.56.6 SignalScan.

SignalScan takes an input sequence such as a putative gene promoter sequence and searches it against TFD (Transcription Factor Database) and Transfac, two transcription factor databases. The databases contain consensus binding sites for transcription factors and transcription associated factors. The program identifies if the input sequence contains any known consensus sites and if present outputs their position and a relevant journal reference (Prestridge, 1991).

2.56.7 Blast (Basic Local Alignment Search Tool).

Blast uses the method of Altschul (Altschul et al., 1990) to search databases for similarities to query sequences. An interactive web based BLAST program at the HGMP-RC was used during this project.

Chapter 3 - Methylation analysis of the Human Pyruvate dehydrogenase E1α gene promoter

3.1 Introduction.

The work described in this chapter involves the analysis of DNA methylation at the promoter of the housekeeping pyruvate dehydrogenase E1α gene. Analysis was carried out on DNA from both tissue culture cells and male and female human lymphocyte samples. The section of the promoter studied was cloned upstream of the EGFP reporter gene and transfected into tissue culture cells to determine whether it contained promoter activity. These constructs were further used to investigate the effect of *ex-vivo* methylation of the promoter section on reporter gene expression in tissue culture cells.

3.1.1 Housekeeping genes on the X-chromosome.

The promoters of housekeeping genes on the X-chromosome share the same features as those of autosomal genes. Typically the promoters are associated with CpG islands at their 5' ends and have proximal promoter sequences rich in consensus binding sites of common transcription factors (Bird, 1987; Bird, 1986). Many housekeeping gene promoters do not contain the TATA and CAAT box binding sites commonly associated with genes transcribed by RNA Polymerase II. Instead they frequently contain binding sites for other common transcription factors such as Sp1 and AP-2 which act on TBP (the TATA binding protein) to enhance the rate of transcription most likely by increasing the rate of formation of the transcription initiation complex (Emili et al., 1994; Williams et al., 1988).

The promoters of a number of X-linked housekeeping genes have been studied in detail at the molecular level including those of the hypoxanthine phosphoribosyl transferase (*HPRT*) (Melton et al., 1986), phosphoglycerate kinase (*PGK-1*) (Pfeifer et al., 1990), pyruvate dehydrogenase E1α (*PDHA1*) (Chang et al., 1993) and X-inactive specific transcript (*XIST*) genes (Hendrich et al., 1997; Pillet et al., 1995; Sheardown et al., 1997a). These studies and others on house keeping gene promoters have indicated that only very small sections of the proximal promoter sequences are required for promoter activity, typically as little as only 50-100bp. Further enhancement of promoter activity is achieved by the binding of transcription factors to enhancer elements upstream or downstream of the transcription start site.

Another common feature of housekeeping gene promoters is the presence of small GC-rich direct repeats in the minimal promoter in which there are frequently many binding sites for transcription factors. These regions can even be conserved between different gene promoters as in the case of the *HPRT* and *PGK-1* genes in which an 8bp direct repeat array is 80% conserved (Melton et al., 1986). Individual deletion of one of these sites is found to be functionally compensated for by transcription factor binding at another nearby site. Only when both sites are deleted is the transcription level reduced. This has been demonstrated at specific Sp1 binding sites in the promoters of the *HPRT* (Melton et al., 1986) and *XIST* (Hendrich et al., 1997) genes and implies that in housekeeping gene promoters there is a degree of flexibility in the specific position of the transcription factors used to enhance the rate of transcription initiation. In addition, many housekeeping gene promoters have been shown to bind different transcription factors in different tissues implying that there is not only flexibility in the position of the sites, but also in the actual nature of the proteins involved. This has been demonstrated for several house-keeping gene promoters including those of the *PDHA1* (Chang et al., 1993; Tan et al., 1998), aspartate aminotransferase (Toussaint et al.,

1994), ferrochelatase (Tugores et al., 1994) and Htf9-associated genes (Somma et al., 1991). This flexibility presumably allows the housekeeping genes to maintain their ubiquitous expression in cells expressing different complements of transcription factors.

There are some features of housekeeping genes that are unique to those of the X-chromosome. The CpG islands of X-linked genes are hyper-methylated on the inactive X-chromosome of female cells and hypo-methylated on the active X-chromosome. Two studies of CpG islands on the X-chromosome showed that 11/13 mouse and 28/28 human islands were exclusively methylated on the inactive X-chromosome (Norris et al., 1991; Tribioli et al., 1992). This methylation has been implicated in the inactivation of the genes on the inactive X-chromosome and in the stable maintenance of inactivation through cell division, forming a crucial part of the X-inactivation process. The precise effect of methylation on the promoters varies by gene, but has been shown to include ablation of transcription factor binding and the recruitment of specific methyl-CpG binding proteins to the sites of methylation (Heard et al., 1997; Kass et al., 1997).

Several housekeeping genes on the X-chromosome have been found to have additional functional homologues elsewhere in the genome. Autosomal copies of the *PDHA1* (Dahl et al., 1990), *PGK-1* (McCarrey and Thomas, 1987) and *G6PD* (Matsuzawa, 1991) genes have been described. These homologues are uniquely expressed in the testis during the process of spermatogenesis in which the haploid post-meiotic spermatocyte either lacks an X-chromosome or has one which is inactivated (Lifschytz and Lindsley, 1972). These autosomal genes have presumably evolved to ensure that essential gene products required for normal spermatogenesis are expressed even when the genes that encode them are X-linked. The post-meiotic spermatocyte undergoes a four week maturation process which is dependent

upon aerobic carbohydrate metabolism (Perey et al., 1961). Expression of pyruvate dehydrogenase, phosphoglycerate kinase which are enzymes of the glycolytic pathway and glucose-6-phoshate dehydrogenase an enzyme of the pentose phosphate shunt is likely to be essential for spermatogenesis to take place. The evolution of autosomal copies of X-linked genes to complement for their absence or inactivity in spermatocytes is a feature of a small number of housekeeping genes on the X-chromosome. It is possible, however, that other X-linked genes required during spermatogenesis also have functional autosomal copies, which have yet to be discovered.

3.1.2 The Pyruvate dehydrogenase E1α gene.

The pyruvate dehydrogenase complex (PDH) converts pyruvate to acetylCoA, an essential process of aerobic glucose metabolism. The complex is composed of multiple copies of three enzymatic units E1(comprising α and β subunits), E2 and E3, the two regulatory components El kinase and phospho-El phosphatase and the non-catalytic protein X (Fouque et al., 1998; Randle et al., 1978). The active site of the PDH complex (the thiamine pyrophosphate site) is shared between the PDH E1α and β-subunits and is primarily regulated by phosphorylation/de-phosphorylation (Linn et al., 1969). Genetic defects of the PDH complex are one of the most common causes of lactic acidosis in childhood with clinical presentations ranging from lethal lactic acidosis to mild lactic acidemia (Robinson et al., 1996). In the majority of cases the defect in the PDH complex lies in the E1 subunit which is an $\alpha_2\beta_2$ tetramer. The genes for both the E1α (Dahl et al., 1987) and E1β (Chun et al., 1990) subunits have been cloned. Mutation analysis in patients has shown that the majority of mutations are in the E1α gene (*PDHA1*). The *PDHA1* gene was localised to Xp22.1-22.2 (Brown et al., 1989) and found to be subject to X-chromosome inactivation (Brown et al., 1990). The variation in clinical presentation of human pyruvate dehydrogenase deficiency has been

explained in part by the effect of X-inactivation ratios on the level of wild type or mutant gene expressed in the tissues of patients (Brown et al., 1990).

The *PDHA1* gene comprises eleven exons and spans approximately 17kb of the genome (Maragos et al., 1989). It is associated with a 5' CpG island and has a promoter rich in consensus binding sites for common transcription factors such as Sp1, AP-2 and CRE (see Figure 3.1). The first major study of the promoter of the *PDHA1* gene was carried out by (Chang et al., 1993) in which sections of the promoter were assayed for their ability to drive chloramphenicol acetyltransferase activity in expression constructs transfected into tissue culture cells. In addition the study detailed DNase I footprinting experiments testing the positions at which proteins bound to the promoter of the gene *in-vitro*. The paper reported that the promoter of the *PDHA1* gene exhibited features of both a housekeeping and a tissue-specific gene although these features were determined solely in tissue-culture studies. These findings are used to describe and analyse the methylation results for the *PDHA1* promoter in this chapter, and their relevance to the control of the genes expression is considered in the discussion.

3.2 Results.

3.2.1 CpG-Plots.

Genomic DNA methylation in mammals occurs predominantly at CpG di-nucleotides (Woodcock et al., 1987) and exclusively on the 5-carbon of cytosine rings (5Me-C). In order to plan a strategy for the analysis of potential methylation at the promoter it was necessary to characterise the position and density of CpG di-nucleotides. This was achieved by using the CpG-Plot computer program which is part of the GCG-Wisconsin suite (see Methods 2.56

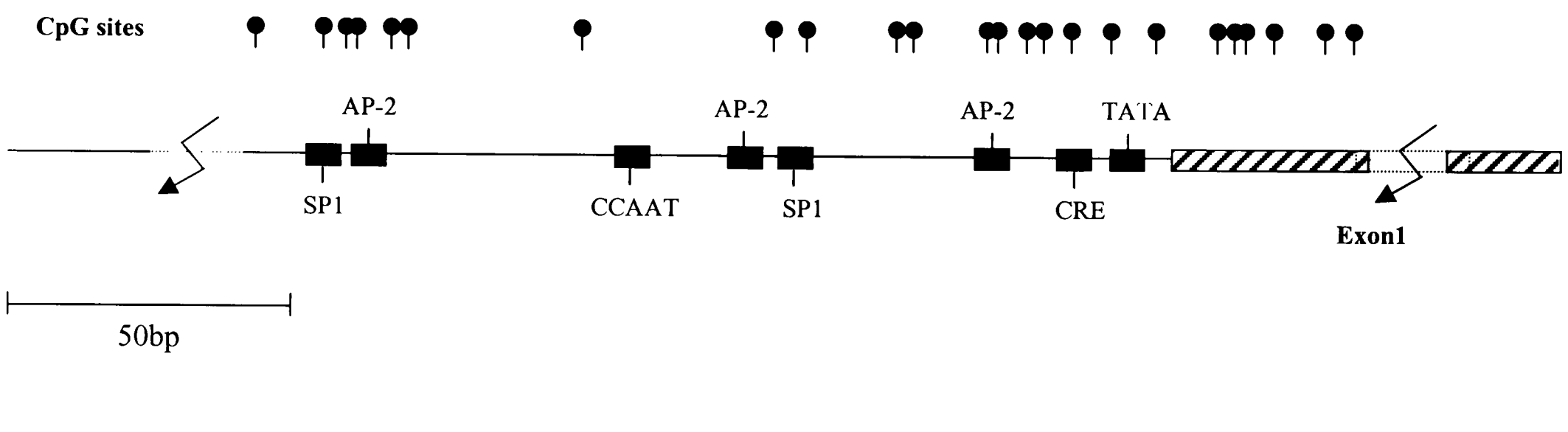


Figure 3.1: Schematic of the Human PDHE1α gene proximal promoter region.
The gene promoter region contains many of the transcription factor consensus sites found in the promoters of widely-expressed genes. A number of sites are indicated by small filled boxes and represent: TATA box (TATA) (-5), Sp1 binding sites (SP1) (-148 and -64), AP-2 binding sites (AP-2) (-140, -73 and -29), CAAT box (CCAAT) (-93) and cAMP responsive element (CRE) (-10). The transcription initiation site and the start of exon1 (shown as a striped box) are as defined by Maragos et al. (1989). The 200bp region shown has a (C+G)% composition of 70.5% and contains 24 CpGs (represented as batons) and is part of a large CpG island for this gene.

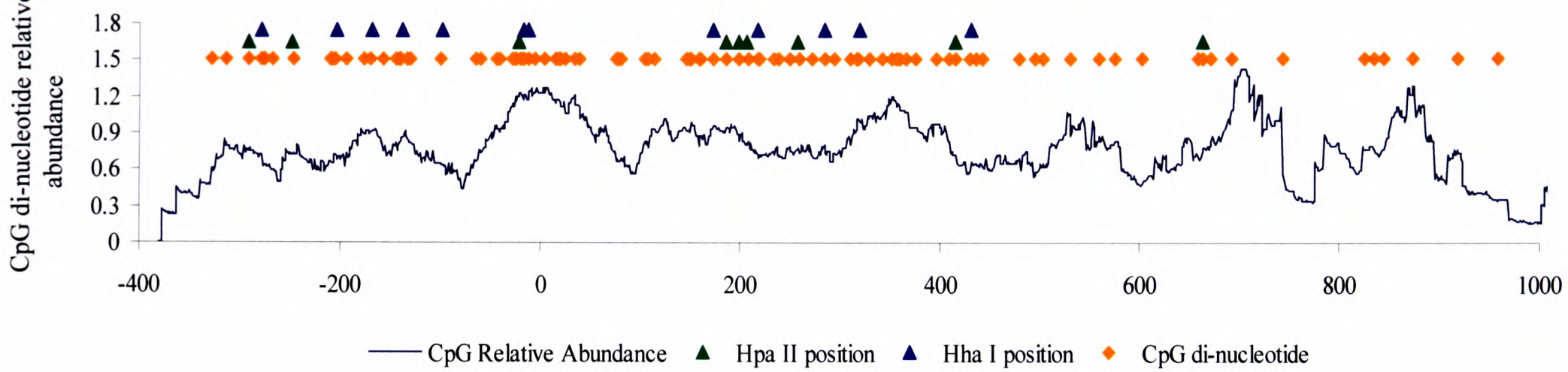
and 2.56.5). The program provides a graphical output which is not well suited to detailed analysis so the output was re-directed to a text file. This data was then edited and imported into Microsoft®Excel97 to allow the plotting of customised graphs showing the positions of individual CpG sites and also of *Hpa II* and *Hha I* restriction endonuclease sites.

In all parts of this chapter positions on the *PDHA1* gene are given relative to the transcription start site as defined by Maragos et al. (1989). Figure 3.2 shows a CpG-plot of a 1.4kb (-400 to +1000bp) section of the promoter which contains 95 CpG di-nucleotides, 9 *Hpa II* and 12 *Hha I* sites. The section contains an 858bp CpG island (-339 to +519bp) with a (C+G) content of 68.4%, 8 *Hpa II* sites, 12 *Hha I* sites and 80 CpG di-nucleotides. Figure 3.3 shows another CpG-Plot of a smaller section of the promoter (-333 to +61bp) which includes the 254bp fragment (-189 to +65bp) that is thought to define the core promoter (Chang et al., 1993). The two graphs were used to help design methylation sensitive PCR (msPCR) experiments and to select the target region to be used in bisulphite sequence analysis.

3.2.2 Methylation sensitive PCR (msPCR).

Primers were designed to amplify two overlapping sections of the promoter sequence. The first or "upstream msPCR" amplifies a 590bp section (-414 to +176bp) containing 3 *Hpa II* and 8 *Hha I* sites using the primers PDHA1F1 (5'-GCAGACCATTTACAGAAACCAG-3') and PDHA1EX1REV (5'-TCTGAGAAGCGCCAGACAG-3'). The second or "downstream msPCR" amplifies a 1912bp section (+18 to 1930bp) containing 6 *Hpa II* and 10 *Hha I* sites using the primers PDHA1EX1F (5'-GACTGAGGCGTGGCGTCT-3') and PDHA1R1 (5'-TTACACACCAAGTGGAAAGATACC-3'). The conditions used for the msPCR reactions are shown in Table 3.1 (below).

A.)



B.)

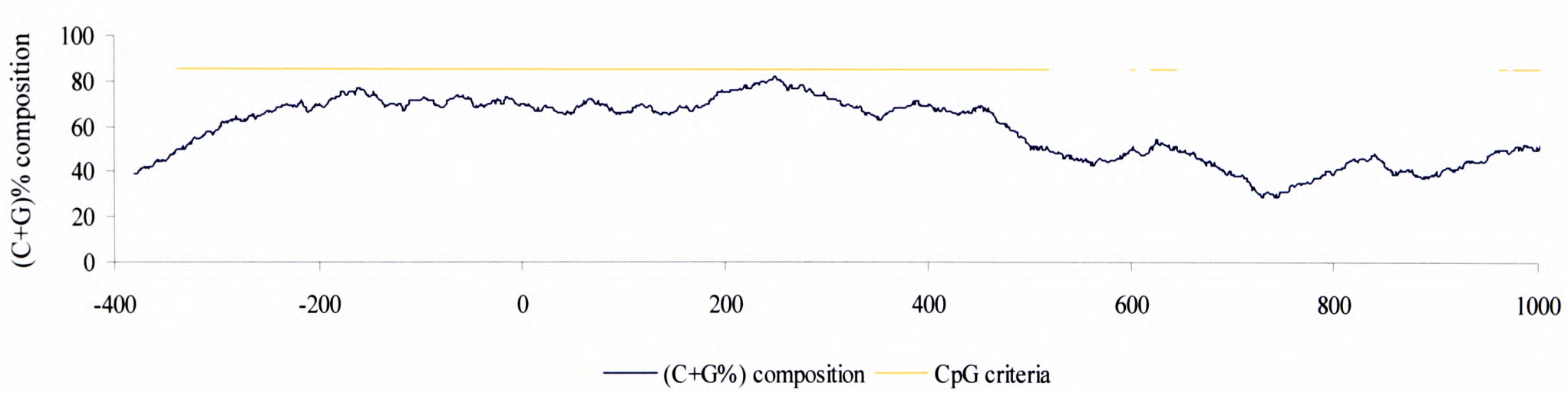


Figure 3.2: CpG-Plot of the 5' end of the Human Pyruvate dehydrogenase E1-α gene (PDHE1A).
The position of 95 CpG sites, 9 HpaII sites and 12 HhaI sites are shown (A). Regions for which the CpG relative abundance is ≥0.6 and the (C+G)% composition is ≥50% are shown in the 'CpGcriteria" line (B). Transcription start is as defined by (Maragos et. al. 1989).

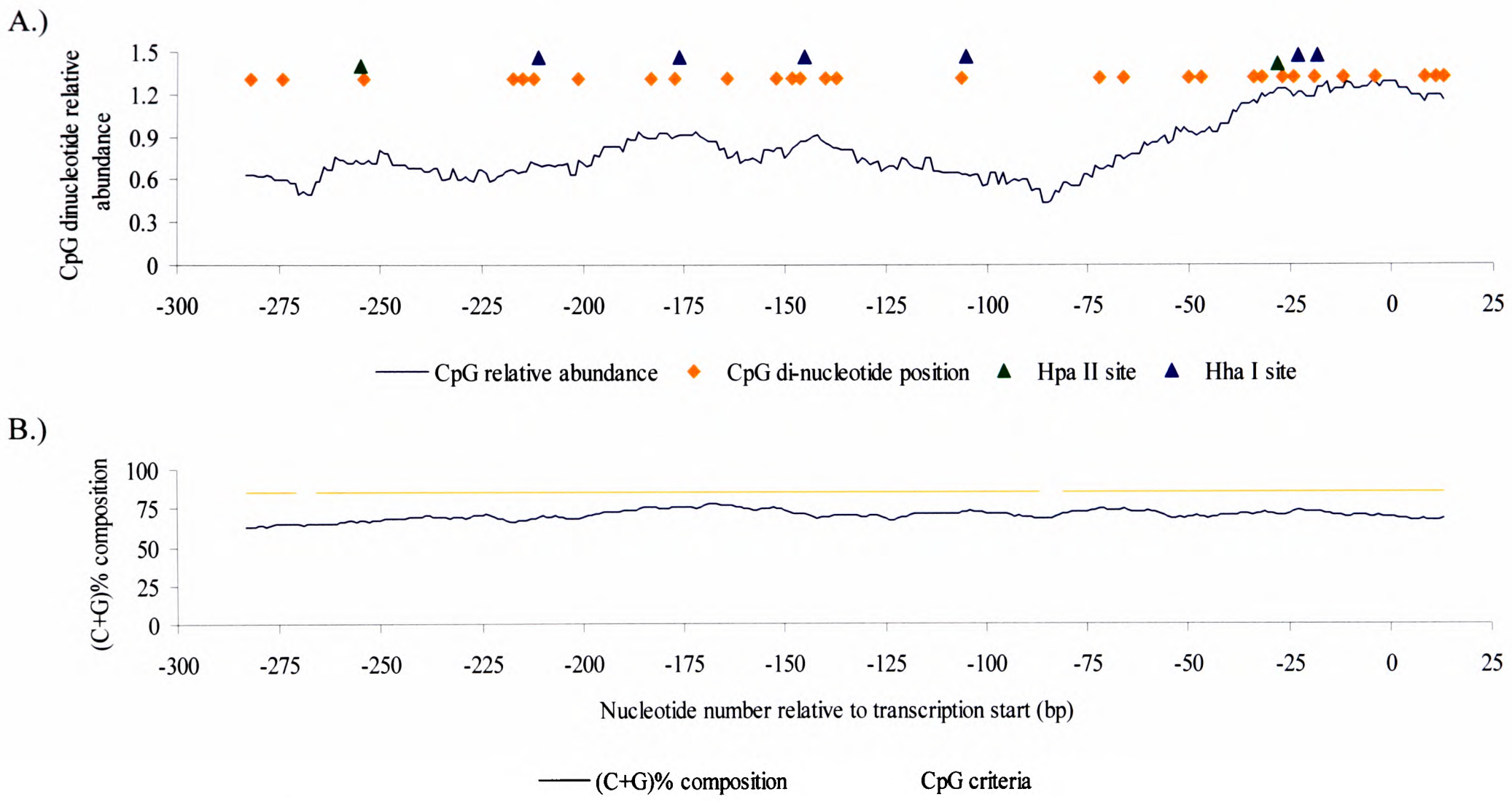


Figure 3.3: CpG-Plot of a sub-section of the Human Pyruvate dehydrogenase E1-α gene (PDHE1A) promoter.
The position of 30 CpG sites, 2 HpaII sites and 6 HhaI sites are shown (A). Regions for which the CpG relative abundance is ≥0.6 and the (C+G)% composition is ≥50% are shown in the 'CpGcriteria" line (B). Transcription start is as defined by (Maragos et. al. 1989).

Upstream msPCR (590bp product) PDHA1F1 : PDHA1EX1REV Qiagen buffer + Q solution	96°C for 2'00'' 96°C for 0'30'' 58°C for 0'30'' 72°C for 0'30'' } 35 cycles
Downstream msPCR (1912bp product) PDHA1EX1F : PDHA1R1 Qiagen buffer + Q solution	96°C for 2'00'' 96°C for 0'30'' 60°C for 0'30'' 72°C for 1'00'' } 35 cycles

Table 3.1: Primers and conditions for the upstream and downstream *PDHA1* msPCR reactions.

i) Methylation sensitive PCR on HeLa cell DNA.

HeLa cell DNA was digested with either *Msp I*, *Hpa II* or *Hha I* and the digests purified by ethanol precipitation. Methylation sensitive PCR was carried out using the upstream and downstream primer sets to determine the methylation status of the two blocks of *Hpa II* and *Hha I* sites. The upstream msPCR reactions show that there are target molecules that are completely methylated at all *Hpa II* and *Hha I* sites (see Fig3.4). The downstream msPCR reactions show that some of the *Hpa II* and *Hha I* sites are unmethylated in all molecules (see Fig3.5). These results suggest that the 590bp section of the promoter contains methylated residues in some molecules of HeLa cell genomic DNA.

ii) Methylation sensitive PCR on lymphocyte DNA.

In order to assess the methylation status of the promoter in DNA from uncultured cells, DNA was isolated from peripheral blood donated by 10 males and 10 females. Figure 3.6 shows the results from the upstream msPCR reactions performed on 4 male and 5 female DNA samples. In the male samples no products were observed in the *Hpa II* lanes implying that at least one of the three *Hpa II* sites is unmethylated in all target molecules. In the female samples only 3 out of 5 produced products in the *Hpa II* lanes. This implies that methylation is

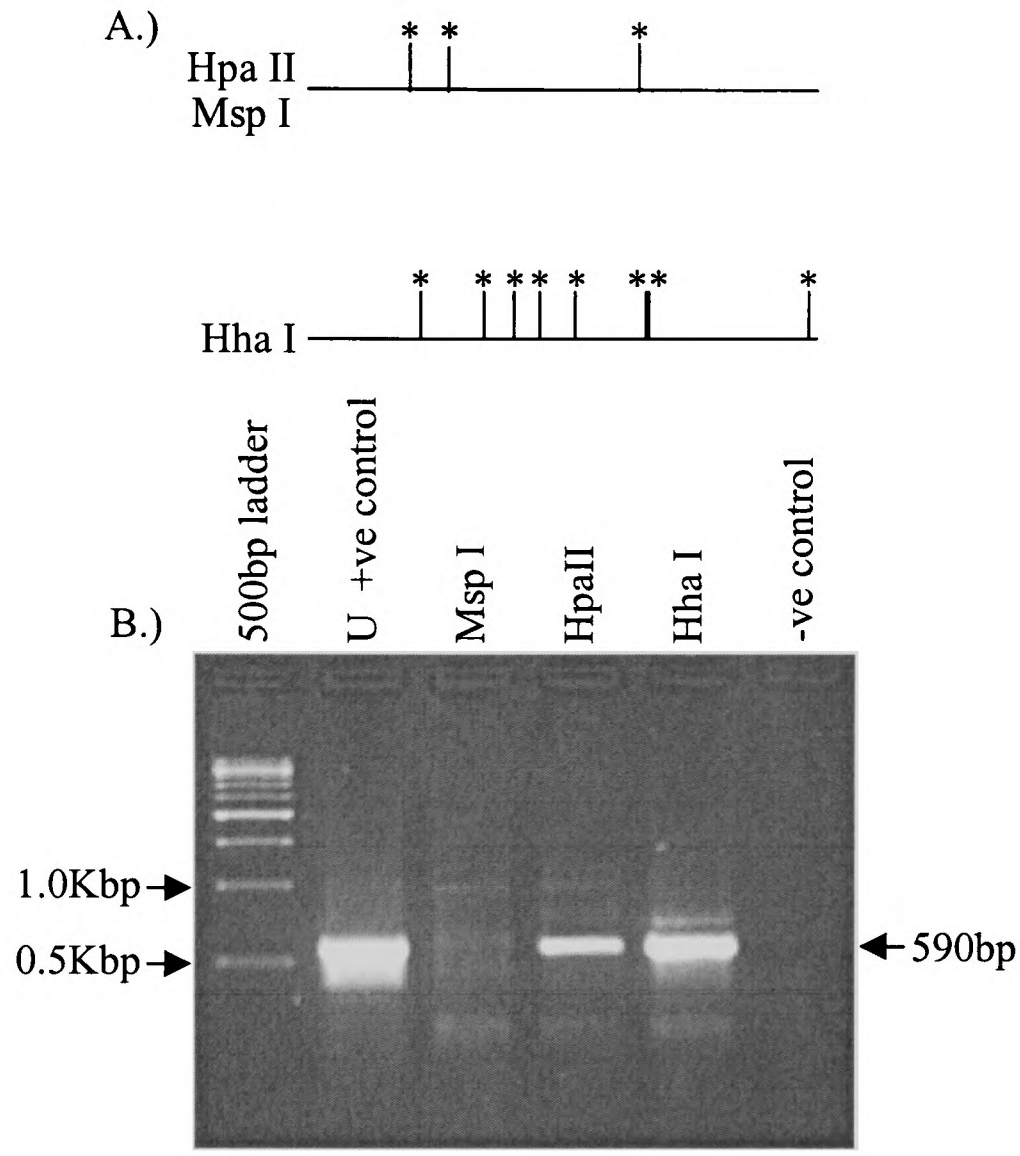


Figure 3.4: Upstream methylation sensitive PCR (msPCR) of the human PDHE1α gene using HeLa cell DNA.

Methylation sensitive PCR (msPCR) was used to determine the bulk methylation status of a region delimited by flanking PCR primers. Genomic DNA was pre-digested with Dra I flanking the target region to facilitate total digestion in subsequent restriction digests. Aliquots of the pre-treated DNA were digested with either Msp I, HpaII or Hha I, the digests were then purified and 125ng of each used in msPCR reactions. The primers PDHA1F1 (5'-GCAGACCATTTACAGAAACCAG-3') and PDHA1EX1REV (5'-TCTGAGAAGCGCCAGACAG-3') were used to amplify a section from -414 to +176bp including 3 MspI/Hpa II and 8 HhaI sites (A). Hpa II and Hha I are restriction enzymes that only cut their target sequences when they are unmethylated, so PCR products will only be generated in DNA pre-digested with them if ALL sites are methylated. Conversely, as MspI cuts its target site irrespective of methylation status it serves as a control for potential to totally digest the template. HeLa cell DNA was used in msPCR reactions and showed that in some target molecules the Hpa II and Hha I sites are totally methylated (B).

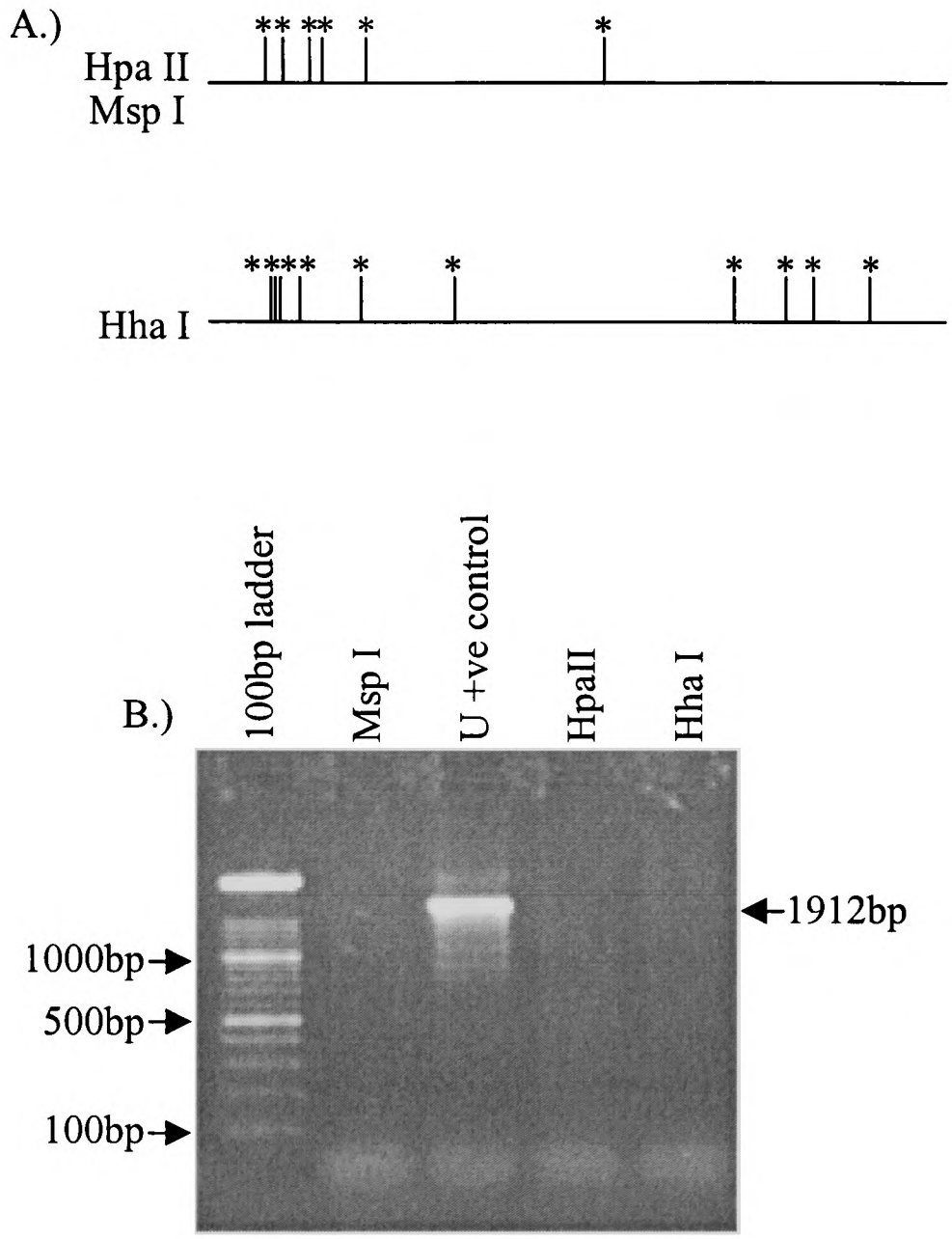


Figure 3.5: Downstream methylation sensitive PCR (msPCR) of the Human PDHE1α gene using HeLa cell DNA.
Methylation sensitive PCR (msPCR) was used to determine the bulk methylation status of a region delimited by flanking PCR primers. Aliquots of Genomic DNA were digested with either Msp I, HpaII or Hha I, the digests were then purified and 125ng of each used in msPCR reactions. The primers PDHA1EX1F (5'-GACTGAGGCGTGGCGTCT-3') and PDHA1R1 (5'-TTACACACCAAGTGGAAAGATACC-3') were used to amplify a section from +18 to +1930bp including 6 MspI/Hpa II and 10 HhaI sites (A). HeLa cell DNA was used in msPCR reactions and showed that both the Hpa II and Hha I sites are either un-methylated or partially methylated in the target molecules (B).

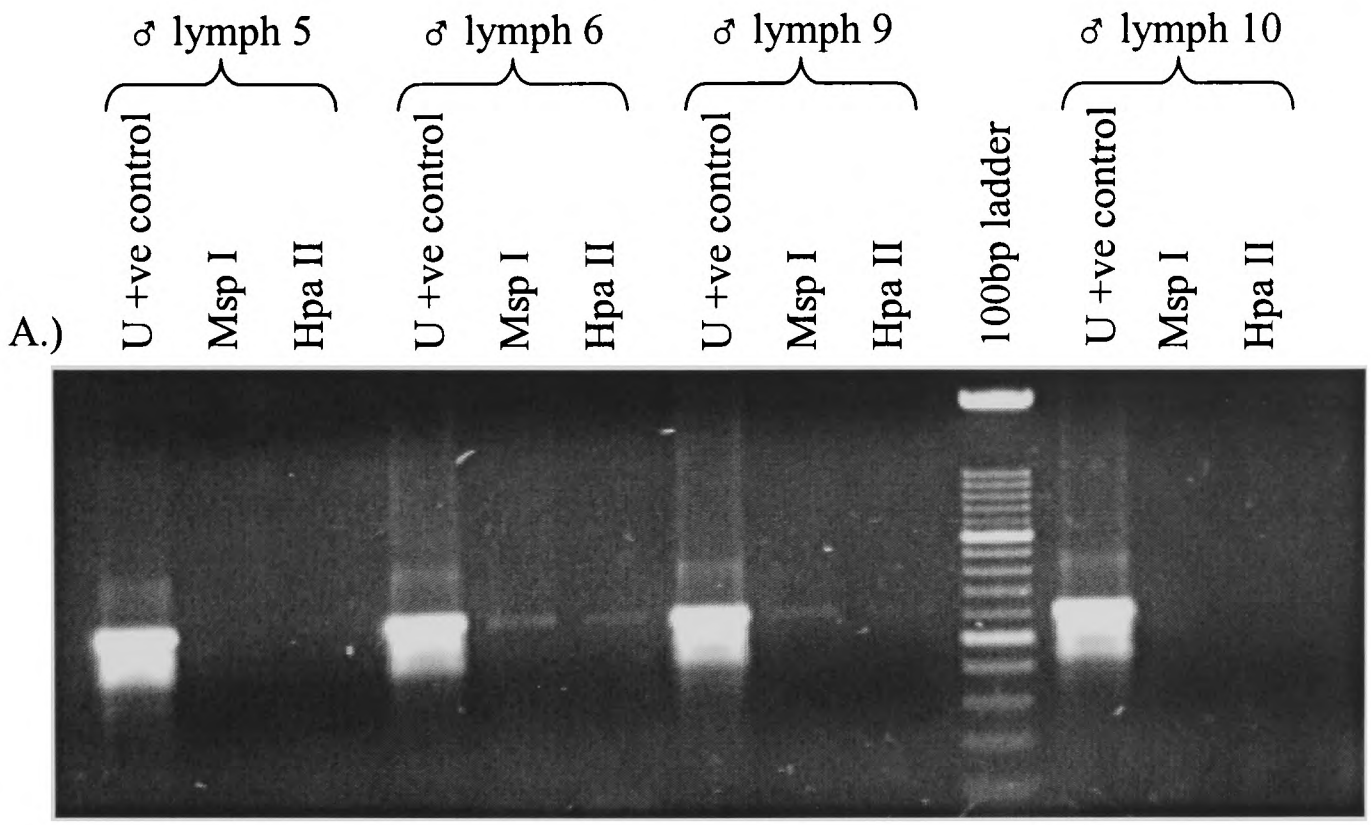


Figure 3.6: Upstream methylation sensitive PCR (msPCR) of the Human PDHE1α gene using human primary peripheral lymphocyte DNA.
Aliquots of peripheral lymphocyte DNA from multiple male and female donors were pre-digested with Dra I and then with either Msp I or HpaII. The digests were purified and 125ng of each used in msPCR reactions. The primers PDHA1F1 (5'-GCAGACCATTTACAGAAACCAG-3') and PDHA1EX1REV (5'-TCTGAGAAGCGCCAGACAG-3') were used to amplify a section from -414 to +176bp spanning 3 MspI/Hpa II sites. Four male samples were analysed by msPCR and no product bands were observed in the Hpa II lanes, indicating that the male target molecules are either totally un-methylated or partially methylated (A). Five female samples were analysed by msPCR and exhibited heterogeneity (B). Samples 1 and 3 showed partial or no methylation and samples 7, 8 and 11 total methylation at Hpa II sites in some target molecules.

heterogeneous at the three *Hpa II* sites between female samples, but some females have target molecules in which all three *Hpa II* sites are methylated.

iii) Methylation sensitive PCR on sperm and chorionic villus DNA.

Sperm cells do not express the *PDHA1* gene as the X-chromosome of sperm is inactive (Lifschytz and Lindsley, 1972). It is possible that this inactivity could be correlated with methylation at the promoter of the gene. In order to test the possibility that the promoter is methylated in sperm cells, msPCR was performed on sperm DNA from a patient donor. In addition, to test the methylation status of the promoter in another example of non-somatic tissue, chorionic villus cells from another male patient were subjected to msPCR. Figure 3.7 shows that in the male CVS sample all of the *Hha I* sites are methylated in some target molecules and that in the sperm sample all *Hpa II* and *Hha I* sites are methylated in some target molecules. These results suggest that the inactivity of the *PDHA1* gene in sperm may be correlated to promoter methylation. The observation that there is some methylation at the promoter of *PDHA1* in male chorionic villus cells suggests that extra-embryonic tissue retains methylation that is most likely not retained in male somatic cells.

3.2.3 Bisulphite sequencing.

In order to establish the precise methylation profiles at the promoter of the *PDHA1* gene in HeLa cells and peripheral lymphocytes the bisulphite genomic methylation sequencing technique was used (Frommer et al., 1992) (Methods 2.47).

i) Primer design for bisulphite PCR.

Primers were designed for analysis of methylation at the promoter of the *PDHA1* gene so as to include the transcription start site, the core promoter as defined by (Chang et al., 1993) and

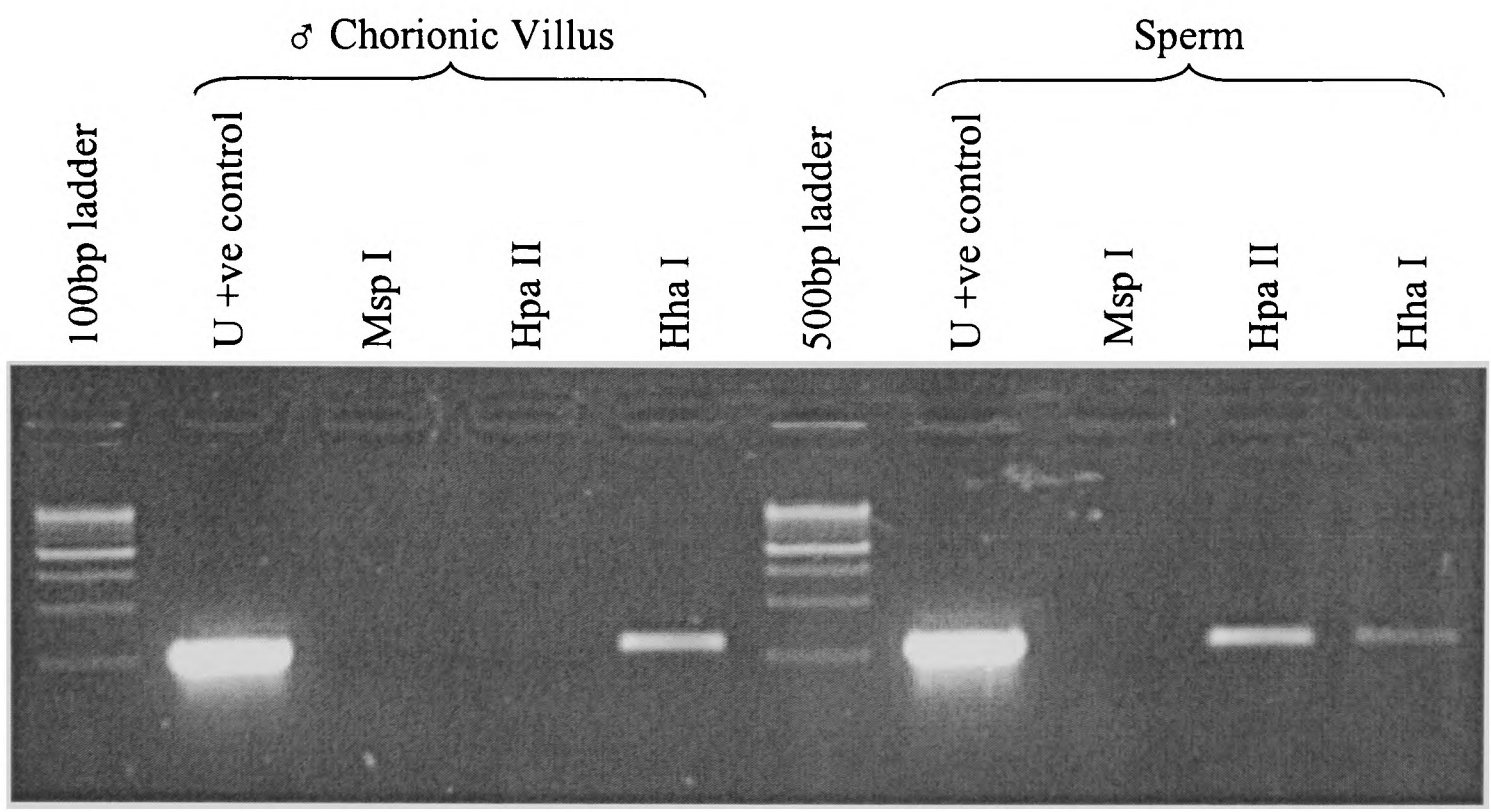


Figure 3.7: Upstream methylation sensitive PCR (msPCR) of the Human PDHE1α gene using human chorionic villus and sperm DNA.
DNA was prepared from sperm and a male chorionic villus sample from patient donors. Aliquots of the DNA were pre-digested with Dra I and then digested with either Msp I, Hpa II or Hha I. The digests were purified and 125ng of each used in msPCR reactions. The primers PDHA1F1 (5'-GCAGACCATTTACAGAAACCAG-3') and PDHA1EX1REV (5'-TCTGAGAAGCGCCAGACAG-3') were used to amplify a section from -414 to +176bp spanning 3 Msp I/Hpa II and 8 Hha I sites. In the chorionic villus msPCRs the Hpa II sites are either un-methylated or partially methylated whereas in some of the target molecules the 10 Hha I sites are totally methylated. In the sperm msPCRs the Hpa II and Hha I sites are totally methylated in some target molecules.

as much further upstream sequence as could be amplified. The nature of the bisulphite template restricts the sites at which primers can be positioned. Primers also need to be designed to allow unbiased amplification of target molecules whatever their methylation status. The following criteria were applied to the design of primers:

a) Primers were of 18-24 bases in length with similar annealing temperatures.

b) CpG di-nucleotides were avoided if possible and were never positioned near the 3' end of the primer

c) Forward primers were designed to be G-rich

d) Reverse primers were designed to be C-rich

e) Primers were chosen to give a product in the size range of 300-600bp

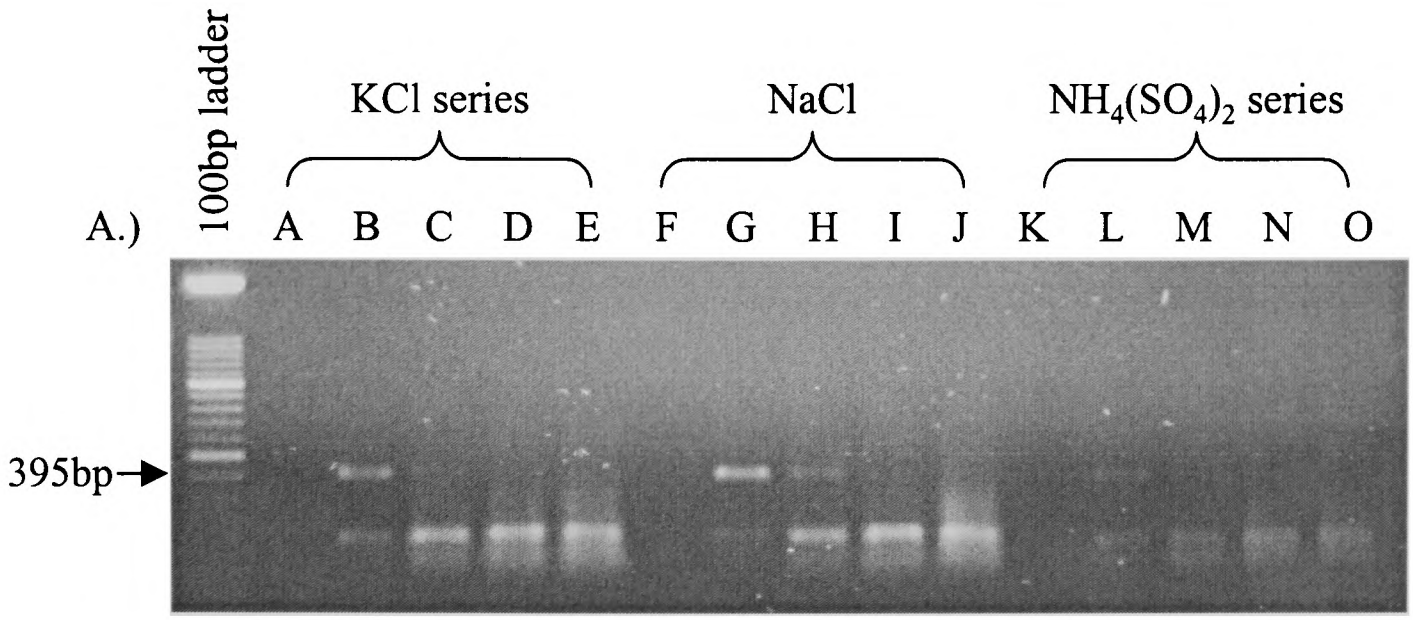
f) Primers were designed to be complimentary to bisulphite treated DNA (i.e. all non-CpG cytosines were treated as thymines).

g) Primers were designed to amplify the coding-strand of the DNA.

Using these criteria, the primers PDHA1BISF1 (5'-AGATGTATTGAACGAGGGTTGTA-3')and PDHA1BISR2 (5'-CCCAAATCTCCTTCAAATACC-3') were designed for the *PDHA1* promoter to amplify a 395bp product (-333 to +61bp) corresponding to the region shown in the CpG-Plot of Figure 3.3.

ii) Optimisation of the *PDHA1* bisulphite PCR reaction.

The *PDHA1* bisulphite PCR reaction was optimised using the 15 buffer series as previously described (see Materials 2.2 and Methods 2.44). The exception was that, due to the nature of the bisulphite treated template, the first annealing temperature used was 45°C irrespective of primer composition. Figure 3.8A shows the optimisation for the primers PDHA1BISF1 and



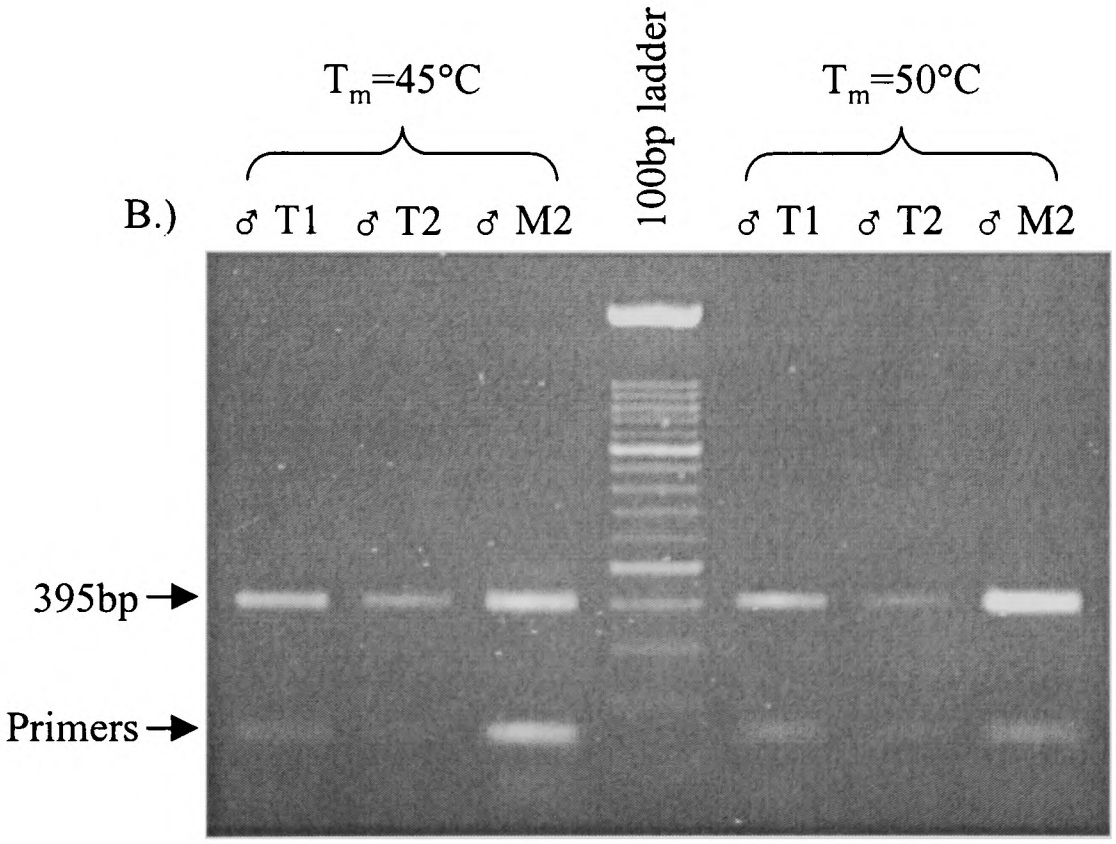


Figure 3.8: Optimisation of the Human PDHE1α bisulphite PCR.
DNA was isolated from HeLa cells and 5μg used in a bisulphite conversion reaction. The DNA was purified and 1μl ($1/20^{th}$) used in PCR reactions with the primers HPDHA1BISF1 (5'-AGATGTATTGAACGAGGGTTGTA-3') and HPDHA1BISR2 (5'-CCCAAATCTCCTTCAAATACC-3') in each of 15 PCR reactions using an optimisation buffer series (A-O). The buffers are made up with three different base salt compositions and 5 different magnesium concentrations per set. The optimsation buffers B and G generated the correct 395bp product (A) with buffer G giving a better yield. Further optimisation was performed by repeating the PCR with buffer G on three different bisulphite treated DNA samples (T1, T2 and M2) at two different annealing temperatures (B).

PDHA1BISR2. Buffers B and G produced a band of the correct size (395bp) with buffer G giving a higher product yield. The PCR was further optimised by investigating the effect of changing the annealing temperature. Figure 3.8B shows the annealing temperature optimisation with bisulphite treated DNA from three different sources T1, T2 (peripheral blood) and M2 (cultured fibroblast). The optimisation revealed little difference in specificity or yield over a 5°C annealing temperature range (two of which are shown). To reduce the possibility of PCR bias in the amplification process the lowest annealing temperature that generated a specific product was used. The optimised conditions for the *PDHA1* bisulphite PCR are shown in table 3.2 (below).

PDHA1 bisulphite PCR PDHA1BISF1 : PDHA1BISR2 Buffer G	96°C for 2'00'' 96°C for 0'30'' 45°C for 0'15'' 72°C for 0'30'' } 35-40 cycles

Table 3.2: Optimised conditions for the *PDHA1* bisulphite PCR reaction.

iii) Bisulphite sequence analysis of HeLa genomic DNA.

DNA prepared from HeLa tissue culture cells was bisulphite treated and used as template for the *PDHA1* bisulphite PCR. The PCR products were separated from primers by agarose gel electrophoresis and then excised from the gel and purified. The amount of purified product isolated was calculated by running an aliquot on an agarose gel against a known quantity of molecular weight standard. If the total amount was less than 100ng the purified product was concentrated by ethanol precipitation. To enable cloning of the *Taq* polymerase generated product it was "polished" by *PfU* polymerase and then blunt ligated into the *Eco RV* site of the vector pZErO2 (see Methods 2.39.2 and 2.39.3). Following transformation and culture on selective plates, individual bacterial clones were picked and screened by colony PCR using

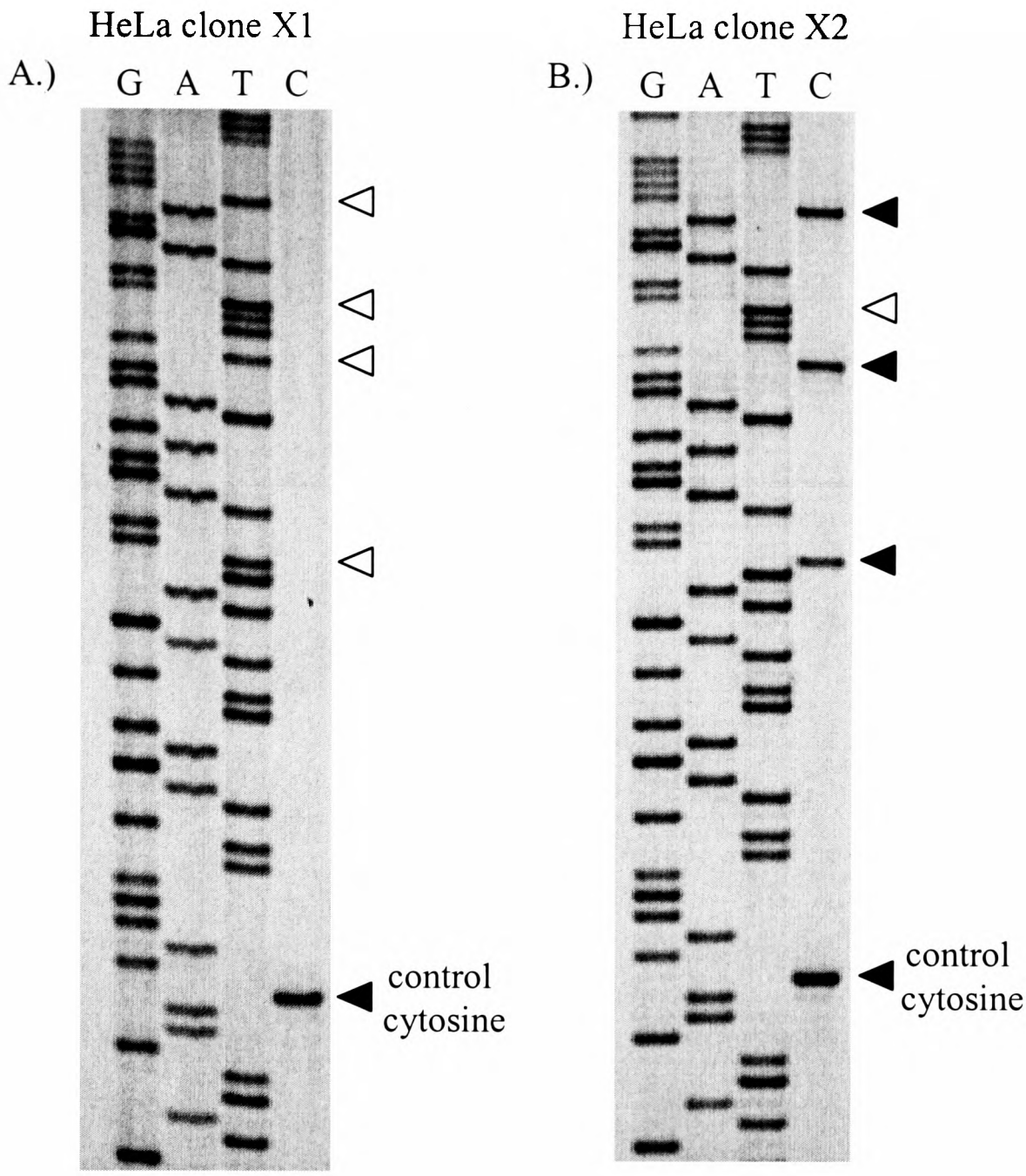
the MCS flanking SP6 and T7 primers. Plasmid DNA was prepared from clones carrying inserts of the correct size and these were sequenced using SP6 and T7 primers.

Sequencing of 10 independent clones revealed that there were two different classes of insert sequence. Figure 3.9 shows part of the sequence autoradiographs of two clones (X1 and X2) which are representative of each class. Figure 3.10 shows the alignment of the whole sequence of clone X1 against the native genomic sequence for the target region with CpG di-nucleotides highlighted. The alignment shows that none of the 34 CpG di-nucleotides are methylated in clones of this class. Figure 3.11 shows the same alignment for clone X2. In this class, 22 out of the 34 CpG di-nucleotides are methylated (64.7 %) which is 8.5x higher than the average CpG site methylation in the genome (Solage and Cedar, 1978) and (Woodcock et al., 1987). The methylation profiles from the two classes have been placed on a promoter map showing the positions of the CpG di-nucleotides, the transcription start site, the "core"-promoter region and a selection of transcription factor sites (Figure 3.12).

iv) Bisulphite sequence analysis of lymphocyte genomic DNA.

Genomic DNA from human peripheral blood was used to establish whether the bisulphite sequences obtained from HeLa cell genomic DNA were representative of those found in non-cultured cell types. The sequences obtained should represent those from the male active X-chromosome (X_{aM}), the female active X-chromosome (X_{aF}) and the female inactive X-chromosome (X_i).

DNA isolated from the peripheral blood of 6 males (5, 6, 9, 10, 12 and 15) was subjected to bisulphite analysis as in iii) (above). Figure 3.13A shows the result of the bisulphite PCR reaction for the first five of these, a 395bp PCR product along with primer/primer-dimer



C.)

GTACTGAACGAGGGCTGCAGAGCTGCAGCACCGGCAGGAGTAGGCGCTCGGTAGGACGGGGCC

Figure 3.9: Bisulphite sequences from treated HeLa cell DNA.
Bisulphite PCR products derived from HeLa cell genomic DNA were blunt-cloned into the EcoRV site of the vector pZErO2 and independent bacterial clones isolated. Plasmid DNA was prepared from the clones and sequenced using either SP6 and/or T7 primers flanking the vector MCS. The resulting bisulphite sequences fell into two categories typified by the clones X1 (A) and X2 (B). The position of a control cytosine in the PCR primer is indicated in both sequences. CpG di-nucleotide positions are indicated with triangles which are either closed or open depending on whether the site is methylated or unmethylated respectively. The genomic sequence of the region included in the autoradiograph images is provided with CpG di-nucleotides highlighted for reference (C).

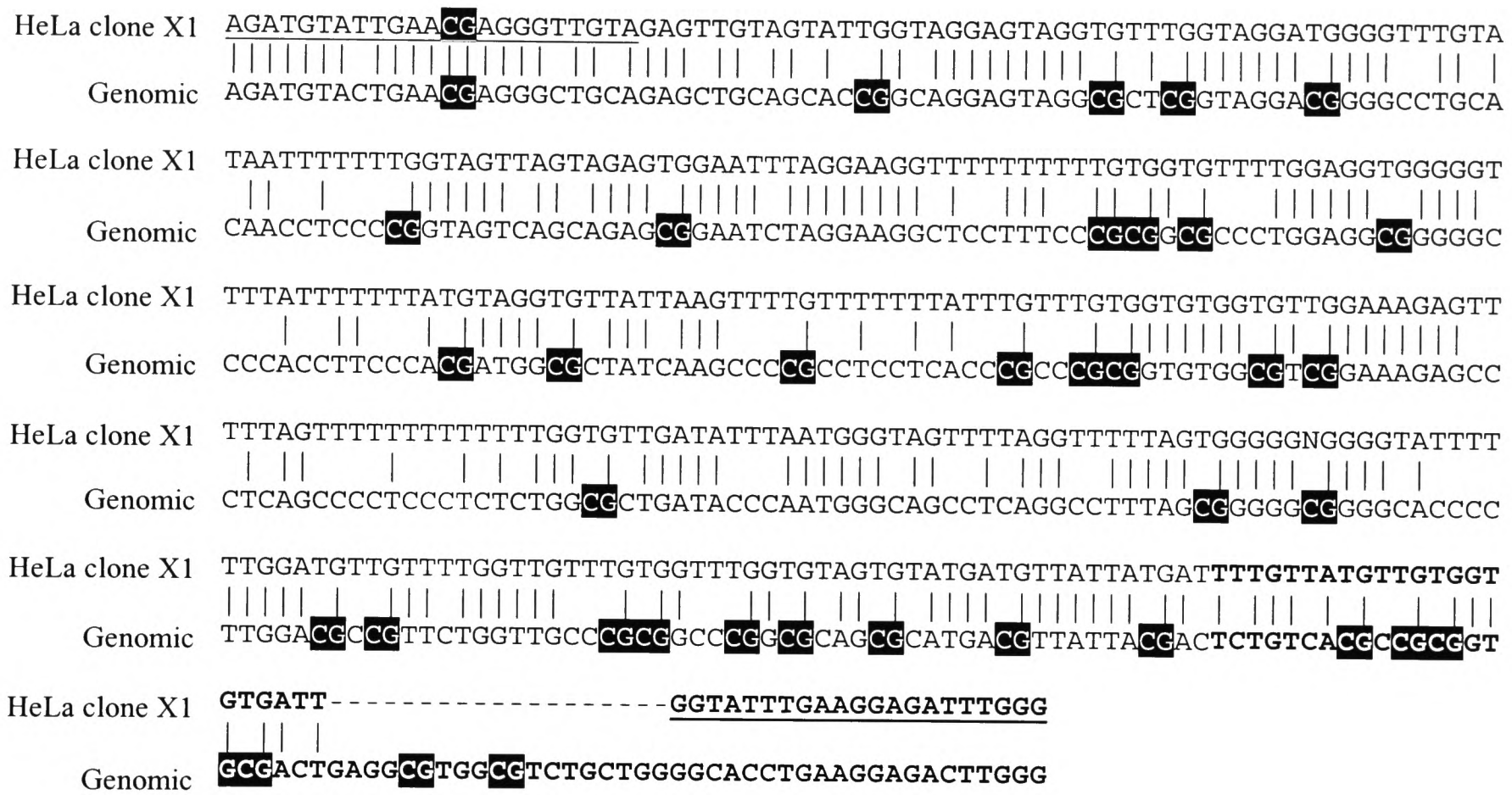


Figure 3.10: Bisulphite sequence of the HeLa clone X1.
The figure shows the bisulphite sequence of HeLa clone X1 which has been aligned against the genomic DNA sequence for the region. Vertical bars represent identical nucleotides between the two sequences. The primer sequences (underlined), the CpG di-nucleotides (highlighted) and exon 1 (bold) are also shown. Dashed lines represent a section of the insert which proved intransigent to manual sequencing and produced bands across all four lanes of the sequencing gel (typical of a CG compression).

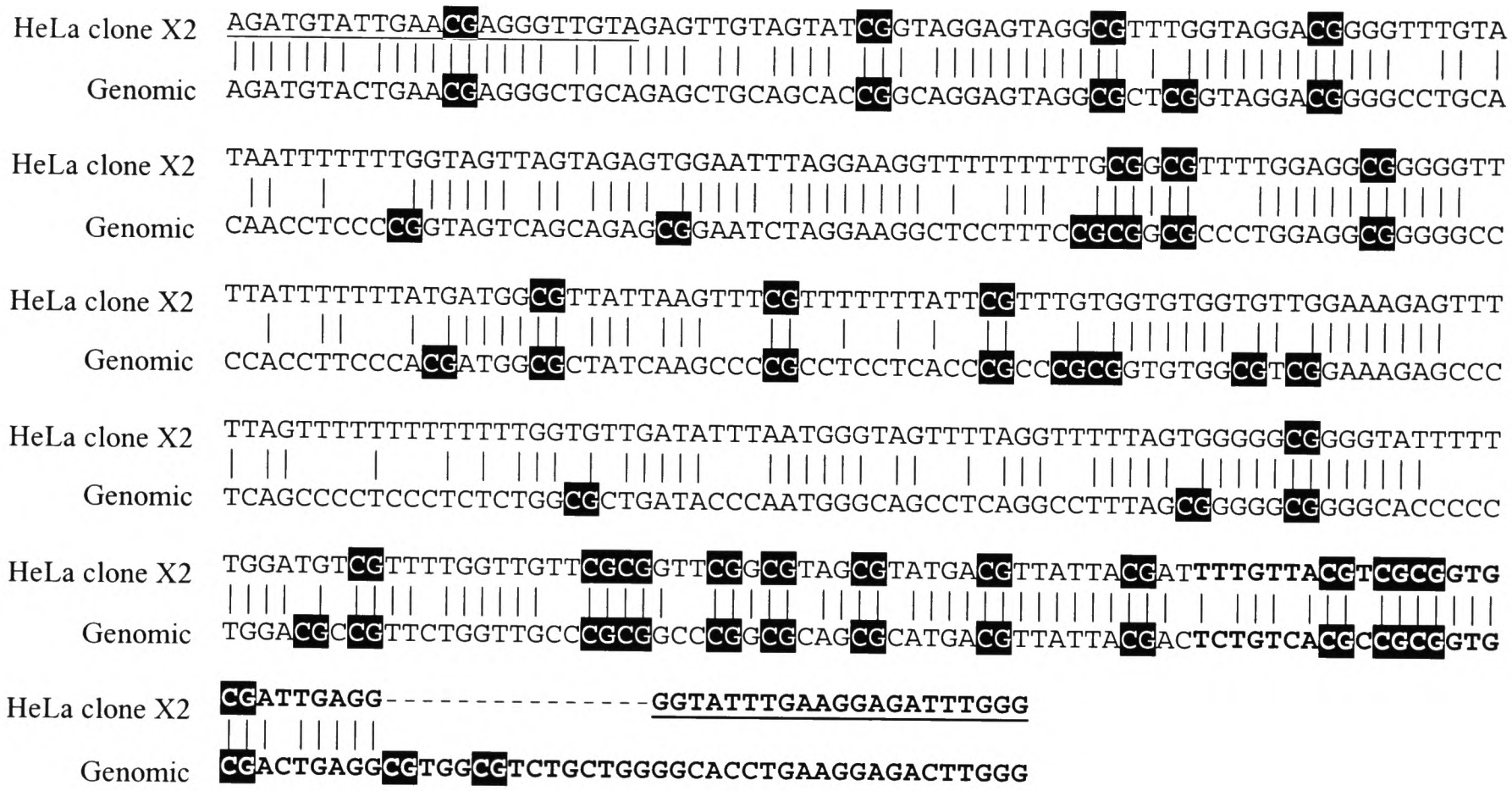


Figure 3.11: Bisulphite sequence of the HeLa clone X2.
The figure shows the bisulphite sequence of the HeLa clone X2 which has been aligned against the genomic DNA sequence for the region. Vertical bars represent identical nucleotides between the two sequences. The primer sequences (underlined), the CpG di-nucleotides (highlighted) and exon 1 (bold) are also shown. Dashed lines represent a section of the insert which proved intransigent to manual sequencing and produced bands across all four lanes of the sequencing gel (typical of a CG compression).

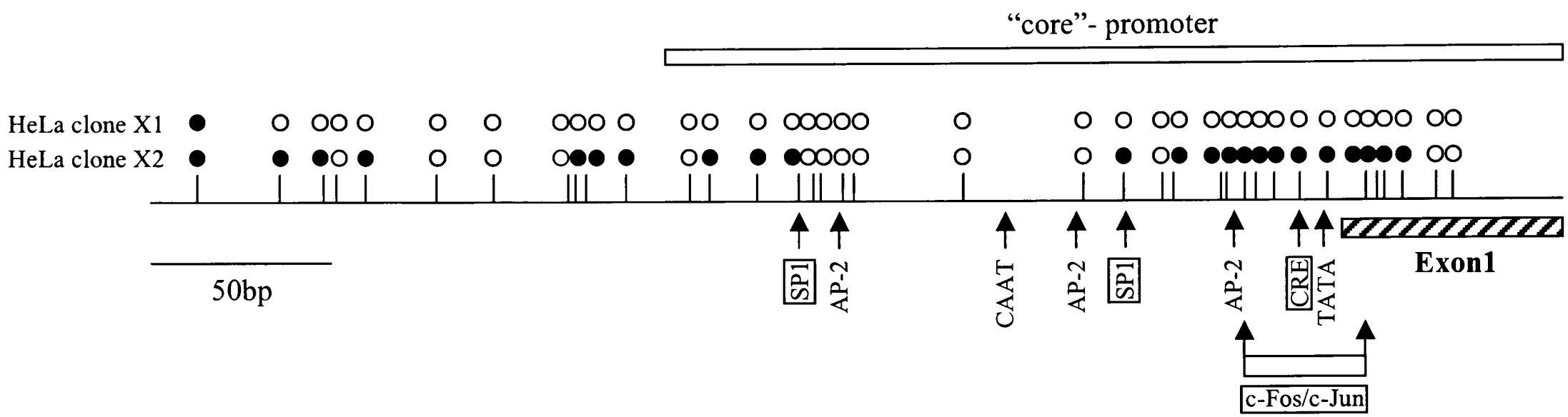


Figure 3.12: Schematic of the two classes of bisulphite sequence generated from HeLa cell DNA.
Unmethylated (○) and methylated (●) CpG di-nucleotide sites are shown as open and closed circles respectively. One class of sequence contains no methylated CpG sites wheras the other class has 22/34 (64.7 %) of CpG sites methylated in the target region of 395bp. The first reference CpG is shown as methylated, but is part of the PCR forward primer (HPDHA1BISF1) and so its methylation status cannot be determined using these primers. The position of the core promoter defined by (Chang et. al. 1993) is shown as are the positions of some of the transcription factor consensus binding sites. These sites are boxed if CpG site(s) in them are methylated.

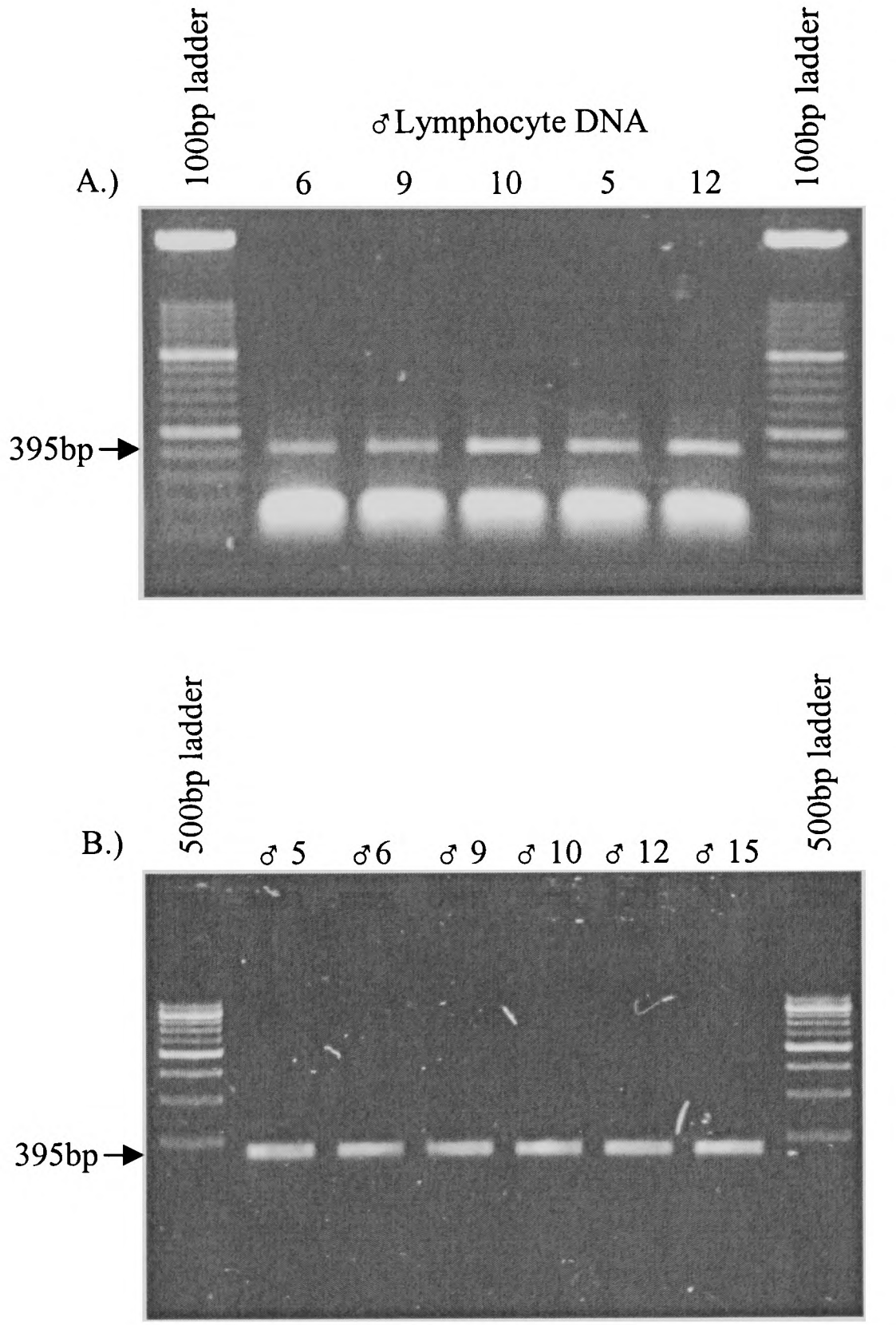


Figure 3.13: Bisulphite PCR on treated male lymphocte DNA samples. Peripheral blood samples were obtained from male donors and DNA isolated by standard methods. Aliquots of DNA (5μg) were subjected to bisulphite treatment and then purified. PCR reactions were performed on 5 male samples (5,6,9,10 and 12) using the primers HPDHA1BISF1 and HPDHA1BISR2 and 1μl ($1/20^{th}$) of the treated DNA (A). PCR products were separated from primers by electrophoresis in a 1% (w/v) agarose gel then purified by excising the bands from the gel and extracting the DNA from the gel slices. Aliquots of the puified PCR products were quantitated on a 1% (w/v) agarose gel (B) prior to "polishing" and cloning into the EcoRV site of the vector pZErO2.

band. These PCR products were purified and aliquots quantitated on an agarose gel (Figure 3.13B). The products were polished with *PfU* polymerase and cloned into the *Eco RV* site of the pZErO2 vector. Individual bacterial clones carrying plasmid with the correct sized insert were identified by colony PCR and then sequenced using SP6 and T7 primers.

DNA from the peripheral blood of 8 females (1, 3, 4, 7, 11, 13, 14 and 16) was subject to bisulphite sequence analysis as described above. Figure 3.14 shows the results of bisulphite PCR reactions with these samples as well as a positive control (bisulphite treated HeLa DNA), a negative control (no DNA) and a control to demonstrate that the bisulphite primers only worked on bisulphite treated DNA (untreated HeLa DNA).

A total of 10 independent clones were manually sequenced from 3 different males (10, 12 and 15) and no methylation was found to be present in any of these. Figure 3.15 shows sequence alignments for clones 10-1 and 15-5 against the native genomic sequence for this region. Fifteen independent clones were manually sequenced from 4 female samples (3, 4, 13 and 14) and no methylation was found to be present in any of these. Figure 3.15 shows sequence alignments for clones 4-2 and 4-4 against two male sequences (10-1 and 10-5) and the native genomic sequence for this region.

(v) Analysis of bisulphite clones by restriction digestion with *Tsc I.*

The absence of methylated sequences from female samples was unexpected, especially as 3 out of 5 samples analysed by msPCR produced bands in the *Hpa II* digested lanes (see Figure 3.6B). Having manually sequenced 15 independent female clones, an alternate and more efficient method was adopted to try to establish whether methylated clones were simply under-represented as opposed to absent. The enzyme *Tsc I* cuts at the site ACGT, so it will

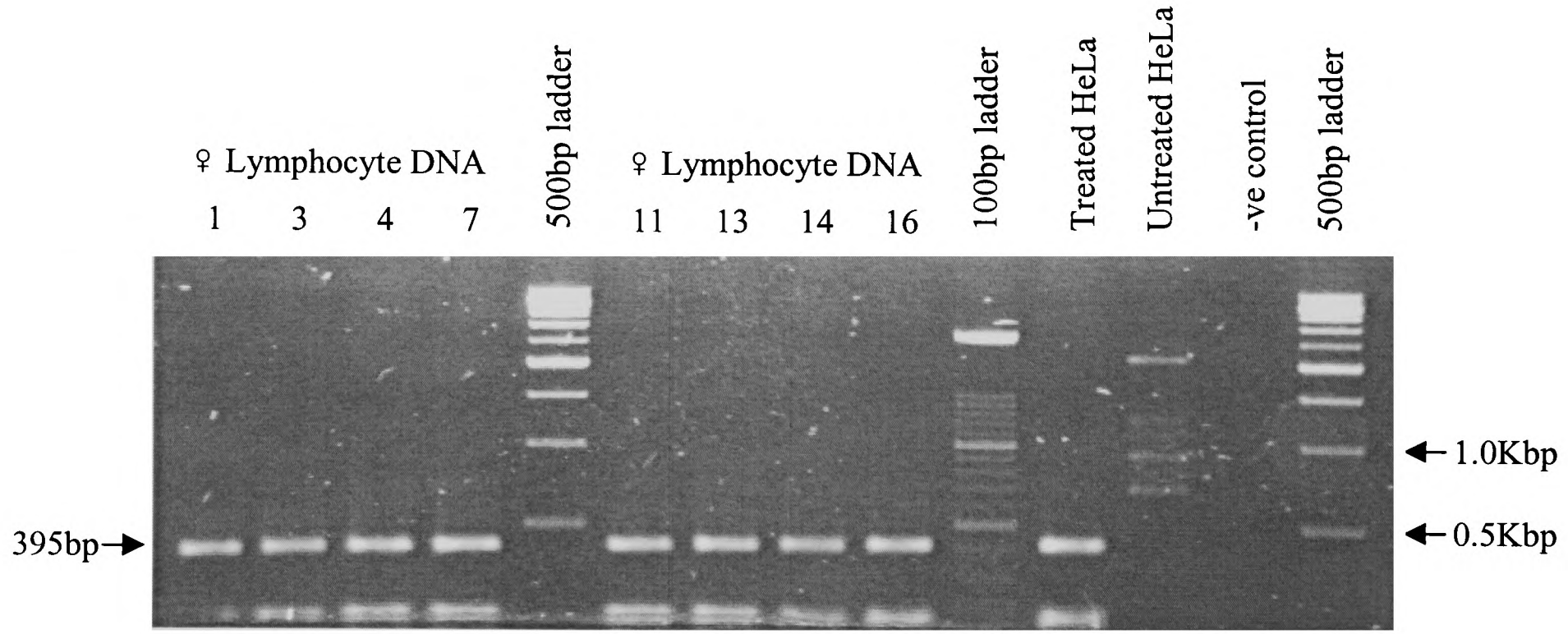


Figure 3.14: Bisulphite PCR on treated female lymphocte DNA samples.
Peripheral blood samples were obtained from female donors and DNA isolated by standard methods. Aliquots of DNA (5μg) were subjected to bisulphite treatment and then purified. PCR reactions were performed on 8 female samples (1,3,4,7,11,13,14 and 16) using the primers HPDHA1BISF1 and HPDHA1BISR2 and 1μl ($1/20^{th}$) of the treated DNA. Bisulphite treated and untreated aliquots of HeLa DNA (5μg) were used as controls to verify that the primers only prime treated and not untreated DNA templates. PCR products were separated from primers by electrophoresis in a 1% (w/v) agarose gel then purified by excising the bands from the gel and extracting the DNA from the gel slices. Aliquots of the purified PCR products were quantitated on a 1% (w/v) agarose gel prior to "polishing" and cloning into the EcoRV site of the vector pZErO2.

```
Genomic  AGATGTACTGAACGAGGGCTGCAGAGCTGCAGCACCGGCAGGAGTAGGCGCTCGGTAGGACGGGGCCTGCA
♀4-2     AGATGTATTGAACGAGGGTTGTAGAGTTGTAGTATTGGTAGGAGTAGGTGTTTGGTAGGATGGGGTTTGTA
♀4-4     AGATGTATTGAACGAGGGTTGTAGAGTTGTAGTATTGGTAGGAGTAGGTGTTTGGTAGGATGGGGTTTGTA
♂10_1    AGATGTATTGAACGAGGGTTGTAGAGTTGTAGTATTGGTAGGAGTAGGTGTTTGGTAGGATGGGGTTTGTA
♂15_5    AGATGTATTGAACGAGGGTTGTAGAGTTGTAGTATTGGTAGGAGTAGGTGTTTGGTAGGATGGGGTTTGTA

Genomic  CAACCTCCCCGGTAGTCAGCAGAGCGGAATCTAGGAAGGCTCCTTTCCTGCGGCGCCCTGGAGGCGGGGGC
♀4-2     TAATTTTTTTGGTAGTTAGTAGAGTGGAATTTAGGAAGGTTTTTTTTTTGTGGTGTTTTGGAGGTGGGGGT
♀4-4     TAATTTTTTTGGTAGTTAGTAGAGTGGAATTTAGGAAGGTTTTTTTTT-GTGGTGTTTTGGAGGTGGGGGT
♂10_1    TAATTTTTTTGGTAGTTAGTAGAGTGGAATTTAGGAAGGTTTTTTTTT-GTGGTGTTTTGGAGGTGGGGGT
♂15_5    TAATTTTTTTGGTAGTTAGTAGAGTGGAATTTAGGAAGGTTTTTTTTTTGTGGTGTTTTGGAGGTGGGGGT

Genomic  CCCACCTTCCCACGATGGCGCTATCAAGCCCCGCCTCCTCACCCGCCCGCGGTGTGGCGTCGGAAAGAGCC
♀4-2     TTTATTTTTTTATGTAGGTGTTATTAAGTTTTGTTTTTTTATTTGTTTGTGGTGTGGTGTTGGAAAGAGTT
♀4-4     TTTATTTTTTTATGTAGGTGTTATTAAGTTTTGTTTTTTTATTTGTTTGTGGTGTGGTGTTGGAAAGAGTT
♂10_1    TTTATTTTTTTATGTAGGTGTTATTAAGTTTTGTTTTTTTATTTGTTTGTGGTGTGGTGTTGGAAAGAGTT
♂15_5    TTTATTTTTTTATGTAGGTGTTATTAAGTTTTGTTTTTTTATTTGTTTGTGGTGTGGTGTTGGAAAGAGTT

Genomic  CTCAGCCCCTCCCTCTCTGGCGCTGATACCCAATGGGCAGCCTCAGGCCTTTAGCGGGGGCGGGGCACCCC
♀4-2     TTTAGTTTTTTTTTTTTTTGGTGTTGATATTTAATGGGTAGTTTTAGGTTTTTAGTGGGGGTGGGGTATTTT
♀4-4     TTTAGTTTTTTTTTTTTTTGGTGTTGATATTTAATGGGTAGTTTTAGGTTTTTAGTGGGGGTGGGGTATTTT
♂10_1    TTTAGTTTTTTTTTTTTTTGGTGTTGATATTTAATGGGTAGTTTTAGGTTTTTAGTGGGGGTGGGGTATTTT
♂15_5    TTTAGTTTTTTTTTTTTTTGGTGTTGATATTTAATGGGTAGTTTTAGGTTTTTAGTGGGGGTGGGGTATTTT

Genomic  TTGGACGCCGTTCTGGTTGCCCGCCGGCCCGGCGCAGCGCATGACGTTATTACGACTCTGTCACGCCGGGTG
♀4-2     TTGGATGTTGTTTTGGTTGTTTGTGGTTTGGTGTAGTGTATGATGTTATTATGATTTTGTTATGTTGGGTG
♀4-4     TTGGATGTTGTTTTGGTTGTTTGTGGTTTGGTGTAGTGTATGATGTTATTATGATTTTGTTATGTTGGGTG
♂10_1    TTGGATGTTGTTTTGGTTGTTTGTGGTTTGGTGTAGTGTATGATGTTATTATGATTTTGTTATGTTGGGTG
♂15_5    TTGGATGTTGTTTTGGTTGTTTGTGGTTTGGTGTAGTGTATGATGTTATTATGATTTTGTTATGTTGGGTG
```

Figure 3.15: Bisulphite sequences from male and female lymphocyte DNA.
Bisulphite PCR products derived from male and female peripheral blood DNA were blunt-cloned into the EcoRV site of the vector pZErO2 and independent bacterial clones isolated. Plasmid DNA was prepared from the clones and sequenced using either SP6 and/or T7 primers flanking the vector MCS. Representative sequences from two males (10 and 15) and one female (4) are shown aligned with a reference genomic sequence. The CpG di-nucleotides are highlighted including the reference CpG site in the HPDHA1BISF1 primer. A single nucleotide polymorphism resulting in a run of either 9 or 10 consecutive Ts is indicated in bold.

only cut bisulphite inserts if the original genomic sequences from which they are derived were methylated at the central cytosine. There are four potential *Tsc I* sites in the 395bp *PDHA1* bisulphite PCR product (at positions 151, 284, 322 and 341). If one or more of these sites is methylated the insert size would be cut down by at least 54bp, an easily observable difference on a 1% (w/v) agarose gel (395bp to 341bp). Thirty independent bisulphite clones from female lymphocyte DNA were tested to see if they were digested by *Tsc I*. The clone HeLa X2 was used as a positive control for digestion as it has two *Tsc I* sites (at 322 and 341bp). The HeLa clone digested with *Tsc I* as expected, but none of the 30 other clones digested (data not shown) suggesting that the original sequences from which they were derived were unmethylated at all four *Tsc I* sites. The fact that no indication of methylated sites was found in sequence and enzyme analysis from a total of 45 randomly selected clones would mean that if any were present within the bisulphite amplicon then they would represent less than 2% (1 out of 46) of all clones. The considerable cost and time involved in manually sequencing enough clones to find methylated sequences was not considered to be worthwhile especially as no evidence for methylation had been found.

Together these data suggest that neither the X_{aM}, X_{aP} or X_i are methylated in this region in human peripheral lymphocytes.

3.2.4 Expression analysis.

i) RT-PCR experiments.

It has been reported that *PDHA1* is a widely expressed gene (Brown et al., 1990). In order to demonstrate this, a variety of RT-PCR reactions were undertaken using RNA extracted from murine and human tissues as well as from HeLa cells.

Total RNA was extracted by an adaptation of the method of (Chomczynski and Sacchi, 1987) from samples of brain, heart, kidney, liver, muscle, spleen and eye dissected from a female laboratory mouse. The RNAs were used to synthesise first strand cDNA and these were then used as substrates in PCR reactions with the primers MPDHCDNAF1 (5'-AAGAACTTCTATGGAGGCAACG-3') and MPDHA1R1 (5-AGAGTTCCTTACGCTGGATGAG-3'). Total RNA was also isolated from tissue culture HeLa cells and from male and female retinal samples which were available as examples of uncultured human tissue. PCR was performed on first strand cDNA made from these RNAs with the primers PDH-C (5'-GATTGCTCTAGCCTGTAAG-3') and PDH-D (5'-TGCATGCAATTACTACCTGC-3'). The conditions used in both sets of PCR reactions are given in Table 3.3 (below).

Mouse *Pdha1* RT-PCR (756bp product) MPDHCDNAF1 : MPDHA1R1 Qiagen buffer	96°C for 2'00'' 96°C for 0'30'' 60°C for 0'30'' 72°C for 0'30'' } for 35 cycles
Human *PDHA1* RT-PCR (597bp product) PDH-C : PDH-D Qiagen buffer	96°C for 2'00'' 96°C for 0'30'' 54°C for 0'30'' 72°C for 0'30'' } for 35 cycles

Table 3.3 Conditions for RT-PCR reactions of the human and murine *PDHA1/Pdha1* genes.

Figure 3.16 shows the results for the RT-PCR reactions on all samples. All of the mouse tissues (Figure 3.16A), the human retinal tissues (Figure 3.16B) and the HeLa cells (Figure 3.16C) express the *PDHA1/Pdha1* gene. It is not possible to quantitate the products directly although equal amounts of total RNA were used in all first strand cDNA synthesis reactions and the same amount of cDNA used in each PCR reaction. In agreement with previous

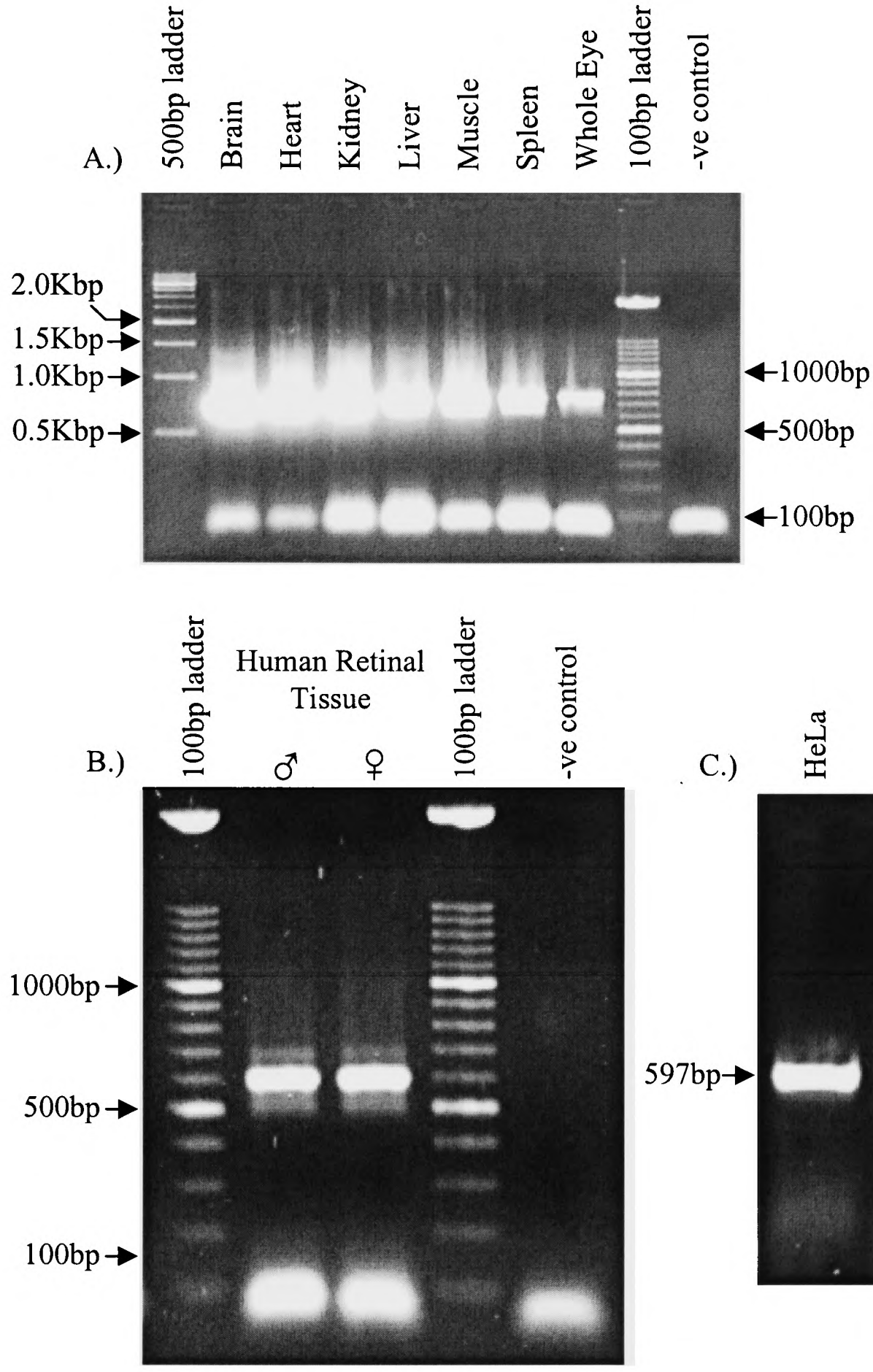


Figure 3.16: Expression of Human PDHE1α in various tissues types and HeLa cells. The expression of the PDHE1α gene was analysed by RT-PCR. Total RNA was isolated from tissues and cell lines and first strand cDNA prepared using a poly d(T) primer. Expression in a wide range of murine tissues was analysed by performing RT-PCR on prepared cDNA with the primers MPDHCDNAF1 (5'-AAGAACTTCTATGGAGGCAACG-3') and MPDHAR1 (5'-AGAGTTCCTTACGCTGGATGAG-3') (A). Expression of the Human PDHE1α gene was by RT-PCR using the primers PDH-C (5'-GATTGCTCTAGCCTGTAAG-3') and PDH-D (5'-TGCATGCAATTACTACCTGC-3') on male and female retinal tissue (B) and HeLa tissue culture cells (C).

reports, these results indicate that *PDHA1/Pdha1* is a widely expressed or "housekeeping" gene.

ii) Generation of EGFP-1 expression constructs.

A section of the *PDHA1* promoter (-414 to +142bp) complementing the regions previously studied by msPCR (-414 to +176bp) and bisulphite sequence analysis (-333 to +61bp) was amplified from male lymphocyte DNA using the primers PDHA1F1 and PDHA1EX1R (5'-GGCGAGCATCTTCCTCAT-3'). The conditions used were the same as for the "upstream" msPCR *PDHA1* reaction in section 3.2.2 (above). The 556bp PCR product was cloned into the *Eco RV* site of the vector pZErO2 and then shuttled into the expression vector pEGFP-1 as described in figure 3.17A. Bacterial clones were picked and screened by colony PCR using the primers PDHA1F1 and PDHA1EX1R to identify those carrying inserts of the correct size. Plasmid DNA was prepared from these clones and then digested with *Bam HI* and *Xho I* to confirm insert size and with *Pst I* to determine the orientation of the promoter relative to the EGFP cassette. Figure 3.17B shows the digests of a forward (PDHPF-EGFP1) and a reverse (PDHPR-EGFP1) clone which were both sequenced to confirm the integrity and orientation of their inserts (data not shown).

iii) Functional analysis of the *PDHA1* promoter-reporter gene constructs.

The –731 to +65bp region of the 5' proximal sequence of the *PDHA1* gene has previously been shown to be capable of driving the CAT (chloramphenicol acetyltransferase) gene in transient transfection experiments in HepG2 cells (Chang et al., 1993). It was important to establish that the –414 to +142bp region cloned into pEGFP-1 in the forward orientation was also capable of driving EGFP expression.

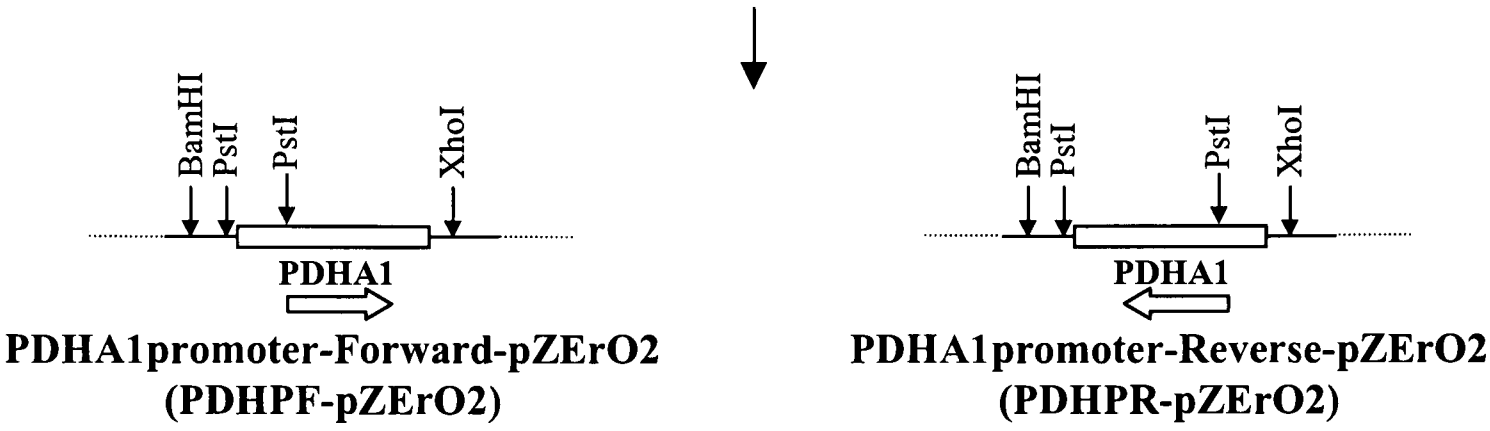


Figure 3.17A: Making the Human PDHE1α promoter EGFP-1 expression constructs.

The genomic DNA primers PDHA1F1 (5'-GCAGACCATTTACAGAAACCAG-3') and PDHA1EX1R (5'-GGCGAGCATCTTCCTCAT-3') were used to amplify a 556bp genomic DNA fragment of the human gene promoter (-414 to +142). The PCR product was polished with *PfU* polymerase and cloned into the EcoRV site of the vector pZErO2. Insert containing clones were screened to determine the orientation of the insert by PstI digestion. Inserts were prepared from a clone of each orientation by double digestion with BamHI and XhoI and cloned into pEGFP-1.

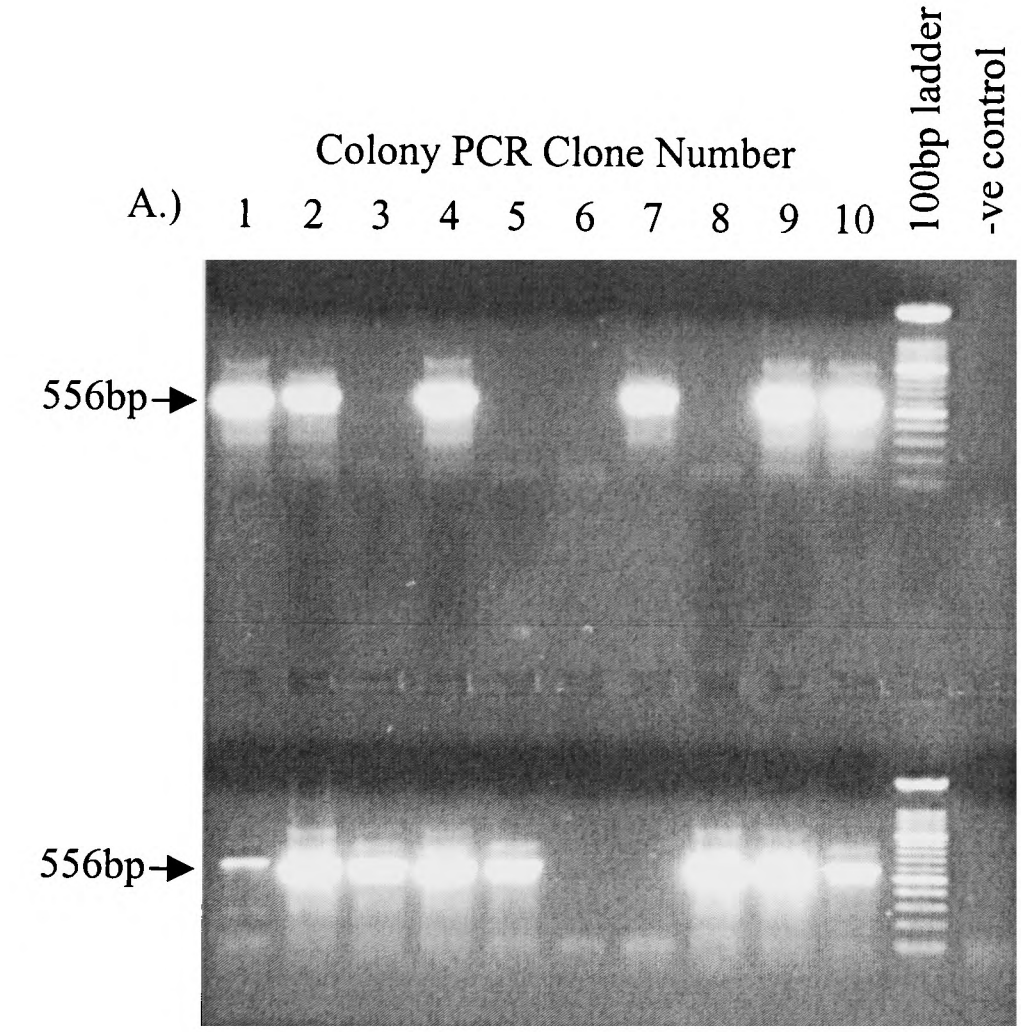


Figure 3.17B: Making the Human PDHE1α promoter EGFP-1 expression constructs.

Insert carrying clones were identified by colony screening PCR using the primers PDHA1F1 and PDHA1EX1R (A). A clone of each orientation was prepared and subjected to restriction digestion with BamHI, BamHI/XhoI and PstI to confirm the insert size and orientation of the clones in pEGFP-1 (B). The insert size is 620bp in both clones shown and the orientation digest with PstI gives a band of 110bp and 458bp for the reverse and forward orientations respectively. The clones were sequenced to confirm that they carried the correct inserts in the correct orientation relative to the EGFP protein cDNA.

HeLa cells were selected as the cell line into which the various construct preparations were to be transfected as they had previously been shown to express the *PDHA1* gene (see Figure 3.16C). Aliquots of $1x10^6$ cells were electroporated with 2μg of the various constructs at 250V and 1500μF, then recovered in wells of an 8-well multi-plate containing sterile cover-slips and 3ml of culture medium (see Methods 2.50). Cells were cultured for 48 hours at 37°C in an atmosphere of 5% (v/v) CO_2 then fixed and mounted in Moviol (containing Propidium iodide at 2μg/ml) on a glass microscope slide (see Methods 2.53 and 2.54).

Slides were visualised with a Zeiss photo-microscope (*Zeiss*) under UV-illumination through a dual pass filter that allowed the simultaneous viewing of red and green signals. Images were captured on Fuji 35mm slide-film at fixed exposure and magnification using a Zeiss automatic mounted camera. Positives were developed from the film which were scanned into a computer using a Minolta Quickscan35 scanner and stored in TIF file format.

In order to assess the effect of DNA methylation on the ability of the promoter to drive EGFP expression, both forward and reverse constructs were artificially methylated prior to transfection by the enzyme *Sss I* methylase (see Methods 2.48). This enzyme catalyses the methylation of DNA at every CpG di-nucleotide under optimal conditions (Hoffman, 1986). To ensure that the competency of the promoter-constructs to drive EGFP expression was not affected by factors other than the actual methylation itself, aliquots of both vectors were subjected to "dummy" methylation reactions. In these reactions the vectors were taken through all of the pre-transfection steps that the methylated constructs had been through except that in the methylation reaction the substrate was left out. Aliquots of both methylated and "dummy" methylated forward and reverse constructs were digested with the enzyme *Hpa II* to confirm their methylation status (Figure 3.18).

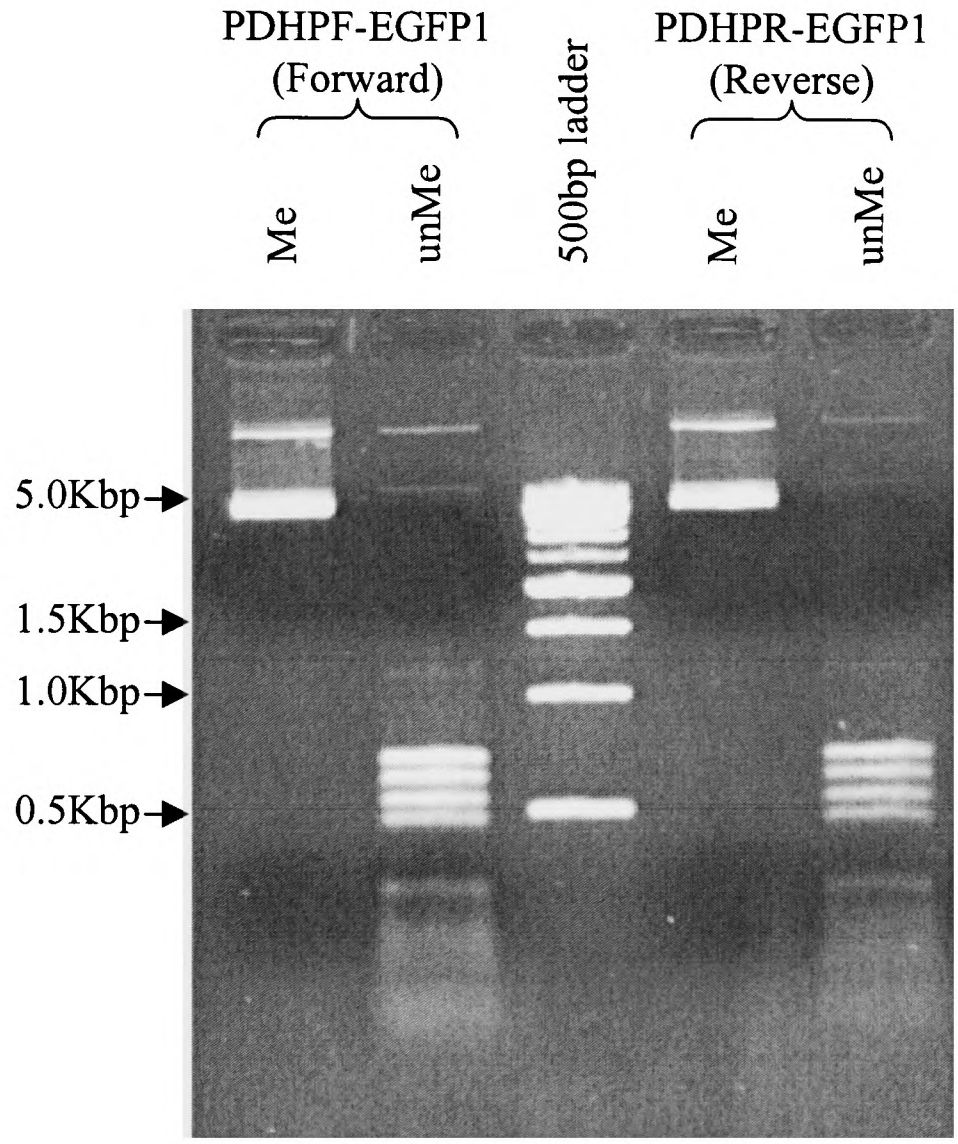


Figure 3.18 HpaII digests of methylated and "dummy" methylated Human PDHE1α promoter EGFP-1 constructs.
In order to ascertain the effect of DNA methylation on the ability of the PDHE1α promoter region (-414 to +142) to drive EGFP reporter gene expression, constructs were artificially methylated prior to transfection. The enzyme SSs I methylase was used to totally methylate aliquots of forward and reverse constructs prior to their transfection into tissue culture cells. As a control, aliquots of forward and reverse constructs were treated in a "dummy" methylation reaction whereby they were subject to the same conditions as during the methylation reaction, but in the absence of one key component (either the substrate or the enzyme). Aliquots of methylated (Me) and "dummy" methylated (unMe) constructs were digested with the methylation sensitive enzyme HpaII to ensure that that the methylation reaction was complete prior to electroporation.

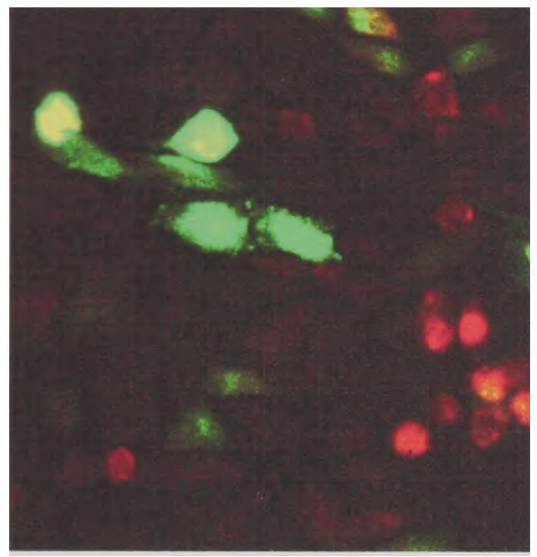
Transfection experiments with the PDHPF-EGFP1 (forward) construct show that the untreated (Figure 3.19A) and "dummy" methylated (Figure 3.19B) constructs direct expression of the EGFP protein resulting in cells with green coloured cytoplasm. The methylated construct does not direct the expression of the EGFP protein in any cells (Figure 3.19C). These results suggest that the –414 to +142bp section of the *PDHA1* promoter is capable of driving reporter gene expression *in-cis* and that methylation of this construct ablates reporter gene expression.

Transfection experiments were also carried out with the PDHPR-EGFP (reverse) construct. The untreated (Figure 3.20A) and "dummy" methylated (Figure 3.20B) constructs directed expression of the EGFP protein, but the methylated construct (Figure 3.20C) did not. These data suggest that some elements in the –414 to +142bp promoter region are capable of directing transcription even though the promoter is in the reverse orientation and that methylation of the construct ablates this function.

3.2.5 Isolation of mouse PAC clones of the murine *Pdha1* gene.

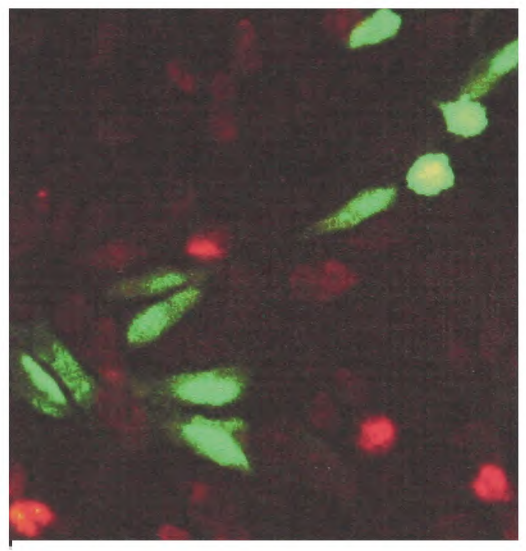
The difficulty of obtaining fresh human tissue for the analysis of promoter methylation makes mouse promoters an attractive alternative. To do this it is of course essential to know the sequence of the mouse promoter. Unfortunately study of the world genome databases reveals that very few human and even fewer murine genes have had their promoters isolated and sequenced. In order to generate a resource for the future study of the murine *Pdha1* gene and its promoter, a mouse PAC library was screened to isolate clones containing the murine *Pdha1* gene.

A.)



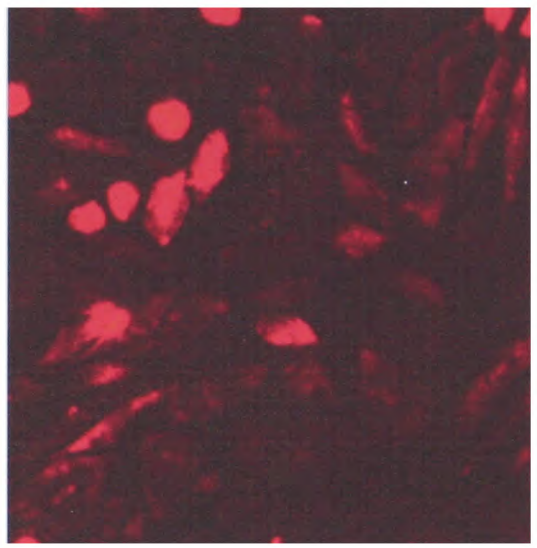
x200

B.)



x200

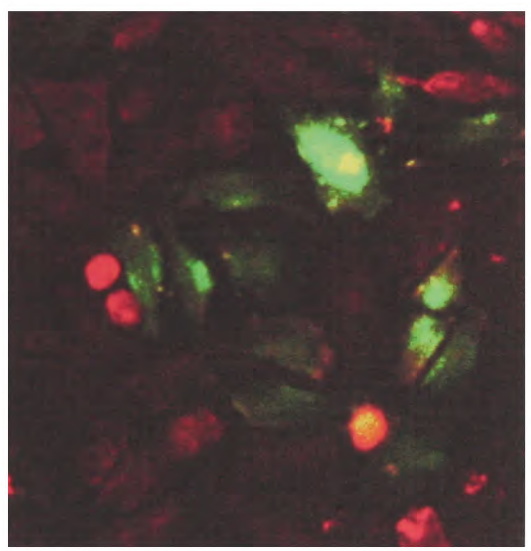
C.)



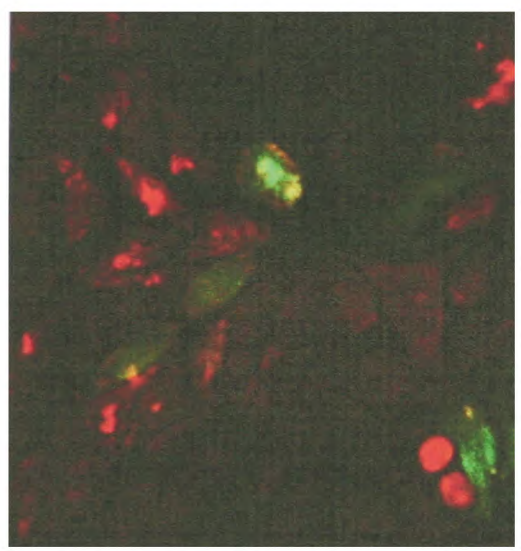
x200

Figure 3.19 Human PDHPF-EGFP1 expression construct analysis in HeLa cells.
The PDHPF-EGFP1 expression construct was used to assess the potential of the cloned region of the human promoter to drive reporter gene expression *in-cis*. In addition the artificial methylation of the construct was used to determine whether methylation was capable of down-regulating promoter driven reporter gene expression. Aliquots of $1x10^6$ HeLa cells were electroporated with 2.0μg of vector and then plated into multi-well plates containing sterile cover-slips and medium. The cells were cultured for 48 hours at 37°C in 5% (v/v) CO_2 then fixed in 4% (w/v) paraformaldehyde and mounted in Moviol aqueous mountant containing Propidium iodide at 2μg/ml. Slides were viewed under UV-illumination through a dual pass filter to allow the simultaneous viewing of red and green signals. Images were captured on Fuji 35mm slide-film and developed positives were scanned with a Minolta Quickscan35 scanner and stored in TIF file format. The HeLa cells were transfected with either untreated vector (A), vector that had been through a dummy methylation reaction (B) or vector that had been totally methylated at all CpG sites with SSs I methylase (C).

A.)



B.)



C.)

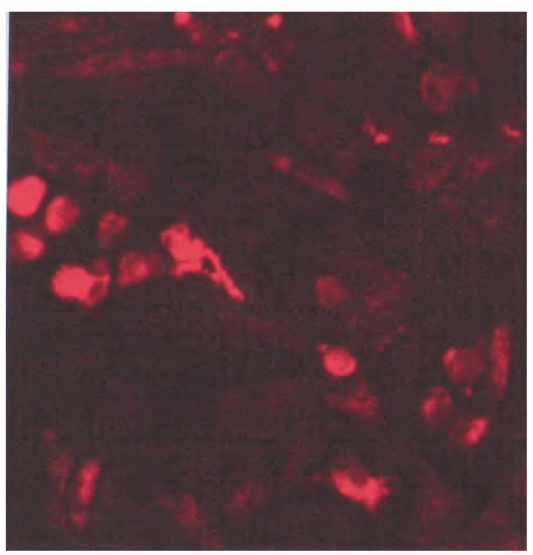


Figure 3.20 Human PDHPR-EGFP1 expression construct analysis in HeLa cells.
The PDHPR-EGFP1 expression construct was used to assess the potential of the reverse orientation cloned region of the human promoter to drive reporter gene expression *in-cis*. In addition the artificial methylation of the construct was used to determine whether methylation was capable of down-regulating reporter gene expression. Aliquots of $1x10^6$ HeLa cells were electroporated with 2.0μg of vector and then plated into multi-well plates containing sterile cover-slips and medium. The cells were cultured for 48 hours at 37°C in 5% (v/v) CO_2 then fixed in 4% (w/v) paraformaldehyde and mounted in Moviol aqueous mountant containing Propidium iodide at 2μg/ml. Slides were viewed under UV-illumination through a dual pass filter to allow the simultaneous viewing of red and green signals. Images were captured on Fuji 35mm slide-film and developed positives were scanned with a Minolta Quickscan35 scanner and stored in TIF file format. The HeLa cells were transfected with either untreated vector (A), vector that had been through a dummy methylation reaction (B) or vector that had been totally methylated at all CpG sites with *SSi* methylase (C).

The mouse genomic PAC library RPCI21 (see Materials 2.11) was screened with the probe PDH1c (Dahl et al., 1990) (Figure 3.21) which contains the entire coding sequence of the human *PDHA1* gene. The library is made up of 7 gridded filters with each clone double spotted in one of 8 different patterns so as to allow accurate identification of positive clones. The library screening generated 19 positives, 2 of which are shown in figure 3.22A. Six of the strongest positives were ordered from the HGMP-RC: clones 496-B16, 621-J6, 543-E8, 391-I14, 440-L12 and 402-I11.

The human PDH-E1α cDNA hybridises to two genomic loci, one on the X-chromosome (Brown et al., 1989) and one on chromosome 4 (Dahl et al., 1990). Similarly the murine cDNA hybridises to two loci, one on the X-chromosome and the other on chromosome 19 (Brown et al., 1990). The human and murine autosomal copies are thought to be intron-less and to behave like functional processed genes. Screening of the mouse PAC library was likely to have isolated clones from both loci, so it was important to distinguish the X-linked from the autosomal clones. This could be achieved by designing PCR primers to intronic sequences flanking an exon of the *Pdha1* gene.

This author has previously isolated exon 8 containing mouse genomic phage clones (unpublished data), so exon 8 flanking intronic sequences were available for primer design. The primers PDHA8F (5'-ATGCATAGACGATACTCGAC-3') and PDHA8R (5'-TGACAGTACAGTAGACATTC-3') were used to screen the 6 PAC clones by PCR using the same conditions as for colony PCR (see Methods 2.40). The PCR produced a product band of 1200bp if the clones were X-linked and no product if they were not. This approach showed clones 621-J6, 543-E8 and 402I-11 to be X-linked (Figure 3.22B). Restriction digestion of DNA from these three clones with *Bam HI*, *Eco RI* and *Hind III* gave different digest patterns

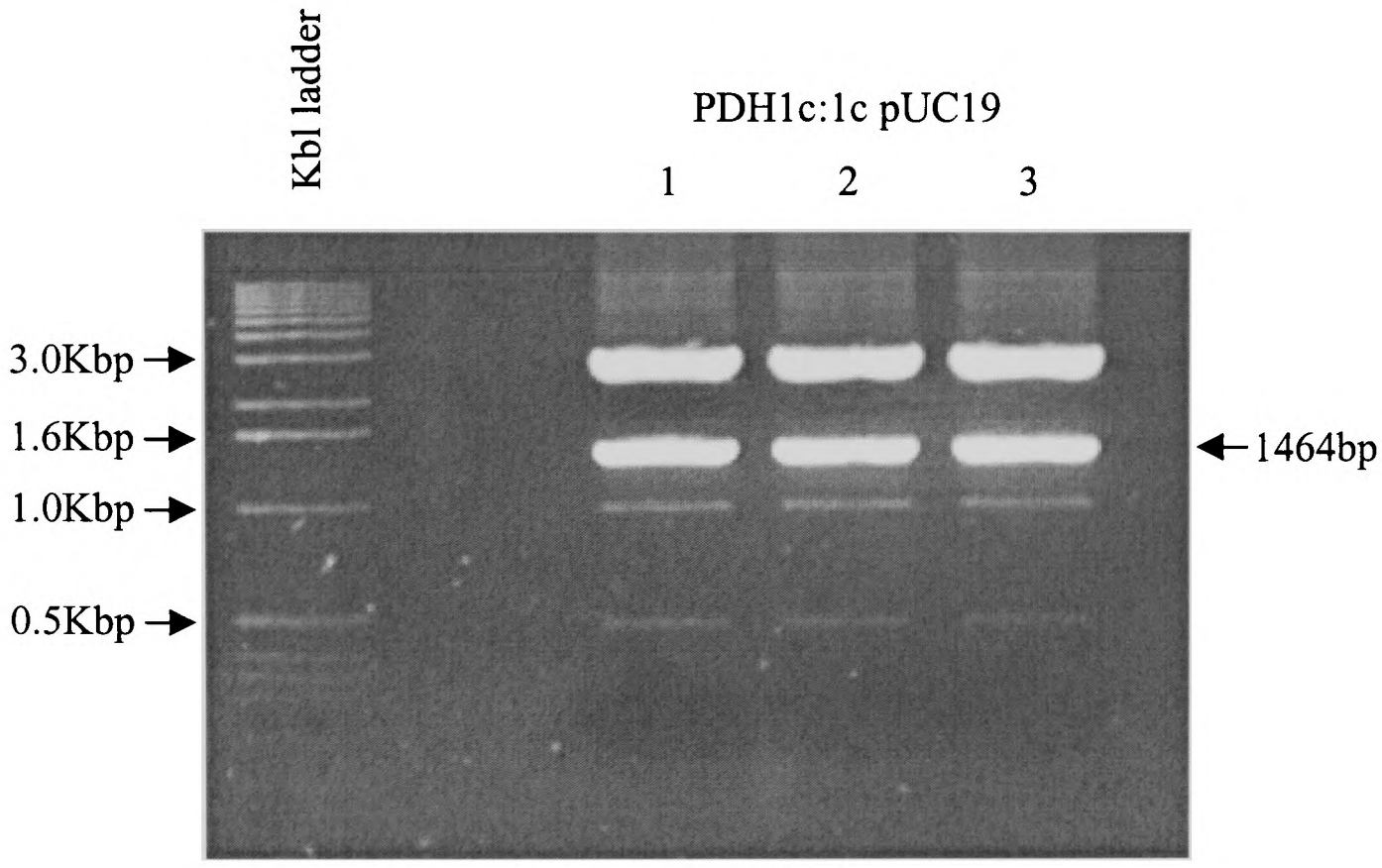
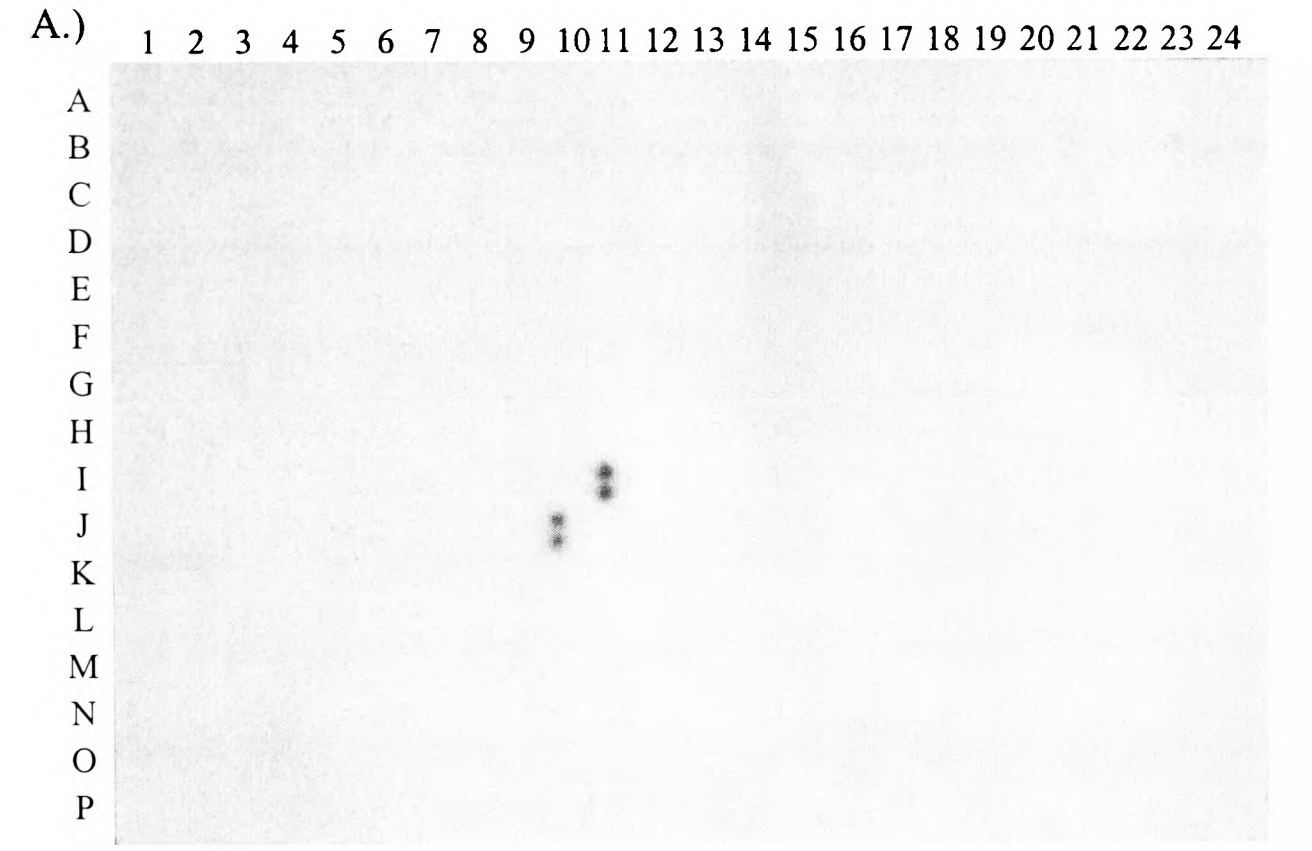


Figure 3.21 Isolation of the PDHA1:1c probe.
The probe PDHA1:1c was used to screen the mouse RPCI21 PAC library. PDHA1:1c is a 1464bp cDNA probe from the human X-linked PDHA1 locus that contains 78bp of the 5' untranslated region and the complete coding and 3' untranslated regions of the gene (Dahl et. al. 1990). The cDNA fragment is cloned into the Eco RI site of the vector pUC19. Prior to labelling, the vector was cut with Eco RI and the 1464bp probe was purified from an agarose gel.



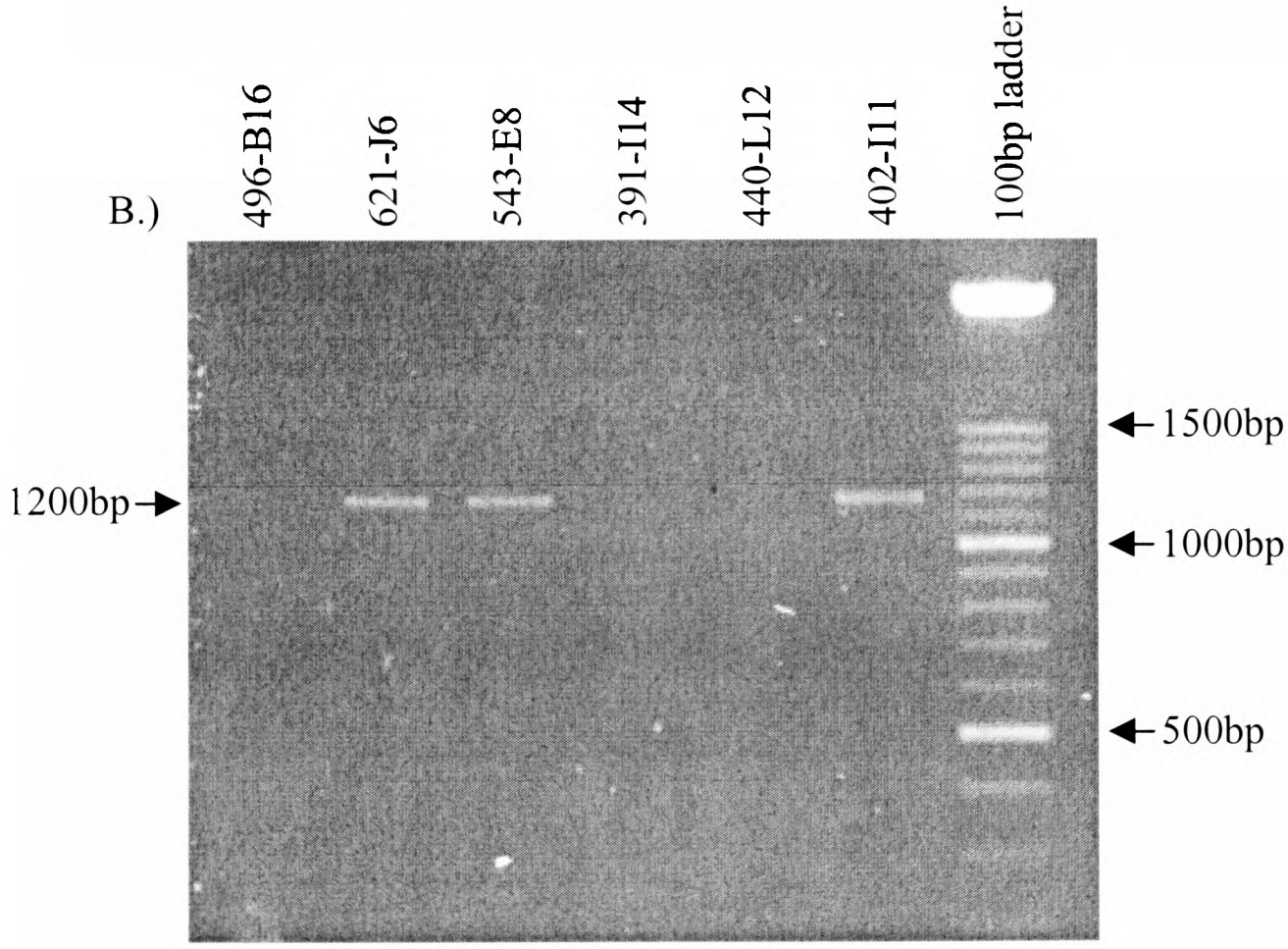


Figure 3.22: Isolation of Pdha1 PAC clones
The Mouse PAC RPCI21 genomic library was screened with the full length human cDNA (PDH1c) to isolate PAC clones containing the gene. A panel containing 2 of the 19 positives is shown (A). The PAC DNAs were screened by PCR using primers specific for X-linked exon8 in order to differentiate X-linked from autosomal clones. Clones 621-J6, 543-E8 and 402-I11 were shown to be X-linked (B).

(data not shown) indicating that they contained different fragments spanning the *Pdha1* gene region. The *PDHA1* gene spans approximately 17.5kb. Southern blotting data suggests that the mouse *Pdha1* locus is smaller (personal communications, R. M. Brown) and is likely to be contained within the PAC clones 621-J6, 543-E8 and 402I-11 as the average insert size of clones from this PAC library is 147kb.

3.3 Discussion.

The aim of this chapter was to develop the bisulphite sequencing procedure to study potential methylation at the promoter of the *PDHA1* gene. Methylation has previously been studied at the promoters of two other X-linked house-keeping genes which are subject to X-inactivation, namely *HPRT* (hypoxanthine phosphoribosyltransferase) (Hornstra and Yang, 1994; Park and Chapman, 1994) and *PGK-1* (phosphoglycerate kinase I) (Pfeifer et al., 1990). These studies demonstrated that the promoters on the inactive X-chromosome were hyper-methylated with respect to those on the active X-chromosome. In the case of *PGK-1* 98.3 % (118/120) of CpG sites in a 450bp section of the promoter were methylated. The two methylation studies at the *HPRT* promoter disagreed about the extent of methylation. Park et. al. (1994) found that 69% (22/32) of CpG sites in a 400bp section of the promoter were methylated whereas Hornstra et al. (1994) found that 82-94% (out of 142) of CpG sites in a 600bp section of the promoter were methylated. The active X-chromosome copies of both genes in males and females were totally unmethylated in the regions studied. As the *PDHA1* gene was also an X-linked house-keeping gene, subject to X-inactivation, it was suspected that its promoter might also be subject to methylation on the inactive X-chromosome.

CpG-plots of the *PDHA1* promoter region revealed an 858bp CpG-island at the 5'-end of the gene (-339 to +519bp) containing 95 CpG di-nucleotides so it was clear that the sites for

potential CpG methylation were present and at high density. The definition of functional elements of the promoter by (Chang et al., 1993) provided valuable information about the region which helped to decide the sections to be studied in msPCR, bisulphite and expression analysis of the promoter. Experiments testing the methylation status of restriction sites across a 590bp (-414 to +176bp) section of the promoter in HeLa cells suggested that methylation was indeed present at all 3 *Hpa II* and 8 *Hha I* sites. This indicated that the level of methylation across this whole region was likely to be high and consistent with that at the *HPRT* and *PGK-1* promoters. As in the *PGK-1* analysis, it was important to look at methylation in lymphocyte DNA samples from males and females to establish the prevailing methylation profile in non-cultured/non-transformed cells.

The msPCR data from lymphocyte DNA samples was consistent with the idea that males have unmethylated targets and that females have methylated targets, but that there was a degree of heterogeneity between females (only 3/5 female lymphocyte msPCR samples produced product bands). The ms-PCR data were however interpreted with caution because they depended upon the complete digestion of unmethylated targets. This was a significant consideration as the DNA was derived from blood which can easily be contaminated prohibiting efficient restriction digestion. The method adopted for DNA isolation from blood (John et al., 1991) consistently yielded DNA that was suitable for successful PCR and generated smears when digested with restriction endonucleases, but it was not possible to conclude that the DNA was completely digested. The sensitivity of PCR as a technique therefore does not exclude the possibility that the three positive bands produced out of the five female samples were a result of incomplete digestion rather than of methylation.

DNA from sperm is thought to be highly methylated at many gene sequences including those of the X-chromosome (Erickson et al., 1993; Groudine and Conkin, 1985; Rahe et al., 1983). The msPCR data presented here suggest that the promoter of the *PDHA1* gene may be highly methylated in sperm. *PDHA1* is known to be inactive in sperm raising the possibility that the observed methylation at the promoter may be involved in the inactivation process. During development, genomic methylation patterns are altered from those of the gametes. In general DNA of extra-embryonic tissues is hypo-methylated and that of the fetus is subject to *de-novo* methylation and is hyper-methylated (Monk et al., 1987; Razin et al., 1984; Sanford et al., 1985). Consistent with this the msPCR performed on a male CVS sample showed evidence of hypo-methylation at the PDHA1 promoter relative to that found in sperm. Methylation analysis of DNA from these cells would be extremely useful in defining which of the methylated residues were important in the control of the gene's expression. The idea of using the limited methylation levels observed in extra-embryonic tissues to help define the importance of particular methylated sites at the promoters of genes has been suggested previously (Sanford et al., 1985).

The bisulphite genomic sequencing technique ("Frommer" sequencing) (Frommer et al., 1992) was chosen as the method by which to analyse methylation at the promoter of the *PDHA1* gene. The analysis was carried out with a variety of modifications on the original method (see Methods 2.47) as described in the literature (Clark and Frommer, 1997; Clark et al., 1994; Feil et al., 1994; Sadri and Hornsby, 1996) and in private communications of the DNA Methylation Society (Woodcock, D. M.). Modern methods of modification of DNA by bisulphite have largely removed concerns about achieving total treatment of the template and have also improved the integrity of treated DNA (Raizis et al., 1995). With the advent of

cheap and reliable means of cloning and sequencing PCR products the last remaining hurdle to successful bisulphite sequencing is the actual PCR reaction itself.

The problem with designing primers for bisulphite PCR is that the nature and composition of the template is very different from those usually encountered in PCR reactions such as genomic DNA, plasmid DNA or first strand cDNA. The composition of the genome is approximately 58%(AT)/42%(CG) (Solage and Cedar, 1978) and only approximately 4% of the cytosines in the genome are methylated (Woodcock et al., 1987). It follows that after bisulphite treatment, the composition of the genome will be on average approximately 78%(AT)/22%(CG) of which only 0.8% will be cytosine. The template for PCR is therefore extremely AT rich and of lower complexity than a normal genomic template. The result is that it is difficult to design primers that will generate specific products. The best way to overcome this problem is to design primers to the relatively sparse G-rich regions of the treated template. The fact that the primers are G-rich and designed towards an A-T rich template means that they can anneal to their correct targets at a temperature where the majority of the treated template will not have annealed. This is why high specificity can be achieved in bisulphite PCR reactions with annealing temperatures as low as 45°C.

Bisulphite sequence analysis of the *PDHA1* promoter from HeLa cell DNA showed that there were two classes of sequence present. The first class had no methylated residues and the second class had 22/34 CpG sites methylated (64.7%) with the density of methylation increasing towards the start of transcription. This data is consistent with three possible conclusions: 1) the active X-chromosome has a methylated promoter, 2) the inactive X-chromosome has a methylated promoter or 3.) Both active and inactive X-chromosomes have methylated and unmethylated promoters. It is highly likely that the second of these

possibilities is correct, that the methylated class of HeLa bisulphite clones are derived from the inactive X-chromosome and that the unmethylated are from the active X-chromosome. This conclusion is supported by the methylation analyses of the *PGK-1* (Pfeifer et al., 1990) and *HPRT* promoters (Hornstra and Yang, 1994). These studies distinguished the methylation patterns of the active and inactive X-chromosomes by using hamster/human somatic cell hybrids containing either the active or inactive human X-chromosome as the only human derived material (Fuscoe et al., 1983; Hansen et al., 1988). The conclusion is further supported by the findings that no male lymphocyte DNA msPCRs or bisulphite sequences showed any evidence of methylation implying that the *PDHA1* promoter is unmethylated on the active X-chromosome. It is important to state that to date none of the publications assessing methylation at the promoters of X-linked genes have formally demonstrated that the promoter of the inactive X-chromosome is hyper-methylated and the promoter of the active X-chromosome is hypo-methylated in normal uncultured cells of either man or mouse. The deduction is always, as is made here, that the appearance of both methylated and unmethylated promoters in female DNA and of only unmethylated ones in male DNA suggests that it is the promoter on the inactive X-chromosome that is methylated.

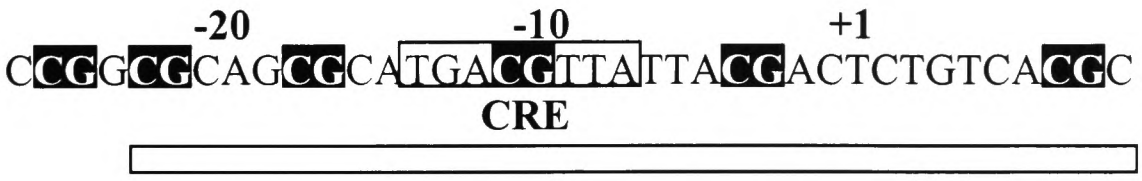
The position of the methylated sites found in bisulphite sequences of HeLa cells were analysed to see if they corresponded to consensus binding sites of known transcription factors (Figure 3.12). A more detailed study was made possible by incorporating data from the work on the *PDHA1* promoter performed by (Chang et al., 1993). This work identified sections of the promoter that were protected from DNase I digestion in footprinting experiments with rat liver nuclear extracts. Table 3.4 (below) shows the positions of four protected sites (P1-4) within the region covered by bisulphite analysis along with the transcription factor binding sites found within them.

P1	-41 to +17bp	TATA, CRE, AP-2
P2	-70 to –47bp	SP1, CACCC
P3	-113 to –76bp	CAAT
P4	-163 to –118bp	CACCC, SP1, AP-2

Table 3.4 Position of protein binding sites in the *PDHA1* promoter.

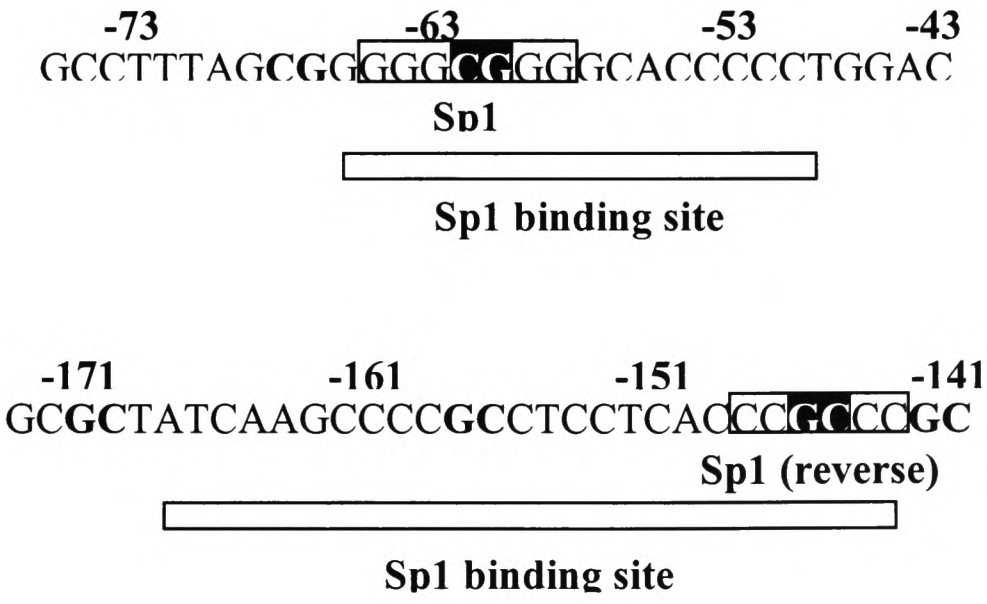
The section of the promoter from –23 to +10bp lies within the P1 protected region and contains a TATA box (TATTA) and an overlapping CRE site (TGACGTTA). CRE (cAMP responsive element) sites can potentially bind a large number of different bZIP protein family members which are regulated by second messengers such as cAMP and TPA (12-O-tetradecanoyl phorbol-13-acetate) (Habener, 1990). The proteins c-Fos and c-Jun are members of the bZIP family and have been shown to bind to the CRE motif (Curran and Franza, 1988) and to interact directly with TBP in a positive manner (Ransone et al., 1993). Purified c-Fos/c-Jun heterodimers were used in DNase I footprinting experiments of the promoter and were found to protect the CRE site (Chang et al., 1993). This supports the idea that c-Fos/c-Jun heterodimers bind to the CRE site in the *PDHA1* promoter and positively interact with the TATA binding protein (TBP) to up-regulate the level of gene expression.

Comparison of the bisulphite sequence of the presumed inactive X-chromosome of HeLa cells to this region of the promoter showed that all 6 of the CpG sites in the c-Fos/c-Jun binding region were methylated (below).



It has been reported that methylation of the CRE element blocks the binding of specific factors and ablates gene activity of the somatostatin gene (Iguchi-Ariga and Scaffner, 1989). If this were also the case for the CRE element in the promoter of the *PDHA1* gene, then this would represent a compelling mechanism for its down-regulation on the inactive X-chromosome.

There are further positive regulatory elements in the *PDHA1* bisulphite target region including two Sp1 sites. Both of these sites are methylated at the central CpG di-nucleotide in HeLa bisulphite sequences from the inactive X-chromosome (below).



The transcription factor Sp1 contains glutamine rich activation domains that positively interact with TBP and other TAFs to enhance levels of transcription (Emili et al., 1994). A

number of studies have investigated the effect of methylation on the ability of Sp1 to bind to its target sequence and found that methylation has no effect on binding (Harrington et al., 1988; Holler et al., 1988) although this finding is contested by a recent report (Clark et al., 1997). The two Sp1 sites in the proximal region of the *PDHA1* promoter have been shown to be protected from DNase I digestion in footprinting studies with both liver cell nuclear extract and purified recombinant protein (Chang et al., 1993). These data suggest that the Sp1 sites may be involved in the enhancement of transcription levels of *PDHA1 in vivo*. It is possible that the methylation of these sites on the inactive X-chromosome of HeLa cells may block Sp1 binding directly.

The final major element of the promoter in the bisulphite target region is an AP-2 site which is methylated at one of three CpG di-nucleotide sites in the binding region and at one further site 5bp distal to it in HeLa bisulphite sequences from the inactive X-chromosome (below).



AP-2 binding site

The transcription factor AP-2 is another glutamine rich transcription factor similar to Sp1 that has a positive regulatory function on enhancer elements (Williams et al., 1988). Methylation has been shown to interfere with the binding of the AP-2 protein to its target sites (Hermann and Doerfler, 1991). The AP-2 site at –148 to –139bp in the *PDHA1* promoter is protected from DNase I digestion in footprinting experiments with both liver cell nuclear extract and purified recombinant AP-2 (Chang et al., 1993). These data suggest that the element represents a genuine target for the AP-2 protein and that the site may be used as an enhancer

of gene expression *in vivo*. The partial methylation of this site on the inactive X-chromosome of HeLa cells may contribute to the down-regulation of the gene's expression.

The function of methylation at the binding sites of transcription factors in the promoter of *PDHA1* has been argued on the basis of its ablation or attenuation of protein binding. An alternate mode of interference of transcription factor binding is widely established in the literature. This mode involves the recruitment of specific CpG binding proteins to sites of DNA methylation (Kass et al., 1997). The effects of such recruitment are potentially two-fold. Firstly, the binding of methyl CpG binding proteins (MBPs) such as MeCP2 to methylated sequences may directly block the access of transcription factors to their sites. Secondly, MBPs can recruit other proteins such as mSin3A and histone deacetylases (e.g. HDAC1 and HDAC2) which directly alter the local chromatin structure to prevent the access of transcription factors (Pazin and Kadonaga, 1997; Ura et al., 1997; Wolffe, 1997). In conclusion, the observed methylation at the CRE, Sp1 and Ap-2 sites of the *PDHA1* promoter on the inactive X-chromosome of HeLa cells is consistent with a role for methylation in the inactivation of the *PDHA1* gene on the inactive X-chromosome.

Bisulphite sequencing of HeLa cell DNA at the *PDHA1* locus suggested that DNA from female lymphocytes would contain methylated and unmethylated alleles. It was expected that bisulphite sequences derived from male lymphocyte DNA samples would show no evidence of methylation consistent with the active X-chromosome being unmethylated at the *PDHA1* promoter. This would have been analogous to methylation at the *HPRT* and *PGK1* housekeeping gene promoters (Hornstra and Yang, 1994; Park and Chapman, 1994; Pfeifer et al., 1990). The male lymphocyte DNA samples were found not to be methylated at the *PDHA1* promoter in either msPCR or bisulphite sequencing experiments. Female lymphocyte DNA samples were found to contain some methylated sites in msPCR experiments, but no

methylation was evident in any bisulphite sequences. This inconsistency could be explained by one of three possibilities; 1) there were no methylated copies of the *PDHA1* promoter in female lymphocytes (the female msPCR results were false positives), 2) methylated copies were present but at a very low frequency or 3) amplification of bisulphite treated samples was biased towards unmethylated targets.

It is impossible to control for the complete digestion of unmethylated target molecules in msPCR experiments making the procedure prone to false positives. DNA isolated from blood is often contaminated with components such as EDTA and Haem that inhibit enzymatic reactions such as restriction digests and ligations. The msPCR products generated from 3/5 female lymphocyte DNA samples could genuinely have been produced from methylated target molecules that were not digested by *Hpa II* and *Hha I*. It is possible however that they were false positives and that there were no methylated *PDHA1* promoters in the female lymphocyte samples used. Consistent with this, sequencing of 15 independent bisulphite clones from the female lymphocyte DNA samples found no evidence of methylation at the *Hpa II*, *Hha I* or any other sites.

To test whether there was methylation on a small number of X-chromosomes from female lymphocyte DNA a complementary approach to direct sequencing was used. Four potential restriction sites for the enzyme *Tsc I* were present in the *PHDA1* bisulphite target region. As the enzymes target is the sequence ACGT, *Tsc I* sites can only exist if the central CpG di-nucleotide is methylated, otherwise bisulphite treatment will convert the sequence to ATGT which the enzyme cannot cut. Although this procedure will clearly only test four out of a possible 34 CpG sites (12%) in the bisulphite target region, the last three of the sites lie within the region of the promoter in which 13/14 (93%) of the CpG sites are methylated on

the HeLa cell inactive X-chromosome. On this basis if there were methylated inactive X-chromosome sequences at low frequency in the bisulphite clones then it would be reasonable to expect at least some of these to be methylated at one or more of the *Tsc I* sites. In total 45 independent clones were tested either by complete sequencing (15) or by *Tsc I* digestion (30) and no evidence of methylation was detected. This suggested that no more than about 2% (1/46) of clones if any contained methylated sites. If this represented the level of methylation in the original DNA sample then the sensitivity of msPCR would be capable of generating products from the few copies which would be resistant to digestion. This could only be confirmed by manually sequencing a hundred or more bisulphite clones which was neither practical or economical in the absence of strong evidence that some methylated clones would be found.

Another possibility was that the bisulphite PCR amplification was skewed so that only unmethylated clones were amplified. This would mean that the msPCR results could be correct and that the bisulphite sequences were not representative of the original DNA samples. The bisulphite sequencing experiments carried out on lymphocyte DNA were exactly the same as for HeLa DNA and the two were frequently carried out in parallel. The procedure generated both methylated and unmethylated classes of bisulphite sequence from HeLa DNA, but only ever produced unmethylated sequences from lymphocyte DNA. This demonstrated that there was no significant amplification bias between methylated and unmethylated targets in the bisulphite PCR reaction itself. The absence of bisulphite sequences representing methylated copies of the *PDHA1* promoter in lymphocyte DNA could therefore only be explained by differences between the DNA samples. This difference was either due to methylation differences between the two sample types or to an inhibitor present in the blood DNA samples that affected amplification of bisulphite products from methylated

but not unmethylated templates. There is no evidence to suggest that any such inhibitor exists so it is reasonable to assume that the amplification is not heavily biased.

The data from male and female lymphocyte DNA samples suggest that the promoter of the *PDHA1* gene was unmethylated on both the active and inactive X-chromosome. It remains possible however that there was methylation at the *PDHA1* promoter in the female lymphocyte samples, but that it was only present in a small proportion of molecules. Methylation levels on the inactive X-chromosome of female lymphocytes have not been extensively studied, but the study of the *PGK-1* gene (Pfeifer et al., 1990) commented that the promoter on the inactive X-chromosome of female lymphocytes was variably hypo-methylated. It would be interesting to determine by other independent methods whether there are unexpected methylation variations in blood derived DNA as this is widely used in diagnostic procedures some of which now involve determination of methylation differences at imprinted loci.

Evidence from HeLa cells suggested that methylation at the promoter of the *PDHA1* gene might be capable of down-regulating gene expression. In order to test this hypothesis reporter gene expression constructs were generated with a section of the promoter, including the bisulphite target region, cloned into the vector pEGFP-1. The vector pEGFP-1 contains the coding sequence for a variant of the *Aequorea victoria* green fluorescent protein which has been modified for maximum expression in mammalian cells and for optimal brightness (Cormack, 1996; Yang, 1996). The vector is promoter-less and has a MCS positioned to allow cloning of DNA fragments upstream of the EGFP-1 coding sequence.

HeLa cells had been shown to express the *PDHA1* gene by RT-PCR (see Figure 3.16C) and as methylation patterns had previously been determined in these cells they were considered ideal for these transfection experiments. The EGFP-1 reporter gene constructs were transiently transfected into HeLa cells by electroporation, then cultured for 48 hours prior to fixation and viewing. The forward and reverse orientation promoter-reporter gene constructs were found to drive reporter gene expression in HeLa cells. Accurate quantitation of relative promoter activity was not necessary as the experiment was solely designed to score the presence or absence of green fluorescence in cells. These experiments demonstrated that the section of the promoter cloned into the pEGFP-1 vector was capable of performing the functions of a promoter. The fact that the reverse construct also generated some green cells was not entirely unexpected as the promoter section contains several promoter elements in the reverse orientation such as the Sp1 site at -148 to –143bp which could be capable of enhancing low-level transcription.

To test the effect of methylation on the ability of the promoter section to direct reporter gene expression, both forward and reverse constructs were methylated and then transfected into HeLa cells. As a control the constructs were also treated in "dummy" methylation reactions where either the methylating enzyme or the substrate was left out. This controls for the possibility that other factors in the preparation of the methylated constructs affected the expression of the reporter gene. In short, the sole difference between methylated and "dummy" methylated constructs was the methylation itself and not the treatment of the vectors. The experiments were carried out three independent times and on no occasion did methylated constructs ever produce green cells whereas "dummy" methylated constructs appeared identical to the results obtained with untreated constructs. In conclusion, the

methylation of the reporter gene constructs down-regulated the level of reporter gene expression to a level undetectable above background.

These results support the idea that, in this system, methylation is capable of repressing the promoter function of this section of DNA. The major concern about this interpretation of the effect of methylation is whether the methylated vector backbone also affects the function of the promoter. It is difficult to envisage a way in which the vector backbone methylation in itself could affect the promoter section, any repressive effects would have to be transmitted through intermediates that acted at distance. The main candidates for such a role would be histone proteins and associated factors, but there is no evidence to suggest that transiently transfected vector DNA actually assembles chromatin. Certainly numerous investigators have used systems identical to those used here and not observed such vector derived effects (Huntriss et al., 1997; Lu and Davies, 1997). In fact it is believed that only once cells become stably transfected and the construct DNA is incorporated into the genome can such effects take place (Paroush et al., 1990).

In conclusion, the transfection experiments demonstrate that the section of the *PDHA1* gene cloned into pEGFP-1 is capable of behaving as a promoter and methylation of this section *ex-vivo* suppresses reporter gene activity to levels undetectable above background.

A more detailed analysis of methylation at the promoter of the *PDHA1* gene in a variety of human somatic tissues would be informative as it would allow definition of the variability of the methylation profile between tissues and individuals. The difficulty of obtaining numbers of different fresh tissue samples from multiple individuals is problematic in humans, but not in mice. The cloning of the mouse *Pdha1* promoter would provide a valuable resource for the

future analysis of the effect of methylation on the genes expression because it would allow a broad analysis to take place. To generate such a resource for future study, independent PAC clones were isolated from a murine genomic library using the full length human *PDHA1* cDNA as a probe. The positive clones were isolated and screened by PCR to differentiate X-linked (*Pdha1*) from autosomal (Pdha2) clones. The primers for this screening PCR were derived from sequences flanking exon8 of the *Pdha1* X-linked gene which are absent from the intron-less Pdha2 gene. Three X-linked clones were identified and were found to have different restriction patterns, indicating that they were overlapping. As the average insert size in the PAC library is 147kb (see Materials 2.11) and the estimated size of the murine gene is less than 17.5kb (personal communication R. M. Brown) it is expected that the X-linked PAC clones contain the entire *Pdha1* gene.

The methylation profiles determined in this chapter suggest that the *PDHA1* promoter on the female inactive X-chromosome is hyper-methylated and that on the active X-chromosome is unmethylated in HeLa cell DNA. HeLa cells may not have the same methylation profiles found *in-vivo* so it is essential that future work defines methylation in primary tissues. Transfection experiments with *PDHA1* promoter-reporter constructs have shown that *ex-vivo* methylation is capable of down-regulating reporter gene expression implying that methylation plays a role in maintaining inactivity of the gene on the inactive X-chromosome. The isolation of PAC clones containing the murine *Pdha1* gene have provided a resource for the isolation and analysis of the murine *Pdha1* promoter to allow detailed comparative study of the promoters of the human and murine pyruvate dehydrogenase E1α genes.

Chapter 4 - Methylation Analysis of the Human and Murine Norrie Disease Protein (*NDP/Ndp*) Gene Promoters

4.1 Introduction.

The work described in this chapter details the analysis of DNA methylation at the human and murine Norrie disease protein gene promoters. Analysis of methylation is carried out using the bisulphite sequencing ("Frommer"-sequencing) technique on genomic DNA isolated from tissue culture cells and tissue samples. The promoters of the human and murine genes are cloned upstream of the EGFP reporter gene and transfected into tissue culture cells known to express the *NDP/Ndp* genes to confirm that the sections studied act as functional promoters. In addition ex-vivo methylation of promoter constructs is carried out to determine the effect of methylation on the expression of the reporter gene in tissue culture cells. These studies represent the first by bisulphite sequencing on an X-linked gene in expressing and non-expressing tissues.

4.1.1 Tissue-specific genes on the X-chromosome.

The promoters of tissue specific genes on the X-chromosome are very similar in composition to their autosomal counterparts. Typically they have TATA and CAAT boxes, but lack the consensus binding sites of transcription factors such as Sp1 sites that are abundant in the promoters of housekeeping genes. Instead promoters of tissue specifically expressed genes tend to contain binding sites for transcription factors that are themselves tissue specifically expressed. This in part helps to achieve the co-ordinated control of groups of tissue specific genes by transcription factors that are switched on during the development of specific cell-lineages. Another feature of tissue specific genes is that approximately 40% are associated

with CpG-islands at their 5'-ends. These islands were thought to be unmethylated, but recent studies at the CpG-islands associated with the galectin-1 and tissue transglutaminase gene promoters have shown them to be methylated in non-expressing tissues and cell-lines (Lu and Davies, 1997; Salvatore et al., 1998).

The murine galectin-1 gene is normally expressed in heart and skeletal muscle, but expression is undetectable in the liver, thyroid and brain. Methylation of the promoter was analysed in these tissues and in tissue culture cell lines and found to correlate with the expression status of the gene (Salvatore et al., 1998). Interestingly in tissue culture cells the methylation was of an all or nothing form whereby expressing cells had totally unmethylated promoters, non-expressing cells were methylated at all promoter CpG sites. In murine tissues the correlation between unmethylated sites and expression was still observed, but the unmethylated sites in expressing tissues were only present in the core promoter area with sites upstream being totally methylated. This was taken to mean that only unmethylated sites in the core promoter section were required to allow expression of the gene *in-vivo*.

The study of methylation at the promoter of the human tissue transglutaminase gene was carried out on DNA from expressing and non-expressing tissue culture cells. The authors suggest that their results are consistent with the gene being controlled by methylation and predict that selective de-methylation may regulate the level of expression of the gene in different tissues *in-vivo* (Lu and Davies, 1997).

In the light of these studies and others, similarity in the composition of promoters of autosomal and X-linked tissue specific genes raises the possibility that methylation might also play a role in the tissue specific expression of X-linked genes. This would be an

extremely complex form of gene control as the tissue-specific control of the gene would have to operate simultaneously with any control imparted as part of the X-inactivation process.

Many tissue specific X-linked genes have been cloned, but few have had the methylation status of their promoters examined. The first such methylation studies were on the promoters of the factor IX (Cullen et al., 1986) and ornithine transcarbamylase (Otc) genes (Mullens et al., 1987). Both promoters contain TATA and CAAT boxes, but neither promoter has a high (C+G)% content or a CpG-island. The analysis of methylation at the factor IX gene promoter by restriction digestion suggested no difference in methylation between active and inactive X-chromosomes or between expressing and non-expressing tissues. Methylation analysis of the murine Otc gene, however, suggested that there might be a relationship between hypo-methylation and tissue specific expression of the gene. More recently methylation has been studied at the promoter of the *FMR1/Fmr1* gene by ligation mediated PCR (Schwemmle et al., 1997). The CpG-island at the promoter of the *FMR1/Fmr1* gene was found to be hyper-methylated on the inactive X-chromosome and hypo-methylated on the active X-chromosome in tissue culture cells. Preliminary methylation analysis in expressing and non-expressing tissues suggested that methylation of the promoter correlates with reduced expression of the gene.

4.1.2 The Norrie disease protein (*NDP/Ndp*) gene.

Norrie disease is a severe X-linked recessive, neuro-developmental disorder characterised by congenital blindness (Norrie, 1927). The disease manifests early during development and results in malformation of the retina. Patients with Norrie disease often have additional symptoms including sensorineural hearing loss, mental retardation, dementia or psychotic features (Warburg, 1966). The *NDP* gene was isolated by screening human adult and fetal

retinal cDNA libraries with fragments of a YAC known to contain a marker closely linked to the disease locus. A 1.8kb evolutionarily conserved cDNA was isolated which contained an ORF that encoded a predicted protein of 133 amino acids (Berger et al., 1992; Chen et al., 1992). The entire gene was isolated and found to span 27kb comprising three exons, the first of which is untranslated. The murine homologue of *NDP* was isolated by screening a murine genomic phage library with the human cDNA and was found to have an organisation almost identical to the human gene. The coding sequence of the gene shares 95% sequence homology to the human over the entire sequence with 99% identity in the third exon.

The NDP is thought to be homologous to the von Willebrand factor (vWF) (Bonthron et al., 1986), the *Drosophila melanogaster* protein *slit* (Rothberg et al., 1990) and connective tissue growth factor (CTGF) (O'Brein et al., 1990). Protein modelling studies suggested that the protein had a three-dimensional structure very similar to transforming growth factor β (TGFβ), nerve growth factor (NGF) and platelet derived growth factor (PDGF) (Meitinger et al., 1993). The predicted structure of the NDP includes a cysteine knot motif which makes it most similar to TGFβ and implies that it also exists as a dimer.

Expression analysis of the *NDP* gene by northern blotting shows that expression is primarily limited to the adult and fetal brain. *In-situ* experiments in murine tissues show that the *Ndp* gene is expressed highly in the inner nuclear and ganglion layers of the retina and in parts of the cerebellum (Chen et al., 1995; Meitinger et al., 1993). The putative role of the NDP as a peptide growth factor and its tissue specific expression are consistent with the classic symptoms of Norrie disease. Since the cloning of the *NDP* gene, more than 30 point mutations have been found in Norrie patients within the coding sequence confirming that the mutations in the *NDP* gene are responsible for Norrie disease (Berger, 1998).

The human and murine *NDP/Ndp* gene promoters have been isolated and used in comparative studies and to define *cis*-regulatory elements involved in the control of the human genes expression (Kenyon and Craig, 1999). Expression constructs containing different sections of the human *NDP* gene promoter were cloned upstream of the luciferase reporter gene and used to assess the effects of different sections of the promoter on luciferase expression. The region –109 to +55bp of the *NDP* promoter was shown to contain the core promoter activity with a further section from –106 to –219bp containing a putative repressing region. The first 160bp of upstream sequence of the human and murine gene promoters were found to be 90% identical and to contain a number of conserved consensus binding sites for known transcription factors. One of these factors was the Neural Retina Leucine zipper (NRL) protein.

The NRL gene is expressed in cells of the human retina and encodes a DNA binding protein of the basic leucine zipper (bZIP) family (Swaroop et al., 1992). The leucine zippers are a family of transcription factors that contain both DNA binding and activation domains that interact directly with the TATA box binding protein. Other members of the bZIP family include the proteins CREB, c-Jun and c-Fos which can heterodimerise in specific ways with each other to modify their sites and modes of action (Kerppola and Curran, 1994). It has been postulated that the NRL protein can heterodimerise with either c-Fos or c-Jun as well as homodimerising with itself. Many of the closely related bZIP factors are nuclear targets of the cAMP signal transduction pathway and have been implicated in many aspects of the physiology of the neuro-endocrine system. The NRL protein has been proposed as a key factor involved in regulating the tissue-specific expression of the human *NDP* gene (Kenyon and Craig, 1999).

4.2 Results.

4.2.1 CpG-Plots.

The CpG-Plot program (see Methods 2.56 and 2.56.6) was used to analyse the promoter sequences of the human and murine *NDP/Ndp* genes. In addition the data generated by the program was extracted and further analysed as described previously (Chapter 3 section 3.2.1). The program calculates the relative CpG di-nucleotide abundance and the (C+G)% composition of a given section of DNA sequence and determines whether a CpG-island is present within it. The definition of a CpG-island is taken to be a section of sequence which has a CpG relative di-nucleotide abundance ≥ 0.5 and a (C+G)% content ≥ 50 for an uninterrupted stretch of 200 nucleotides (Gardiner-Garden and Frommer, 1987). If these criteria are satisfied the program reports the exact position of any such islands. This method of defining CpG-islands will certainly preclude the assignation of a CpG-island to a sequence which satisfies the criteria over a large area, but which drops below the criteria for just a few nucleotides. This could potentially lead to regions which are of unusually high CpG content being overlooked. To avoid this a different method of representing the data generated by CpG-Plot was adopted. The program analyses each sequence in windows of fixed size which are moved one nucleotide at a time across the whole sequence. For each window a value is returned for both CpG di-nucleotide relative abundance and (C+G)% composition. These window results are then checked against the criteria for a CpG-island and if the criteria are satisfied by 200 consecutive windows an island is declared. Normally any section which does not meet these criteria is essentially ignored and no information is provided about sequences that narrowly miss them. To acquire this additional information from the program the window results were extracted and checked individually to see whether any single window satisfied the individual criteria for abundance and composition, if so then a single point was

plotted for that window position on a graph. This approach was applied across the whole sequence thereby giving a graphical representation of the "island-potential". These results of course would include any islands as defined by the original program, but would also include all other points satisfying the criteria.

Figure 4.1 shows the analysis described above for the whole human *NDP* promoter region available spanning the region –1380 to +438bp. The section contains 32 CpG sites, has an overall (C+G)% content of 39.0 % and contains 1 *Hpa II* and 3 *Hha I* sites. The window analysis of the island criteria demonstrates that there is a section of the promoter from +265 to +213bp (52bp) and a region from –167 to +5bp (172bp) which continuously satisfy the criteria. The second region is very nearly a CpG-island by the definition of Gardiner-Garden (1987), it is 172bp long, contains 13 CpG sites and has a (C+G)% content of 55.8 %. The preceding 28bp of sequence have an average CpG relative abundance of 0.77 and an average (C+G)% content of 42.5 %. This 28bp stretch would meet the criteria to be classified as a CpG island if it contained an additional two cytosine or guanine residues.

Figure 4.2 shows the same analysis for the murine *Ndp* gene promoter in the region –1300 to +412bp the section homologous to the human section shown in figure 4.1. The section contains 18 CpG sites has an overall (C+G)% content of 39.5 % and contains no *Hpa II* and 2 *Hha I* sites. The window analysis of the island criteria reveals only one sizeable section from –130bp to –44bp (86bp) which continuously satisfy the criteria. This section lies within the 172bp part of the human promoter that narrowly misses classification as a CpG-island. Comparing the human 172bp section to the homologous mouse section reveals that the average (C+G)% content has fallen by 5.2 % (55.8 to 50.6) and approximately 40 % of the CpG sites have disappeared (13 to 8). This strikingly illustrates that the CpG sites are being

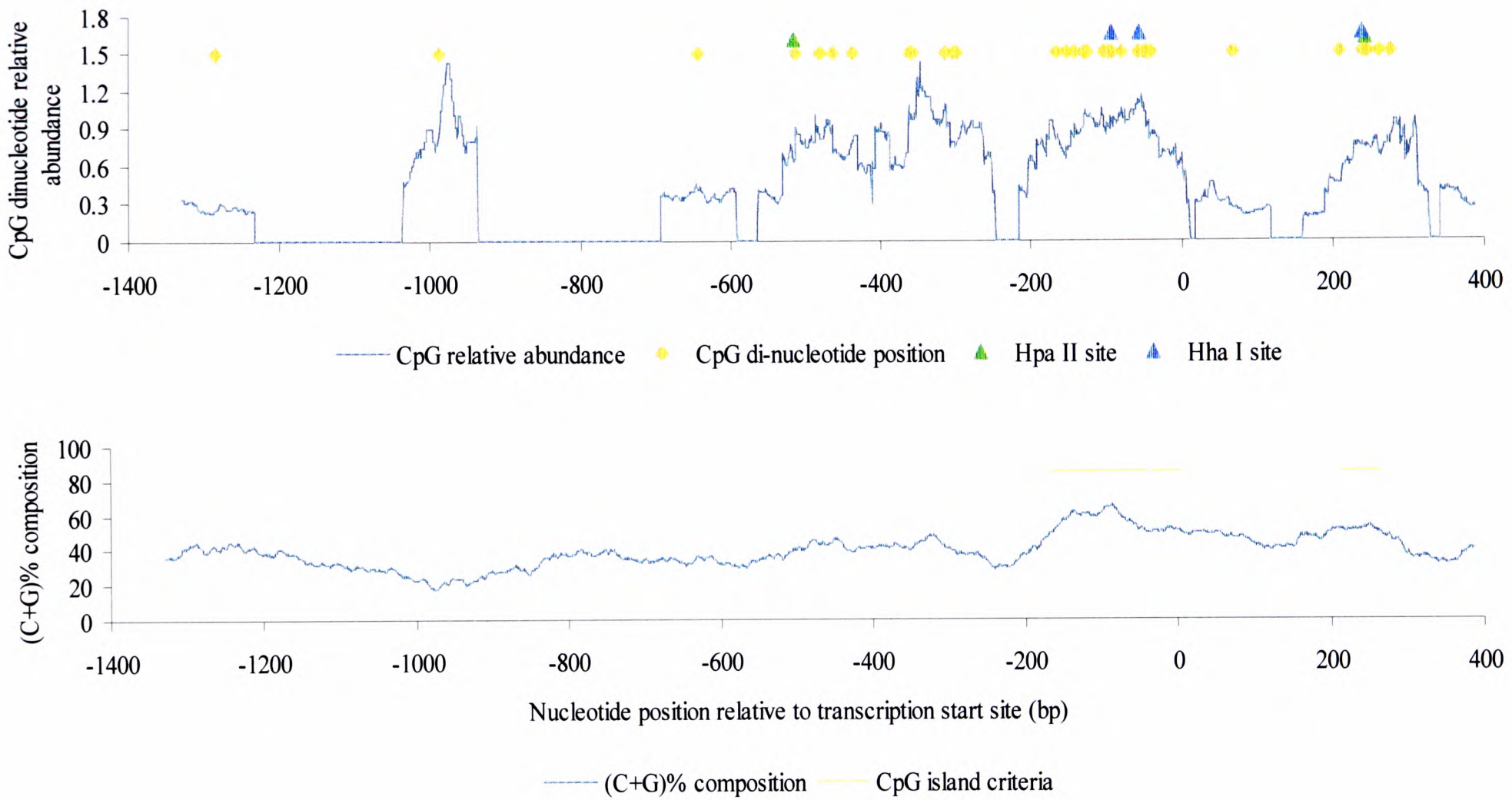


Figure 4.1: CpG-Plot analysis of the Human Norrie Disease Protein (NDP) gene promoter.
The program CpG-Plot (GCG-Wisconsin HGMP-RC) was used to analyse the promoter of the human NDP gene from -1380 to +438bp. The section of DNA has no regions which adhere to the criteria required of a CpG-island, but independent analysis of data extracted from the program (orange line) shows that the region from -167 to +5bp narrowly misses such requirements. The promoter is evidently CpG-rich with a concentration of CpG sites in the proximal promoter region. In total it contains 32 CpG di-nucleotides and 1 Hpa II and 3 Hha I sites.

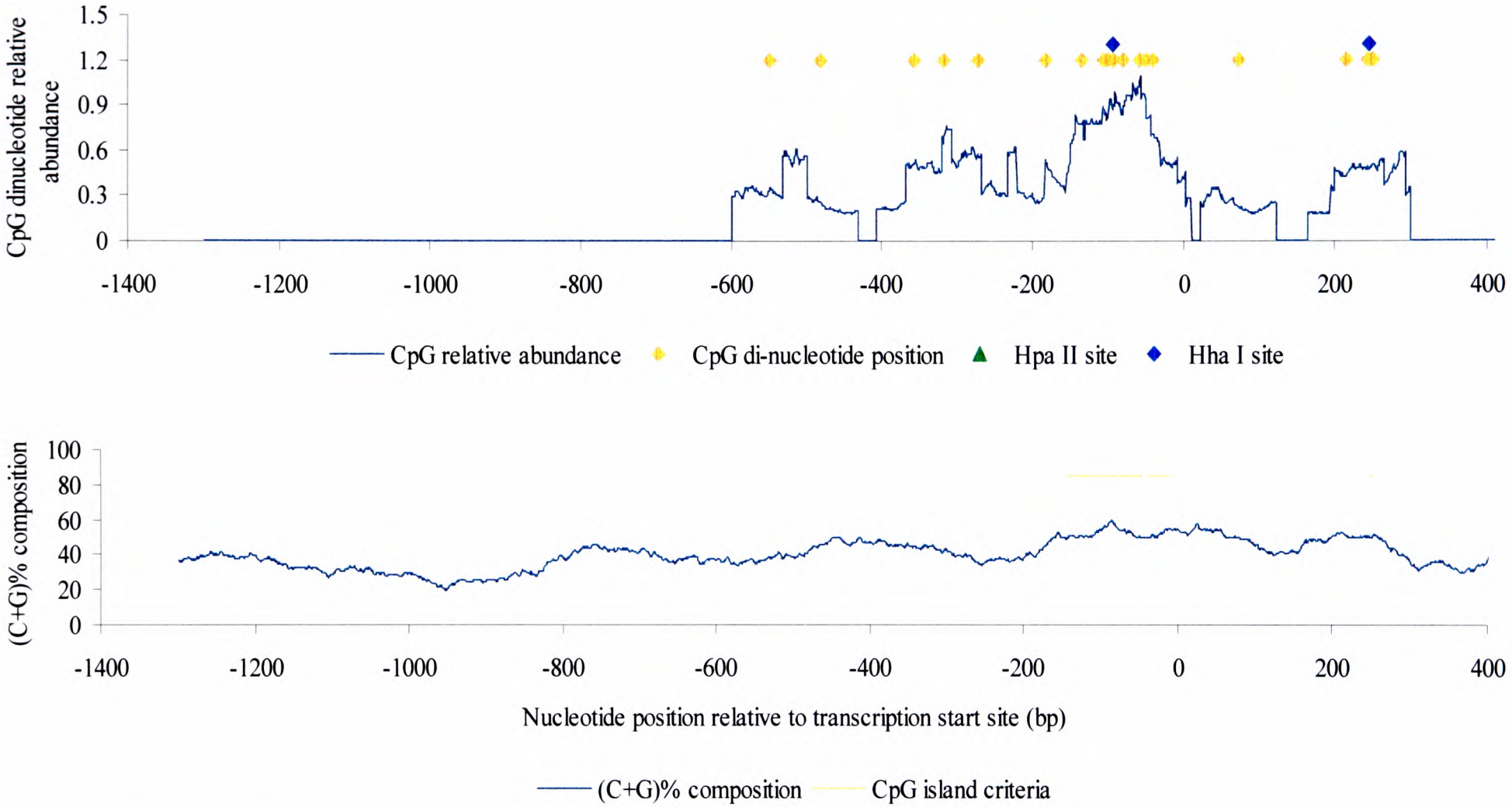


Figure 4.2: CpG-Plot analysis of the Murine Norrie Disease Protein (Ndp) gene promoter.
The program CpG-Plot (GCG-Wisconsin HGMP-RC) was used to analyse the promoter of the murine NDP gene from -1300 to +412bp which is homologous to the section represented in figure 4.1. The section of DNA has no regions which adhere to the criteria required of a CpG-island. Independent analysis of data extracted from the program shows that the only sizeable region that satisifes the criteria is from -130 to -44bp (orange line). The promoter is evidently CpG-rich with a concentration of CpG sites in the proximal promoter region. In total it contains 18 CpG di-nucleotides and 2 Hha I sites. The (C+G)% composition of the proximal 200bp of the promoter has fallen by 5.2% compared to the homologous human seuqnece and 40% of the CpG sites have disappeared.

lost at a rate approximately 7x greater than would be expected if the (C+G) content loss was evenly distributed. Over half (56%) of the (C+G) loss can be accounted for by the loss of C residues from CpG di-nucleotides.

Analysis of the promoter activity of the human *NDP* gene promoter has been carried out by collaborators using expression constructs where sections of the promoter drive the luciferase reporter gene (Kenyon and Craig, 1999). The core promoter activity was found to lie within a construct containing the region –106 to +55bp. A decrease in luciferase activity in constructs containing more upstream sequence led to the suggestion that the region –106 to –219bp contained one or more repressor elements. No further change in reporter gene expression was found in constructs containing more sequence upstream of –219bp (Kenyon and Craig, 1999).

The information derived from the CpG-plots of the two promoters and from the expression analysis of the human promoter was used to select a region for analysis by bisulphite sequencing.

4.2.2 Bisulphite sequence analysis of the human *NDP* gene promoter.

PCR primers were designed for the analysis of methylation at the promoter of the human *NDP* gene. The primers HNDPBISNF1 (5'-GTTTGAAAAAGAATTAAATTTAG-3') and HNDPBISR1 (5'-ACTTCCAAATACTTTTATTCTTC-3') amplify a 431bp product of the region –273 to +158bp which includes the core promoter of the human *NDP* gene. Figure 4.3 shows a CpG-plot of the bisulphite target region which includes 14 CpG sites, no *Hpa II* sites and 2 *Hha I* sites.

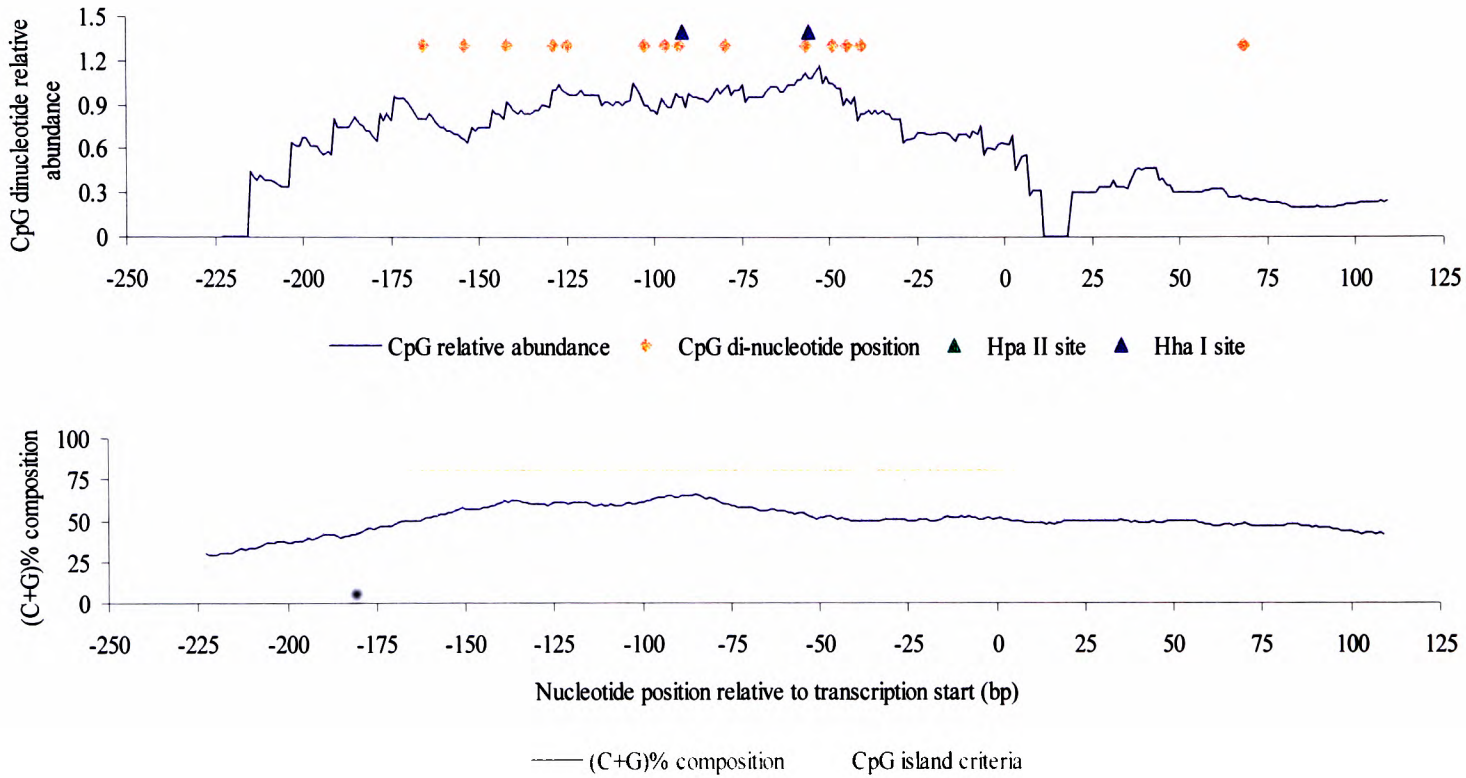


Figure 4.3: CpG-Plot of the Human NDP gene promoter bisulphite target region.
The region selected for bisulphite sequencing (-273 to +158bp) was analysed by the CpG-plot program. Independent analysis of data extracted from the program was used to identify windows of the sequence which adhered to the criteria for CpG abundance and (C+G)% composition required of a CpG island. When these two criteria were satisfied a point was plotted at the position of the window. This analysis was performed on the whole sequence (orange line) revealing regions of high CpG abundance. The bisulphite target region includes 14 CpG sites and 2 Hha I sites and very closely resembles the profile of a CpG-island containing promoter.

The *NDP* gene is expressed in a limited number of tissues including parts of the brain and retina . If methylation profiles from expressing and non-expressing alleles of the gene were to be defined it was essential to use DNA that was isolated from female sources known to express the gene. In order to distinguish the active and inactive X-chromosome patterns it was also necessary to isolate DNA from male cells or tissue and compare it to those sequences from female sources. The cell line Weri-Rb-1 (see Materials 2.14) is a female retinoblastoma cell line known to express the *NDP* gene (Kenyon and Craig, 1999). This cell line was used as a source of DNA for establishing the methylation pattern of the active and inactive alleles of the *NDP* gene in tissue culture cells.

i) Bisulphite sequence analysis of Weri-Rb-1 cell genomic DNA.

DNA was isolated from the human retinoblastoma continuous cell line Weri-Rb-1 (ATCC No. HTB-169) and this was used in PCR optimisation reactions with the 15 buffer series (see Materials 2.2 and Methods 2.44). Table 4.1 (below) describes the conditions of the optimised *NDP* gene promoter bisulphite PCR.

NDP bisulphite PCR (431bp product) HNDPBISNF1 : HNDPBISR1 Buffer B	96°C for 2'00'' 96°C for 0'30'' 45°C for 0'15'' 72°C for 0'30'' } 35 cycles

Table 4.1: Optimised PCR conditions for the human *NDP* promoter bisulphite PCR reaction.

Bisulphite PCR reactions were performed on WeriRb-1 DNA samples using the optimised conditions. These products were purified, *PfU* "polished" and blunt-cloned into the EcoRV site of the vector pZErO-2 as described previously (see Methods 2.39.2 and 2.39.3). The

ligations were purified and 10ng transformed by electroporation into electro-competent TOP10 *E.coli* cells. After plating onto selective plates and culturing, individual bacterial colonies were picked and screened by colony PCR using Sp6 and T7 primers. Insert carrying clones were cultured and plasmid DNA prepared from them ready for sequencing. Sequencing was carried out using the ABI BigDye™ fluorescent cycle sequencing kit (see Methods 2.42) with the Sp6 and T7 primers flanking the multiple cloning site of pZErO-2.

Figure 4.4 shows an ABI sequence trace of the Weri-Rb-1 bisulphite clone cl02 in which all CpG sites are unmethylated. Figure 4.5 shows a trace of the clone cl04 in which methylated CpG sites are present. These two sequences represent the two major classes of bisulphite sequence generated from Weri-Rb-1 DNA. Figure 4.6 shows an alignment of 8 sequences of independent bisulphite clones against the untreated genomic sequence. Sequences cl02, cl05 and cl15 are notably hypo-methylated whereas the other clones show almost complete methylation at all 14 CpG sites.

The Weri-Rb-1 cell line has been karyotyped and found to contain two X-chromosomes with no evident aberrations (McFall, 1977). Analysis of the bisulphite sequences generated from the cell line indicate that there is a single base (A/T) polymorphism at position –158bp. This allows the allocation of the sequences to one or other of the X-chromosomes of this cell line. Interestingly the hypo-methylated sequences all come from the sequence containing the "A" allele and all the hyper-methylated sequences from those containing the "T" allele. It is tempting to speculate that this cell line is derived from a clonal cancer in which the same X-chromosome is inactive in each cell. If this is indeed the case it seems likely that the hypo-methylated sequences are from the active X-chromosome and the hyper-methylated from the inactive X-chromosome.

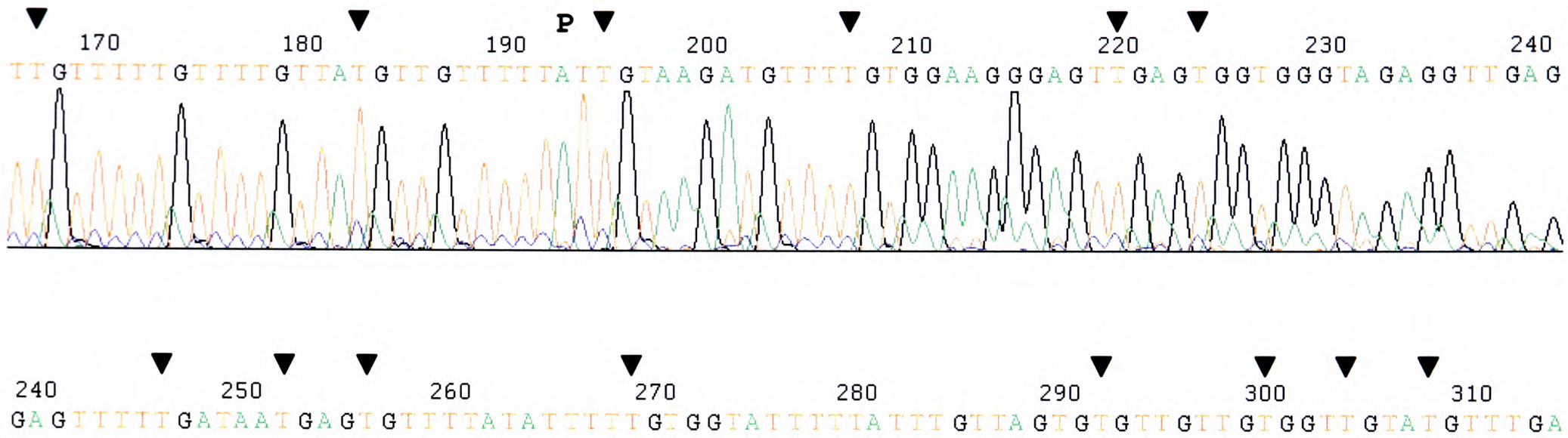


Figure 4.4: ABI sequence trace of the Weri-Rb-1 cell NDP bisulphite clone cl02
Weri-Rb-1 cell genomic DNA was used to analyse methylation at the promoter of the human NDP gene promoter. PCR was performed on bisulphite treated DNA and the resulting PCR products were cloned into the vector pZErO2. Independent insert carrying clones were sequenced with the Sp6 and T7 pZErO2 MCS flanking primers using the ABI BigDye fluorescent cycle-sequencing kit. Sequencing reactions were purified and run on an ABI377 DNA sequencing machine. The sequence above is of one of the clones cl02 which has none of its CpG sites (indicated with black triangles) methylated. The "A" allele of an (A/T) polymorphism present in this region of the promoter in the Weri-Rb-1 cell line is indicated (marked **P**).

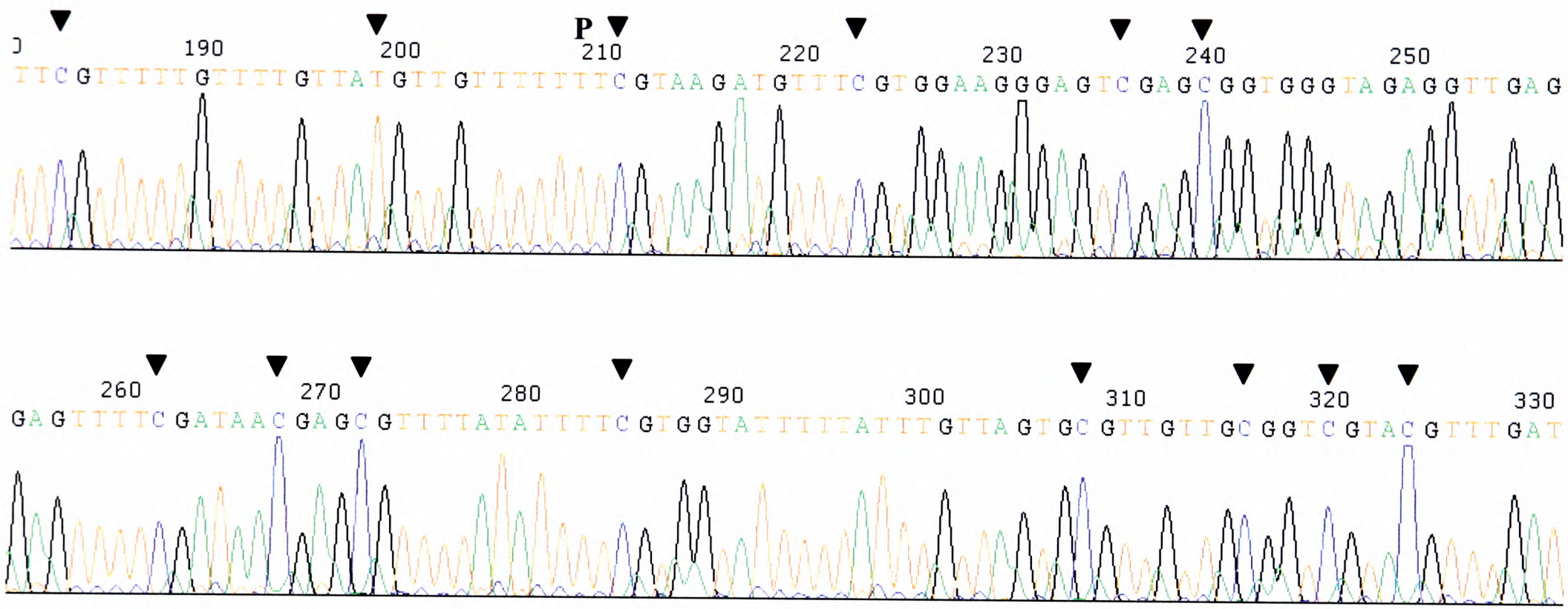


Figure 4.5: ABI sequence trace of the Weri-Rb-1 cell NDP bisulphite clone cl04
Weri-Rb-1 cell genomic DNA was used to analyse methylation at the promoter of the human NDP gene promoter. PCR was performed on bisulphite treated DNA and the resulting PCR products were cloned into the vector pZErO2. Independent insert carrying clones were sequenced with the Sp6 and T7 pZErO2 MCS flanking primers using the ABI BigDye fluorescent cycle-sequencing kit. Sequencing reactions were purified and run on an ABI377 DNA sequencing machine. The sequence above is of one of the clones cl04 which has 13 out of 14 of its CpG sites methylated (indicated with black triangles). The "T" allele of an (A/T) polymorphism present in this region of the promoter in the Weri-Rb-1 cell line is indicated (marked **P**).

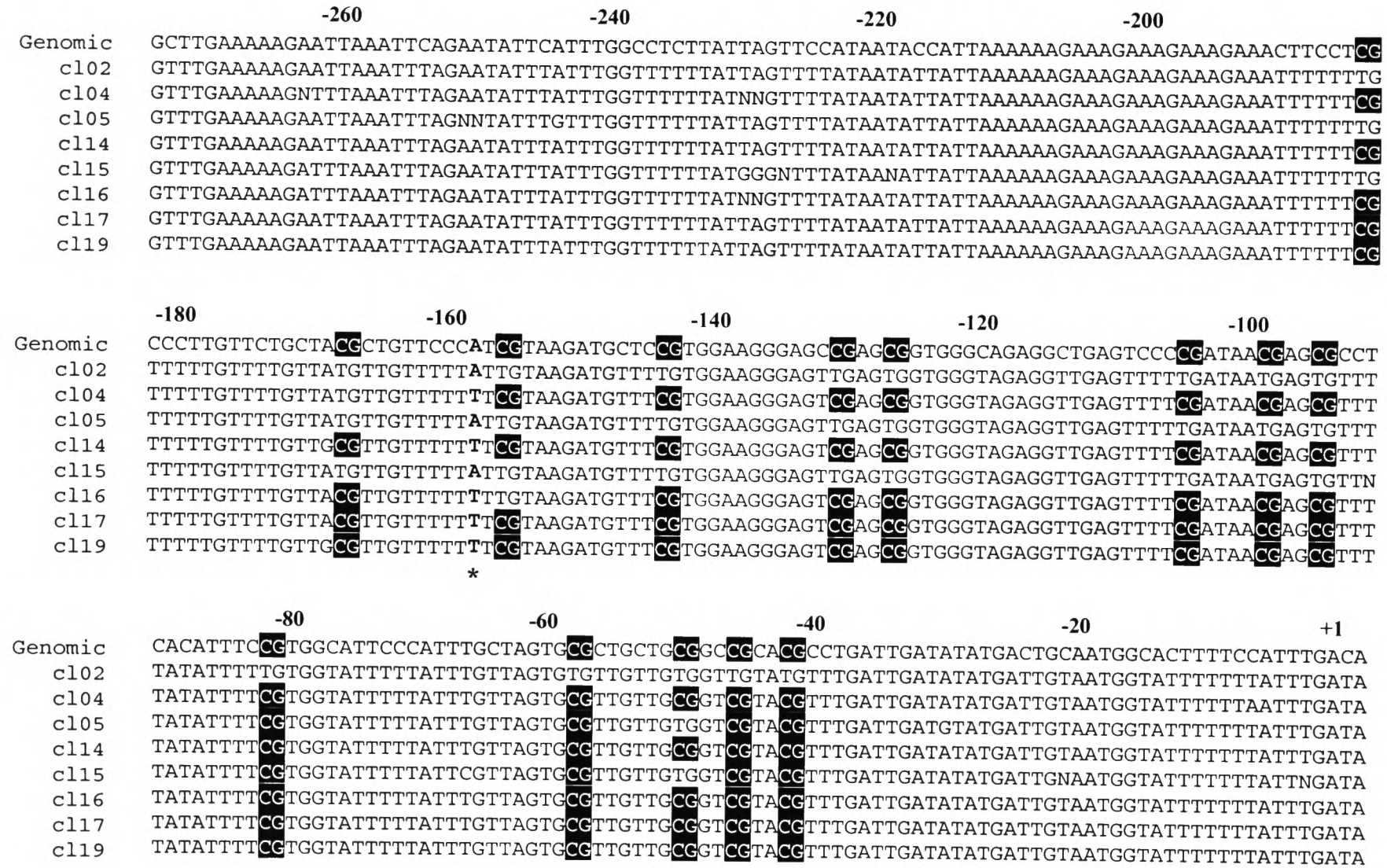


Figure 4.6: Alignment of Weri-Rb-1 cell bisulphite sequences of the human NDP gene promoter
Bisulphite sequences from independent clones are aligned against the genomic target sequence with CpG sites highlighted in black. The position of the polymorphism in the promoter sequence is indicated (*) and both A and T alleles are represented in the sequences. Unmethylated and hypo-methylated sequences are only found with the "A" allele whereas methylated sequences are found uniquely with the "T" allele.

It was also interesting to note that clones cl05 and cl15 had partial methylation at the CpG sites close to the promoter. This was unexpected as the sites were within the core promoter section of the sequence and methylation here might be expected to affect the expression of the gene. A possible explanation for these partially methylated sequences is considered in section 4.3 (below).

ii) *NDP* expression in human retinal samples.

Samples of retinal tissue from males and females were considered to be a good potential source of uncultured tissue that expressed the *NDP* gene. The expression of *NDP* in these tissues was confirmed by RT-PCR as described below.

Retinal tissue from male and female donors (kind gift of Drs. Sally Davies and Mike Boulton, Royal Eye Hospital Manchester) was isolated and frozen immediately at -70°C. Approximately 100mg of tissue was sliced from the frozen samples and total RNA isolated from it by an adaptation of the method of (Chomczynski and Sacchi, 1987) (see Methods 2.19). The RNA was used as substrate to synthesise first strand cDNA using random hexamer primers (see Methods 2.45). The primers NDPCDNAF (5'-TGGATTGCAAGTGCAAAGAG-3') and NDPCDNAR (5'-AGTGACAGGAGGAACGGAAG-3') were designed to amplify a 437bp fragment of the human *NDP* cDNA using the conditions shown in table 4.2 (below).

NDP cDNA PCR (467bp product) NDPCDNAF : NDPCDNAR Qiagen buffer	96°C for 2'00'' 96°C for 0'30'' 60°C for 0'30'' 72°C for 0'30'' } 35 cycles

Table 4.2 Conditions for the human *NDP* cDNA PCR reaction.

Figure 4.7 shows the results of RT-PCR experiments on a male and a female retinal tissue sample alongside control RT-PCR reactions for the ubiquitously expressed *PDHA1* gene (Chapter 3 section 3.24). The figure shows that the *NDP* gene is expressed in both the male and female retinal tissue samples.

iii) Bisulphite sequence analysis of human retinal genomic DNA.

In order to determine whether hyper-methylated promoter sequences were derived from the inactive X-chromosome, DNA from male and female retinal tissue was subjected to bisulphite sequencing as described above. By deduction the sequences from active and inactive X-chromosomes were distinguished, a method used previously for the human *PDHA1* (Chapter 3), *HPRT* (Hornstra and Yang, 1994; Park and Chapman, 1994) and *PGK-1* (Pfeifer et al., 1990) genes.

Figure 4.8 shows an ABI trace of the male retinal bisulphite clone m6809 in which none of the CpG sites are methylated. Figure 4.9 shows the ABI trace of the female retinal clone f6612 which is methylated at 6 of 13 CpG sites. An alignment of the untreated genomic sequence against 5 bisulphite sequences from male and female retinal tissues is shown in figure 4.10. The male sequences are unmethylated at all residues consistent with the idea that the unmethylated bisulphite sequences of Weri-Rb-1 cells are from the active X-chromosome. The methylated sequences found in the female retinal samples were notably less methylated than those of the inactive X-chromosome of Weri-Rb-1 cells. The higher level of methylation in the Weri-Rb-1 cells could be a result of culturing the cells through

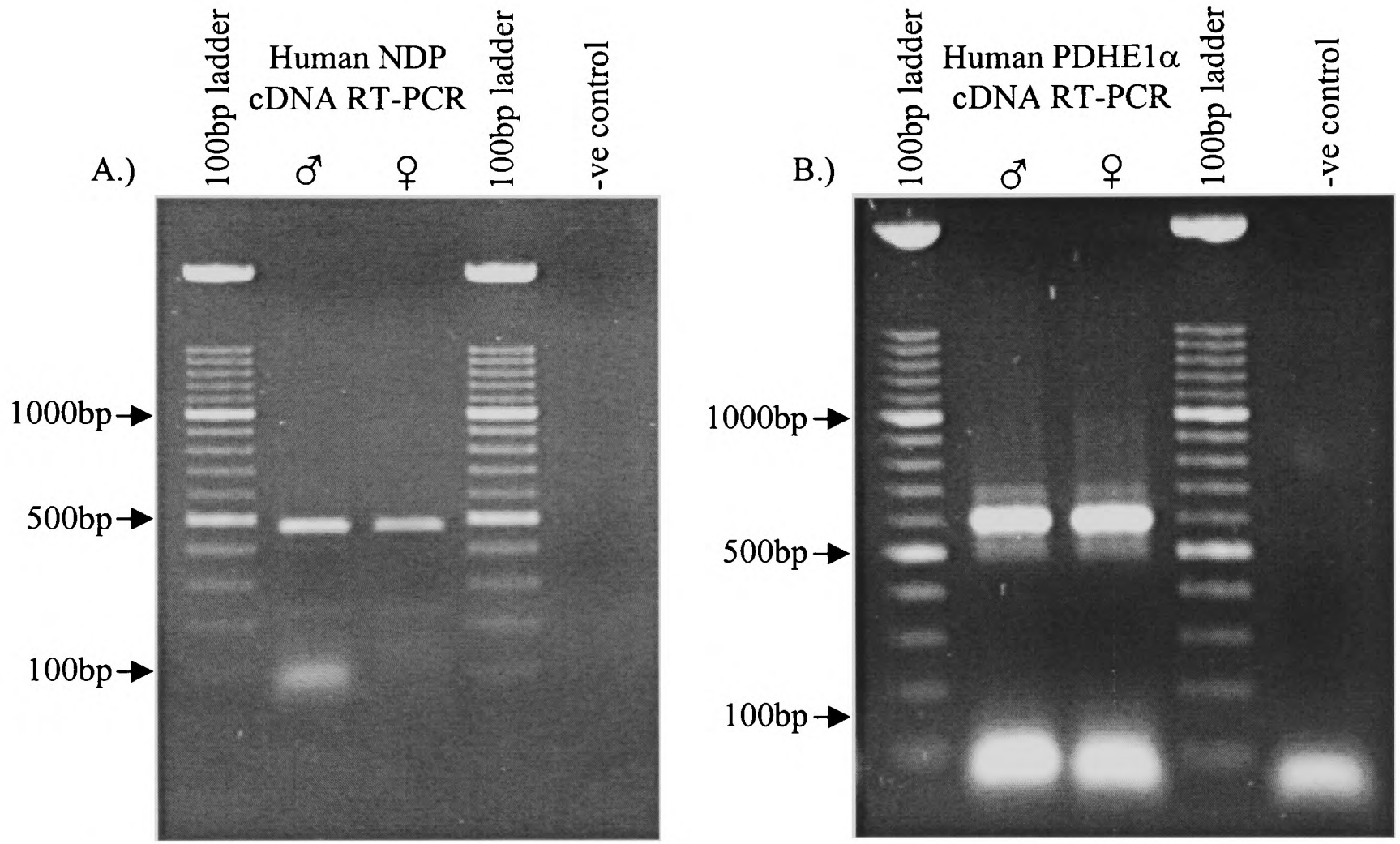


Figure 4.7: Expression of the NDP gene in male and female human retinal tissues.
Total RNA was isolated from 100mg samples of human male and female retinal tissues and used in the synthesis of first strand cDNA. PCR experiments performed on the cDNAs using the primers HNDPCDNAF and HNDPCDNAR demonstrated that the retinal tissues expressed the NDP gene (A). As a control for the quality and quantity of cDNA used in the experiments a parallel PCR reaction was performed using the primers PDH-C and PDH-D (see Chapter 3 section 3.2.4) designed to the ubiquitously expressed PDHA1 gene (B).

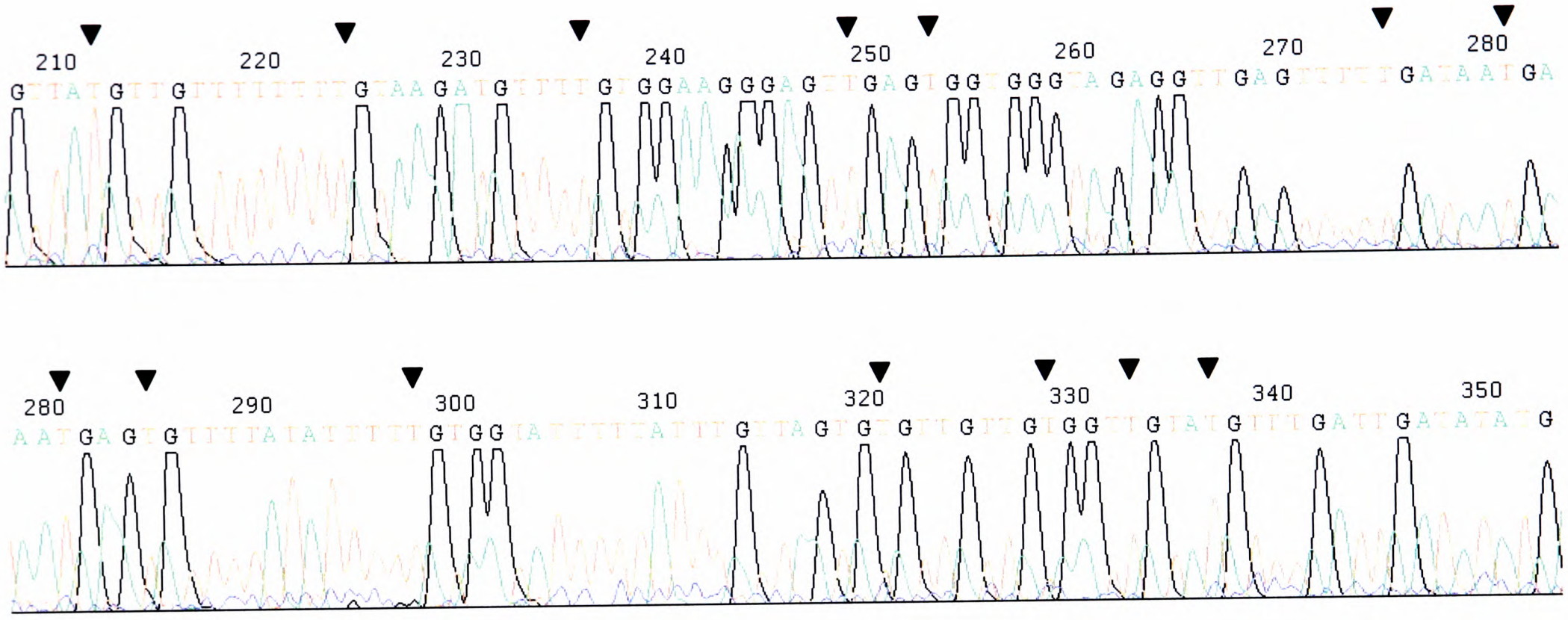


Figure 4.8: ABI sequence trace of the male retinal tissue NDP bisulphite clone m6809.
Human retinal tissue genomic DNA was used to analyse methylation at the promoter of the human NDP gene. PCR was performed on bisulphite treated DNA and the resulting PCR products were cloned into the vector pZErO2. Independent insert carrying clones were sequenced with the Sp6 and T7 pZErO2 MCS flanking primers using the ABI BigDye fluorescent cycle-sequencing kit. Sequencing reactions were purified and run on an ABI377 DNA sequencing machine. The sequence above is of the male bisulphite clone m6809 which has none of its CpG sites methylated (indicated with black triangles).

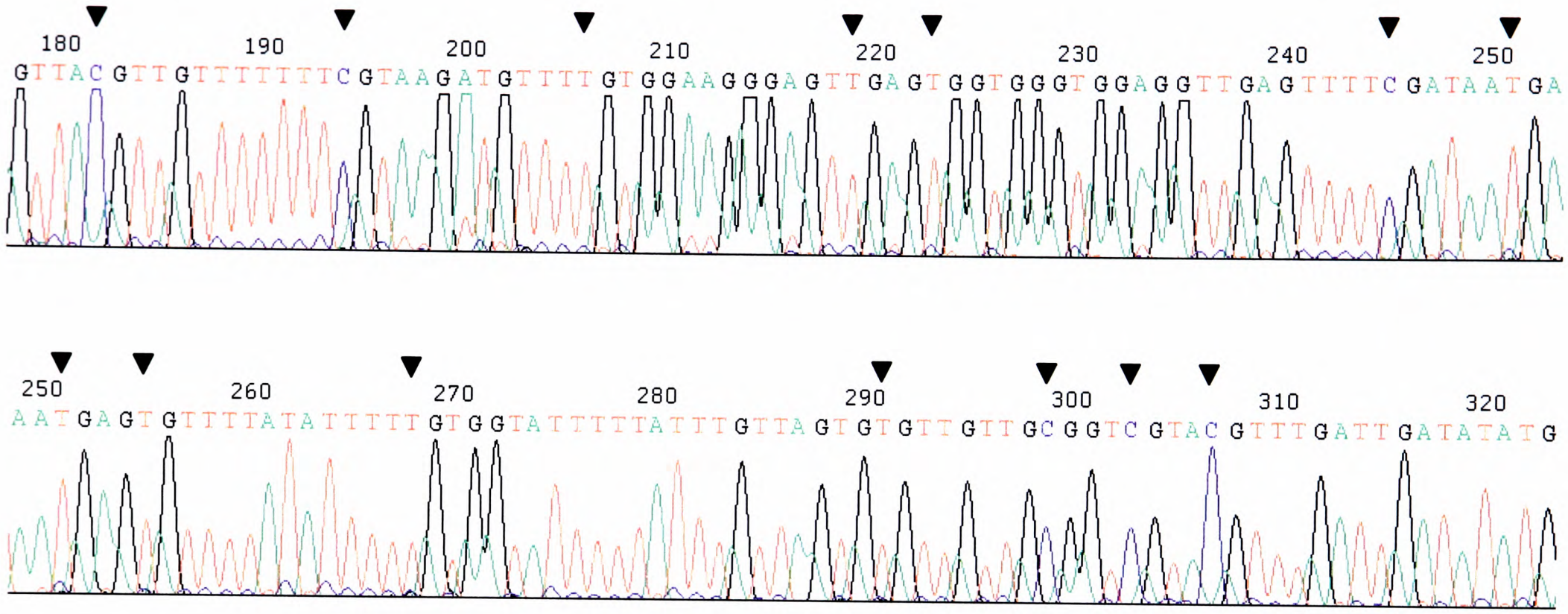


Figure 4.9: ABI sequence trace of the female retinal tissue NDP bisulphite clone f6612.
Human retinal tissue genomic DNA was used to analyse methylation at the promoter of the human NDP gene. PCR was performed on bisulphite treated DNA and the resulting PCR products were cloned into the vector pZErO2. Independent insert carrying clones were sequenced with the Sp6 and T7 pZErO2 MCS flanking primers using the ABI BigDye fluorescent cycle-sequencing kit. Sequencing reactions were purified and run on an ABI377 DNA sequencing machine. The sequence above is of the female bisulphite clone f6612 which has 6 of 13 CpG sites methylated (indicated with black triangles).

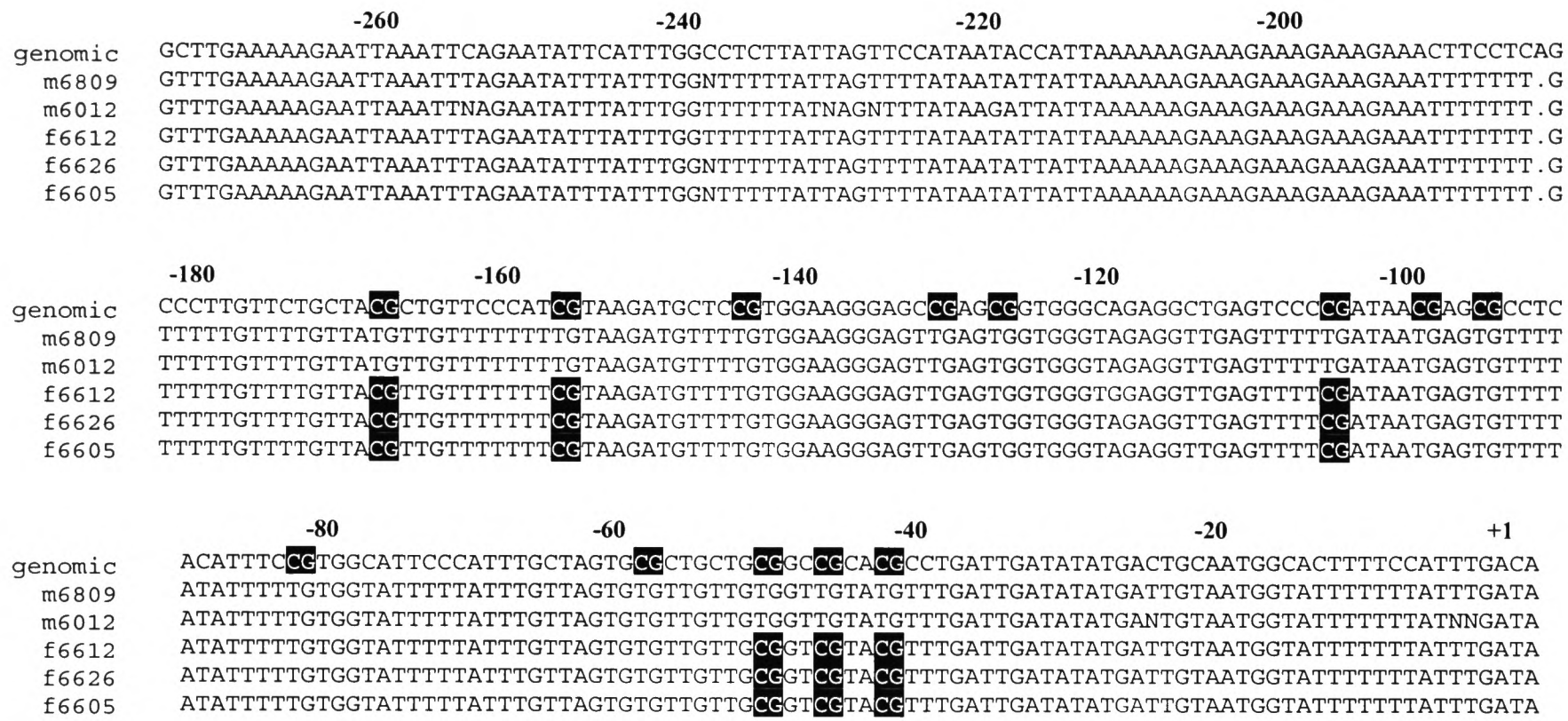


Figure 4.10: Alignment of male and female retinal tissue bisulphite sequences of the human NDP gene promoter. Bisulphite sequences from independent clones are aligned against the genomic target sequence with CpG sites highlighted in black. Male sequences are totally unmethylated whereas female sequences are methylated at 6 of the 13 CpG sites. As the NDP gene is expressed in male retinal tissue (see Figure 4.7) the bisulphite sequence results suggest that the active X-chromosome NDP gene promoter is unmethylated in expressing tissues. The female retinal bisulphite sequences may only represent the inactive X-chromosome sequences as none of the three sequences (f6612, f6626 and f6605) are totally unmethylated as would be expected of the female active X-chromosome.

many passages or might have originally occurred during the oncogenic process of transformation.

Figure 4.11 shows schematics of the different methylation patterns found on the active and inactive X-chromosomes of both Weri-Rb-1 cells and human male and female retinal tissues. This illustrates that the levels of *NDP* promoter methylation on both the active and inactive X-chromosomes of Weri-Rb-1 cells is higher than that found in human retinal tissues which express the *NDP* gene.

4.2.3 Bisulphite sequence analysis of the mouse *Ndp* gene promoter.

PCR primers were designed to analyse methylation at the murine *Ndp* gene promoter in the region homologous to the section studied in the human gene. The primers MNDPBISNF1 (5'-GTTCGTAAAAGAATTAAATTTAG-3') and MNDPBISR1 (5-ACTTCCAAATACTTTTATTCTTC-3') were designed to amplify a 395bp product of the region –273 to +122bp according to the previously described criteria (Chapter 3 section 3.2.3). Figure 4.12 shows a CpG-plot of the murine bisulphite target region which includes 10 CpG sites, no *Hpa II* and one *Hha I* site. The primers were used in PCR optimisation reactions using the 15 buffer series (see Materials 2.2 and Methods 2.44) and the optimal conditions were found to be identical to those used for the human *NDP* bisulphite PCR (see Table 4.1).

i) Bisulphite sequence analysis on genomic DNA from multiple mouse tissues.

The analysis of methylation at a murine gene promoter provides the opportunity to assess methylation in as many tissues as necessary. This is particularly useful for a tissue-specific gene for which suitable cell-lines may not be available. Tissues were dissected out from 1 male and 2 female laboratory mice from the same litter. The small number of mice used was

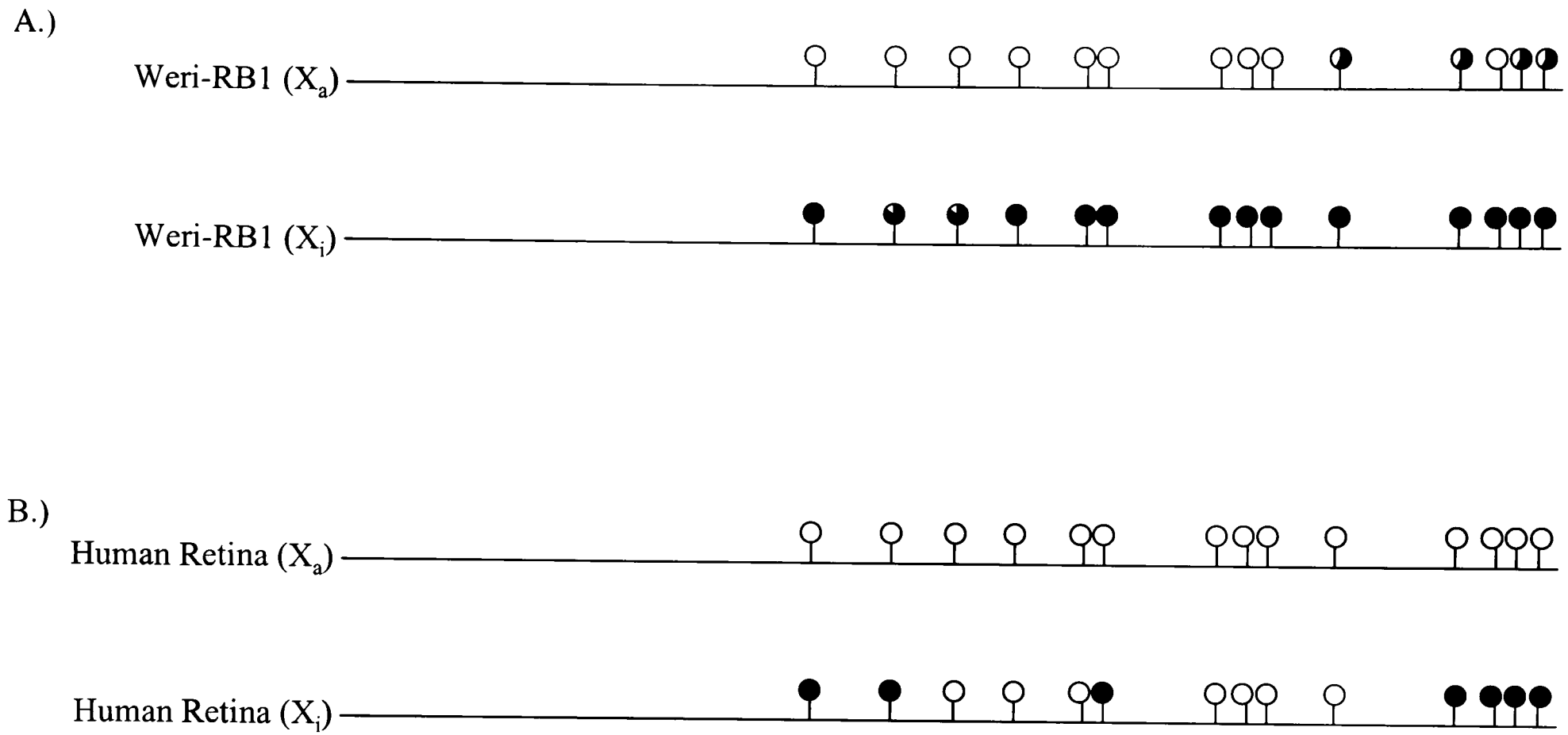


Figure 4.11: Schematics of the human Weri-Rb-1 and retinal tissue bisulphite sequences.
The figure shows the composite methylation status of CpG sites on the deduced active and inactive X-chromosomes of Weri-Rb-1cells (A) and human retinal tissues (B). The individual sites are represented by empty, filled or partially filled circles representing unmethylated, methylated and partially methylated sites. The extent of filling is equal to the fraction of sequences methylated at the site. Both of the sample types express the NDP gene and have hypo-methylated active X-chromosomes. The inactive X-chromosome sequences are hyper-methylated at the NDP promoter making a correlation between expression and methylation status of the promoter. The significance of the difference in the amount of methylation between the Weri-Rb-1 and retinal samples is discussed in the text.

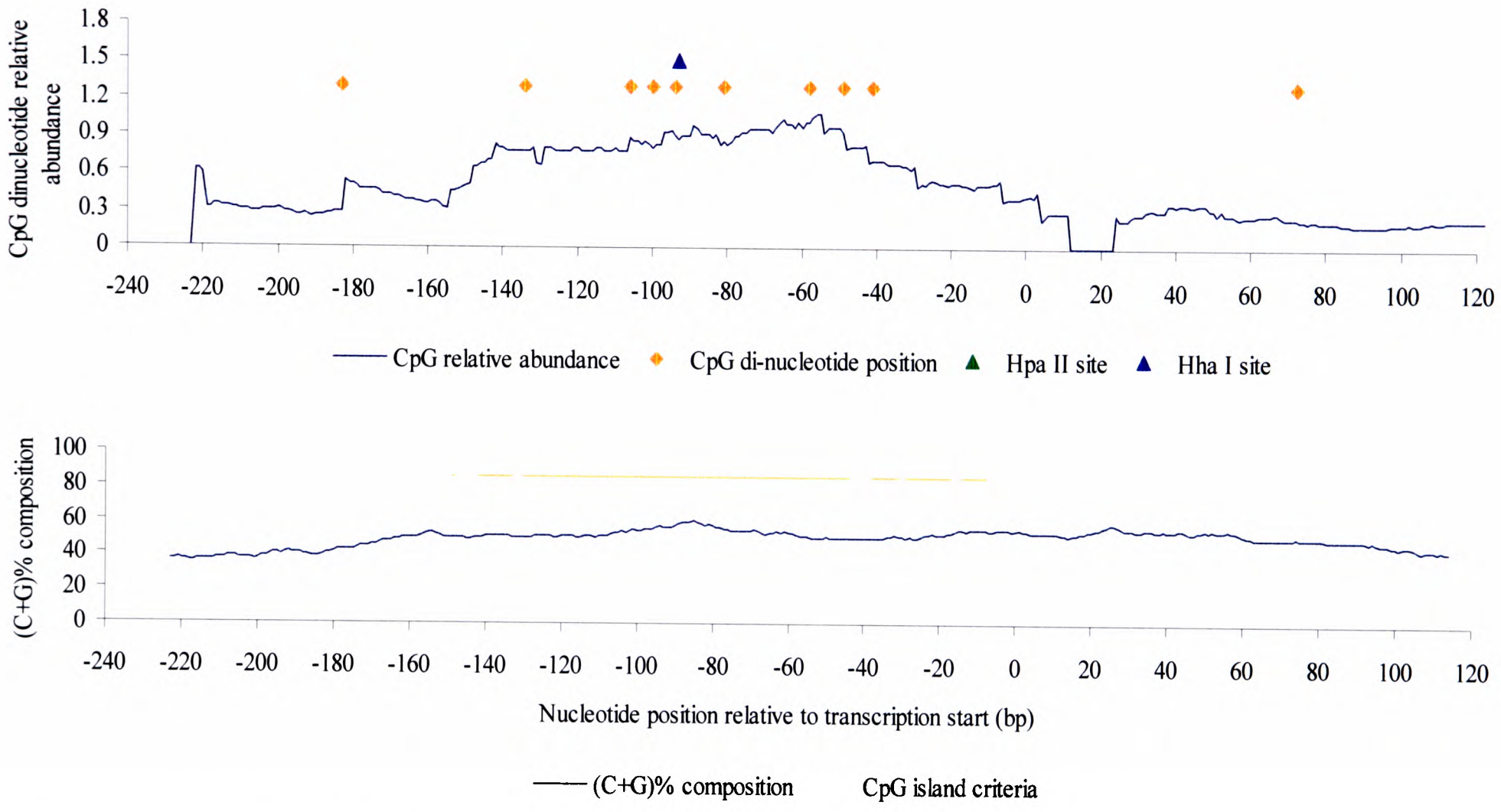
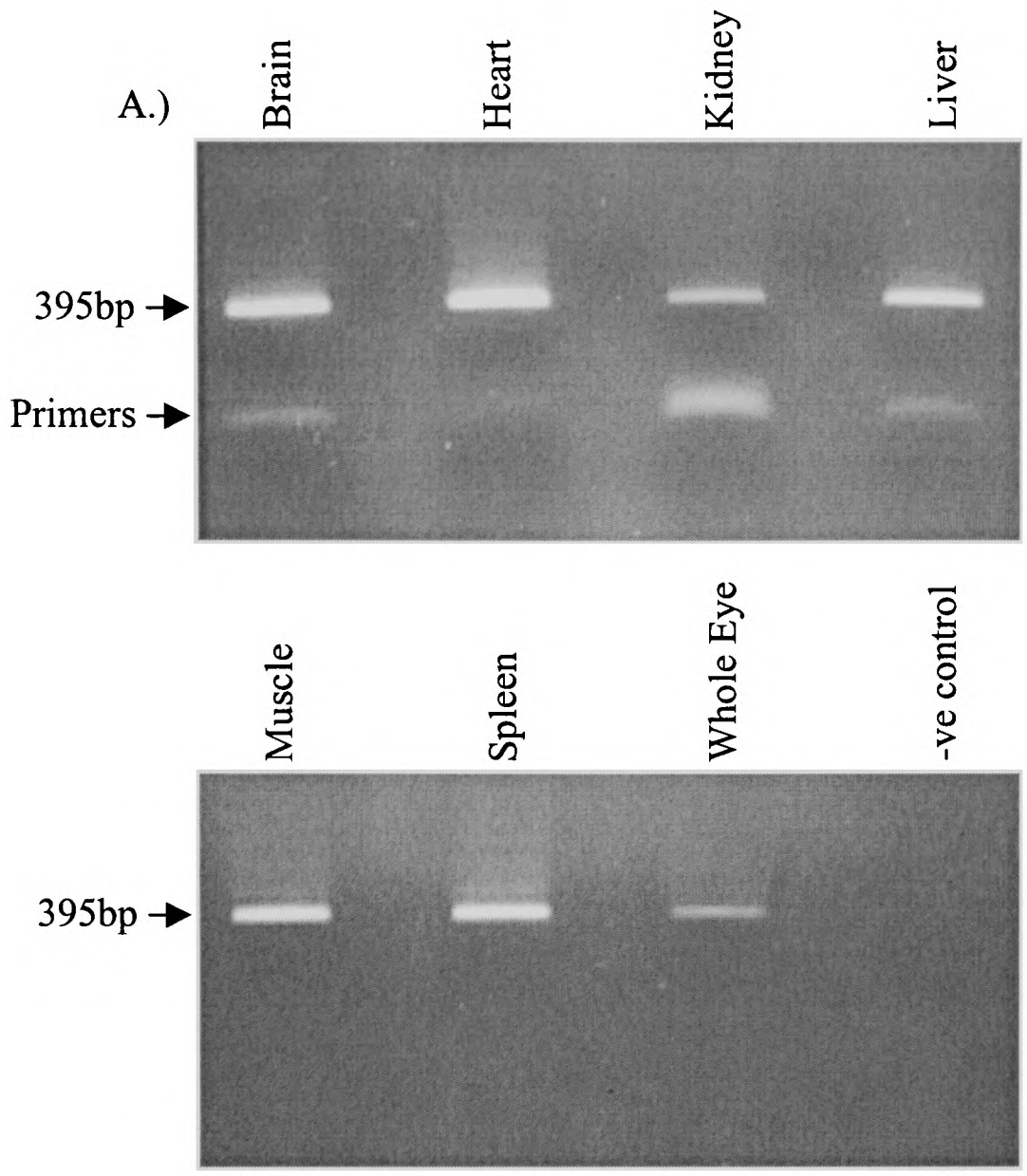


Figure 4.12: CpG-Plot of the Murine Ndp gene promoter bisulphite target region.
The region selected for bisulphite sequencing (-273 to +122bp) was analysed by the CpG-plot program. Independent analysis of data extracted from the program was used to identify windows of the sequence which adhered to the criteria for CpG abundance and (C+G)% composition required of a CpG island. When these two criteria were satisfied a point was plotted at the position of the window. This analysis was performed on the whole sequence (orange line) revealing regions of high CpG abundance. The bisulphite target region includes 10 CpG sites and 1 Hha I site and closely resembles the plot of the homologous human region of the promoter (see Figure 4.3).

primarily due to the large amount of data generated from each mouse. It would not have been possible to sequence enough clones to get representative methylation profiles from more mice within the resources available. The tissues isolated included those of the Brain, Heart, Kidney, Liver, Muscle, Spleen and Eye. Tissues were frozen immediately on dry ice and then transferred for short term storage at -70°C. Genomic DNA was isolated from approximately 100mg of tissue or one whole eye as described in (Methods 2.17). DNA was also isolated from peripheral blood as described in (Methods 2.16). DNA from each source (5μg) was subjected to bisulphite treatment and the products used as substrates in the *Ndp* bisulphite PCR reaction. PCR products (Figure 4.13A) were "polished" with *PfU* polymerase and cloned into the *Eco RV* site of pZErO-2. Ligation reactions were transformed into TOP10 strain *E.coli* by electroporation and plated on selective plates overnight. The resulting bacterial clones were picked and screened by colony PCR (Figure 4.13B) and clones containing the correct sized inserts were sequenced with Sp6 and T7 primers using the ABI BigDye™ fluorescent cycle sequencing kit (see Methods 2.42).

Examples of typical ABI sequencing traces of several bisulphite sequences from independent clones are shown for Heart (clone 4f2), Kidney (clone 9t7), Liver (clone 3f2) and peripheral blood (clone L129-4) in figures 4.14-17 respectively. The alignments of bisulphite sequences for female-Brain, female-Heart, male-Heart, female-Kidney, male-Kidney, female-Liver, male/female-Eye, female-peripheral blood, female-Muscle and female-Spleen against untreated genomic target are shown in figures 4.18-26 respectively.

In all of the bisulphite sequences, there are no examples of totally unmethylated clones in contrast to those found in both human Weri-Rb-1 and human retinal samples. Importantly the methylation analysis of the only male tissues to be studied namely those of the heart, kidney



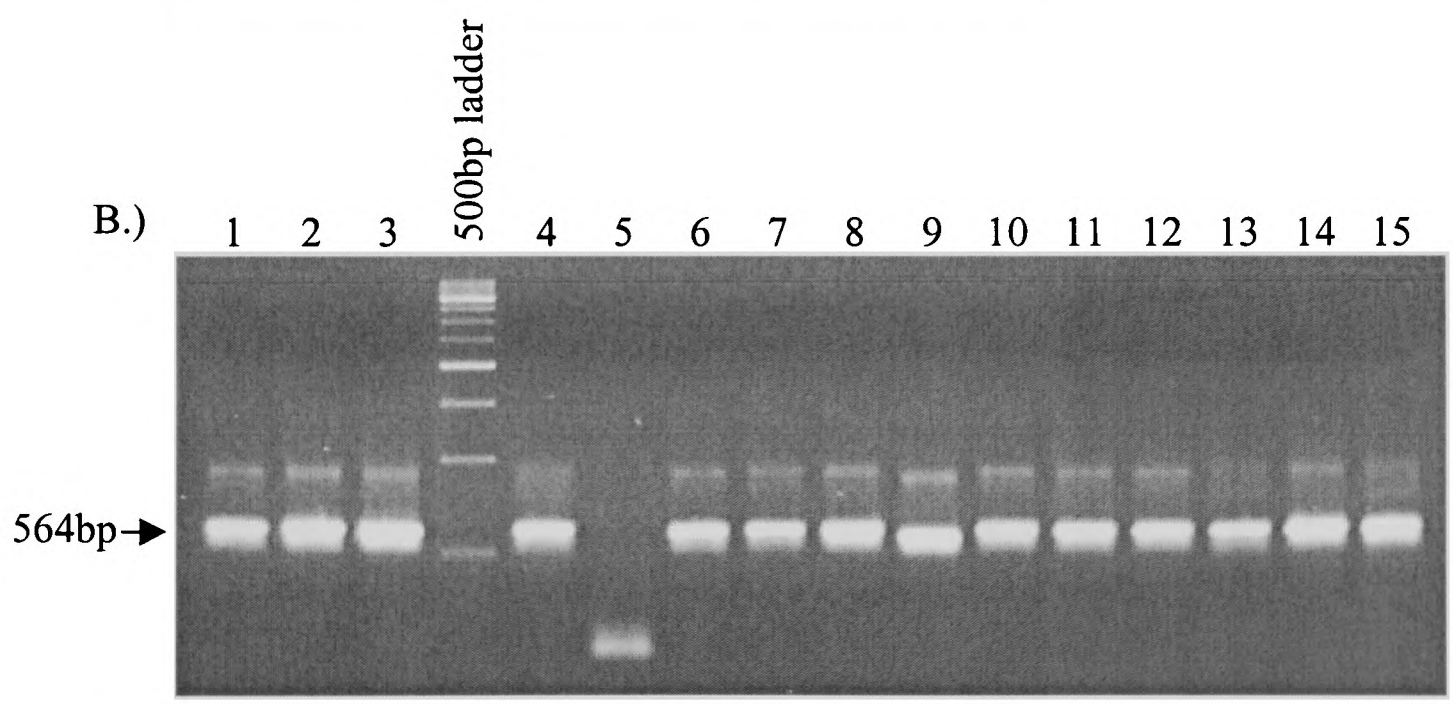


Figure 4.13: Murine tissue bisulphite PCR and colony screening
Multiple murine tissues were isolated and genomic DNA extracted from them for methylation analysis of the murine Ndp gene promoter. PCR reactions on bisulphite treated DNA with the primers MNDPBISNF1 and MNDPBISR1 amplified a 395bp product in all samples. The PCR products were loaded onto a 1.0% (w/v) agarose gel and separated from primers by electrophoresis (A). Product fragments were purified from the gel and blunt-cloned into the *Eco RV* site of the vector pZErO2 (see Methods 2.39.2 and 2.39.3). Insert-containing bacterial clones were identified by colony PCR using the pZErO2 MCS flanking primers Sp6 and T7 (see Methods 2.40)(B).

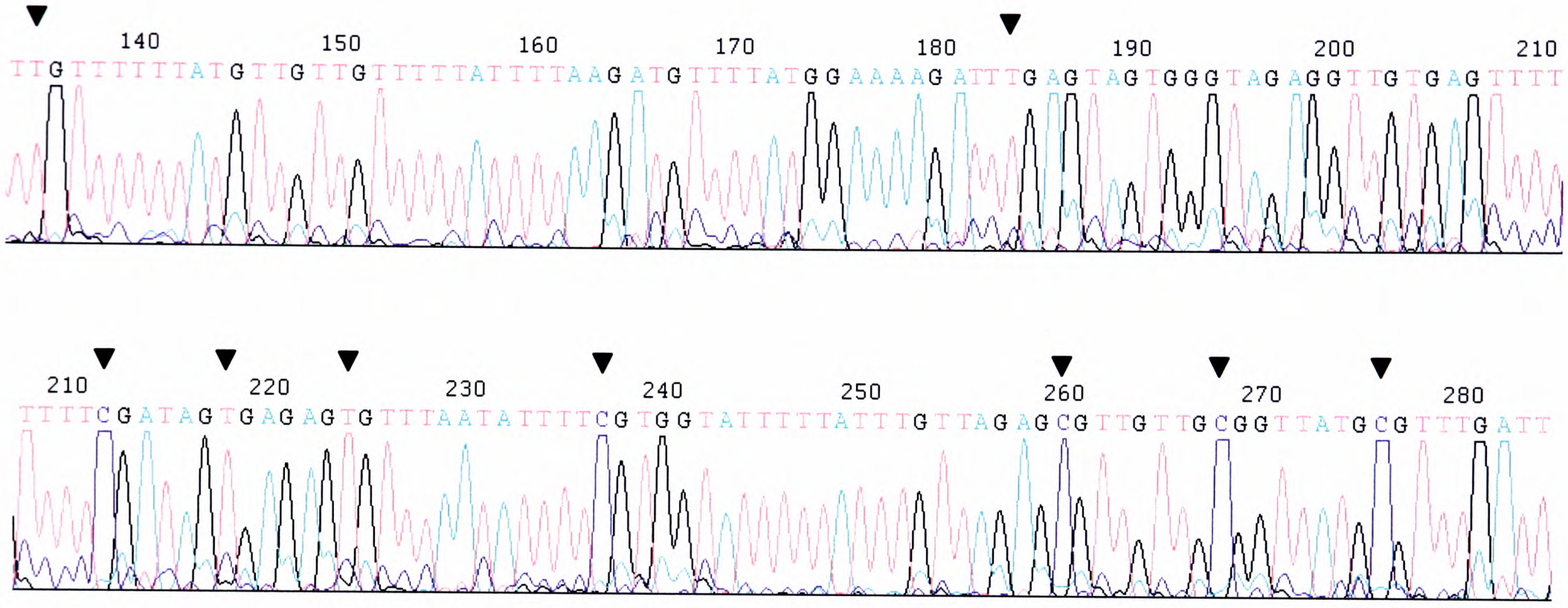


Figure 4.14: ABI sequence trace of the female heart bisulphite clone 4f2
Murine heart tissue genomic DNA was used to analyse methylation at the promoter of the murine Ndp gene promoter. PCR was performed on bisulphite treated DNA and the resulting PCR products were cloned into the vector pZErO2. Independent insert carrying clones were sequenced with the Sp6 and T7 pZErO2 MCS flanking primers using the ABI BigDye fluorescent cycle-sequencing kit. Sequencing reactions were purified and run on an ABI377 DNA sequencing machine. The sequence above is of the clone 4f2 which has 5 of 9 CpG sites methylated (indicated by black triangles).

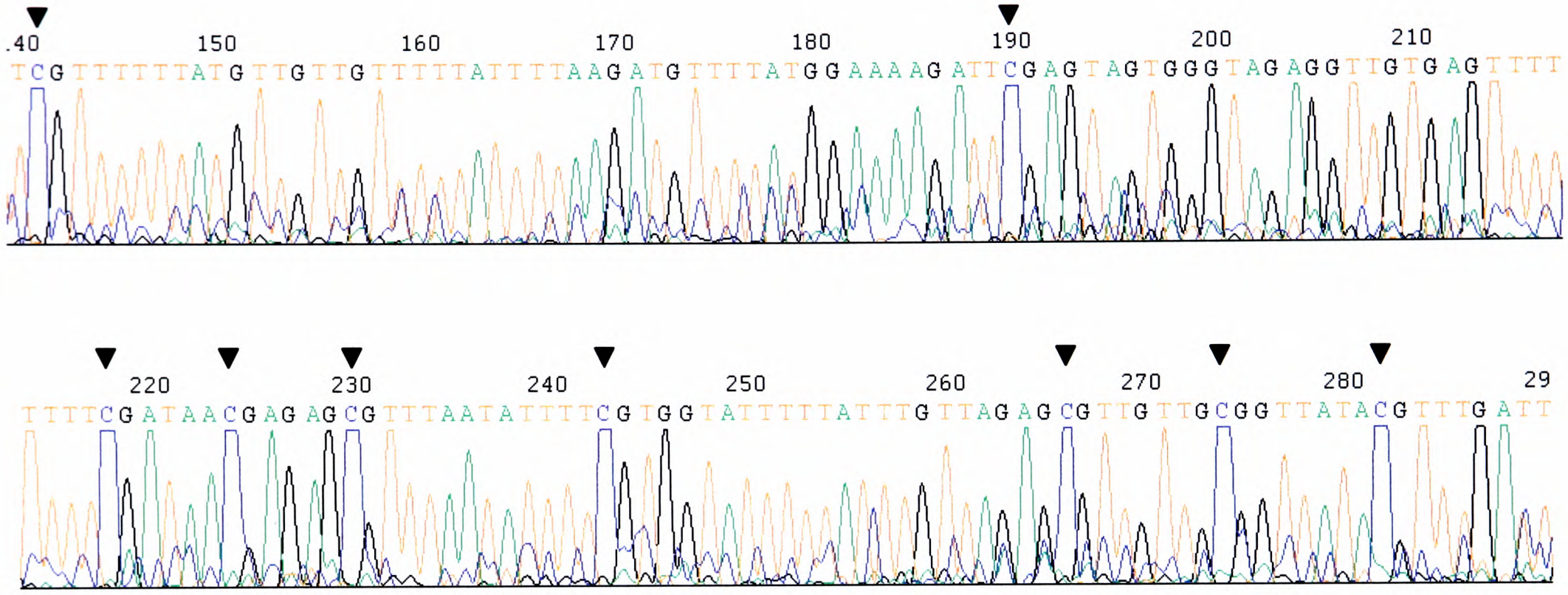


Figure 4.15: ABI sequence trace of the female kidney bisulphite clone 9t7
Murine kidney tissue genomic DNA was used to analyse methylation at the promoter of the murine Ndp gene promoter. PCR was performed on bisulphite treated DNA and the resulting PCR products were cloned into the vector pZErO2. Independent insert carrying clones were sequenced with the Sp6 and T7 pZErO2 MCS flanking primers using the ABI BigDye fluorescent cycle-sequencing kit. Sequencing reactions were purified and run on an ABI377 DNA sequencing machine. The sequence above is of the clone 9t7 which has all CpG sites methylated (indicated by black triangles).

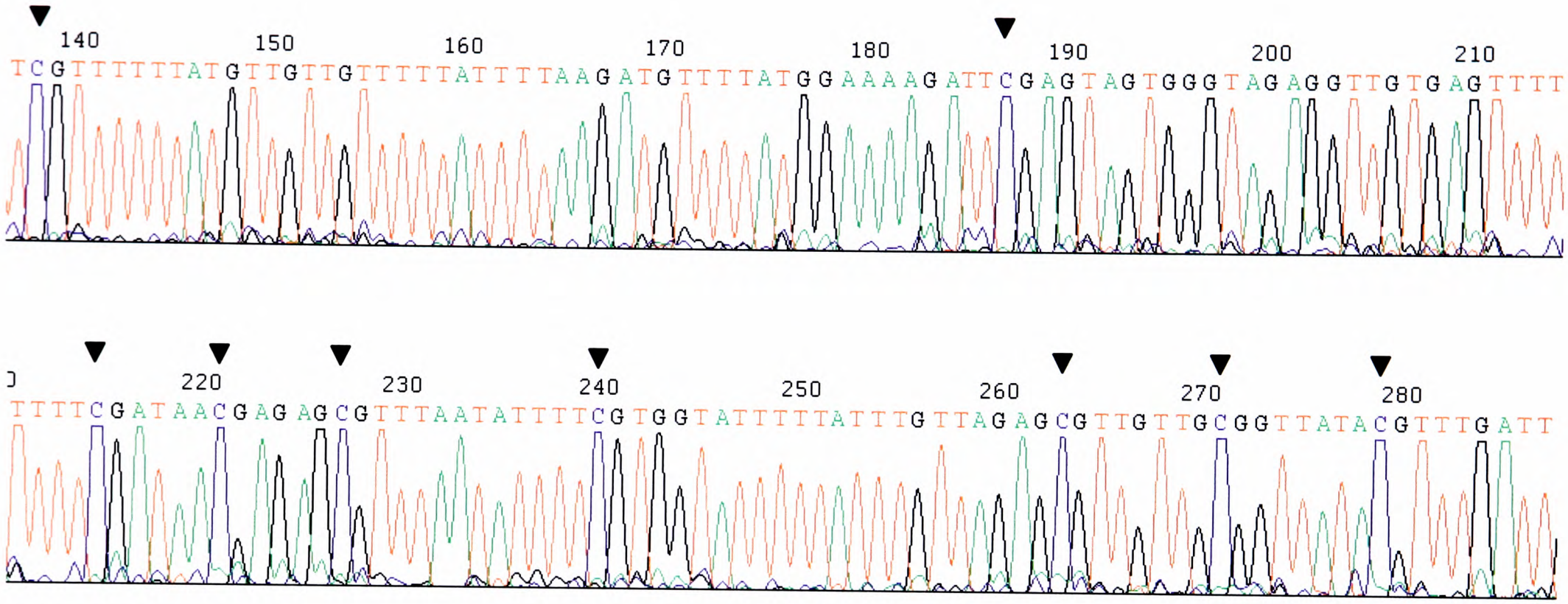


Figure 4.16: ABI sequence trace of the female liver bisulphite clone 3f2
Murine liver tissue genomic DNA was used to analyse methylation at the promoter of the murine Ndp gene promoter. PCR was performed on bisulphite treated DNA and the resulting PCR products were cloned into the vector pZErO2. Independent insert carrying clones were sequenced with the Sp6 and T7 pZErO2 MCS flanking primers using the ABI BigDye fluorescent cycle-sequencing kit. Sequencing reactions were purified and run on an ABI377 DNA sequencing machine. The sequence above is of the clone 3f2 which has all CpG sites methylated (indicated by black triangles).

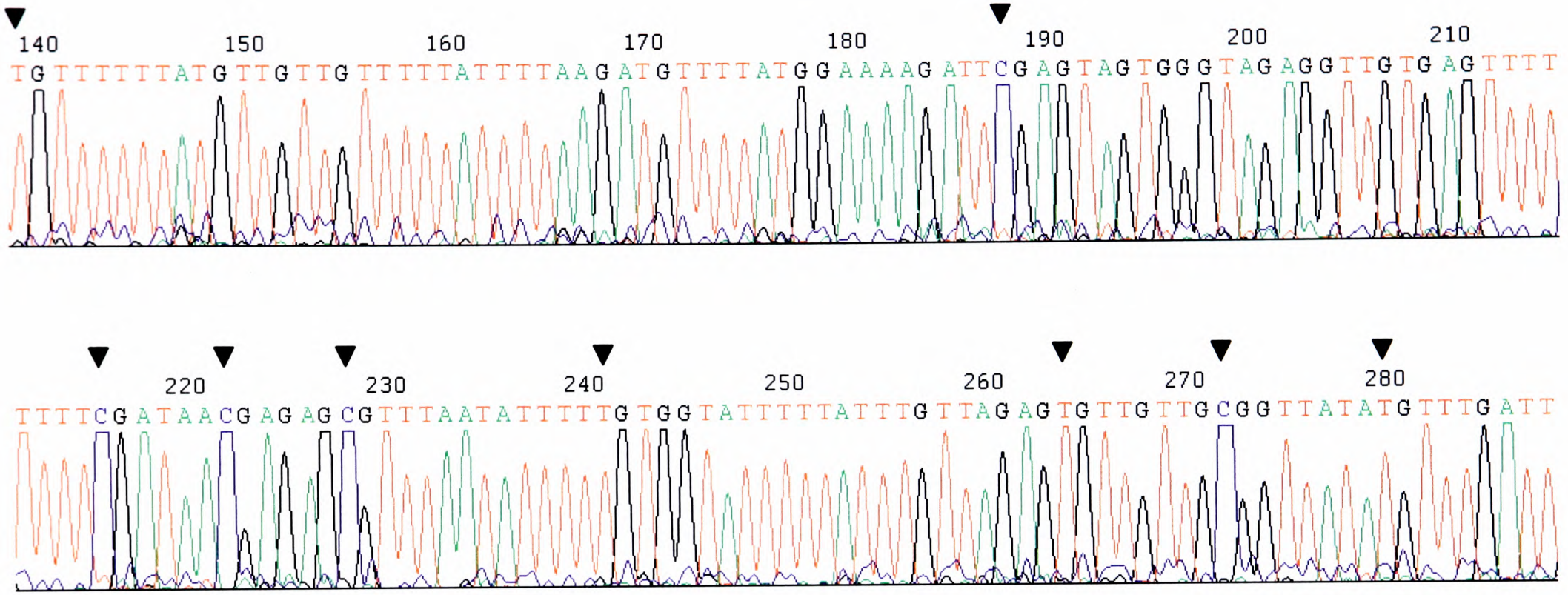


Figure 4.17: ABI sequence trace of the female peripheral blood bisulphite clone l129-4
Murine peripheral blood genomic DNA was used to analyse methylation at the promoter of the murine Ndp gene promoter. PCR was performed on bisulphite treated DNA and the resulting PCR products were cloned into the vector pZErO2. Independent insert carrying clones were sequenced with the Sp6 and T7 pZErO2 MCS flanking primers using the ABI BigDye fluorescent cycle-sequencing kit. Sequencing reactions were purified and run on an ABI377 DNA sequencing machine. The sequence above is of the clone l129-4 which has 5 of 9 CpG sites methylated (indicated by black triangles).

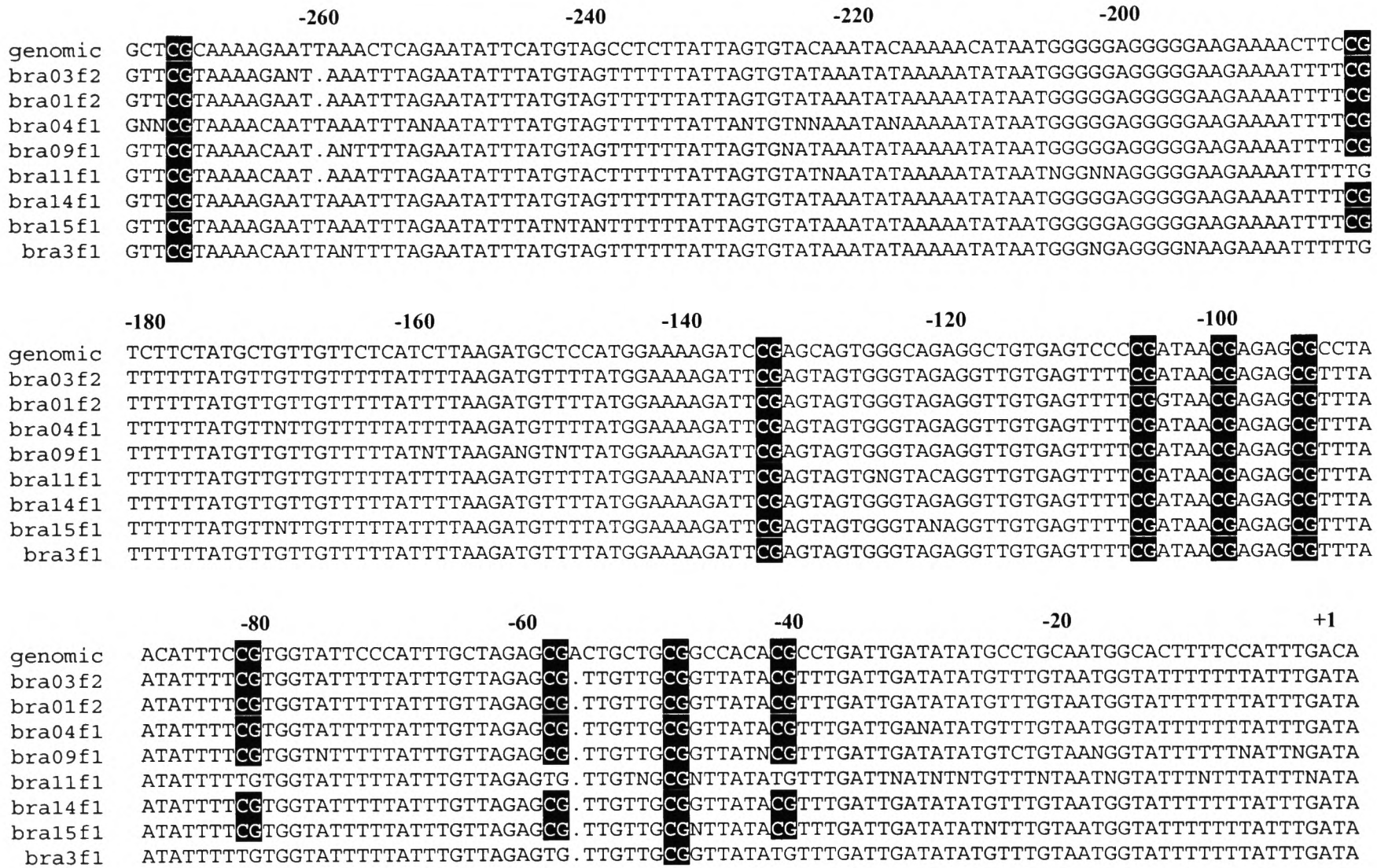


Figure 4.18: Alignment of bisulphite sequences from the genomic DNA of female-brain tissue.
Bisulphite sequences from 8 independent clones are aligned against the genomic target sequence with CpG sites highlighted in black. The first CpG site is part of the PCR primer and is highlighted for reference only. All sequences are heavily methylated with a small number of sequences containing unmethylated sites concentrated in the last 4 CpG sites closest to the transcription initiation site.

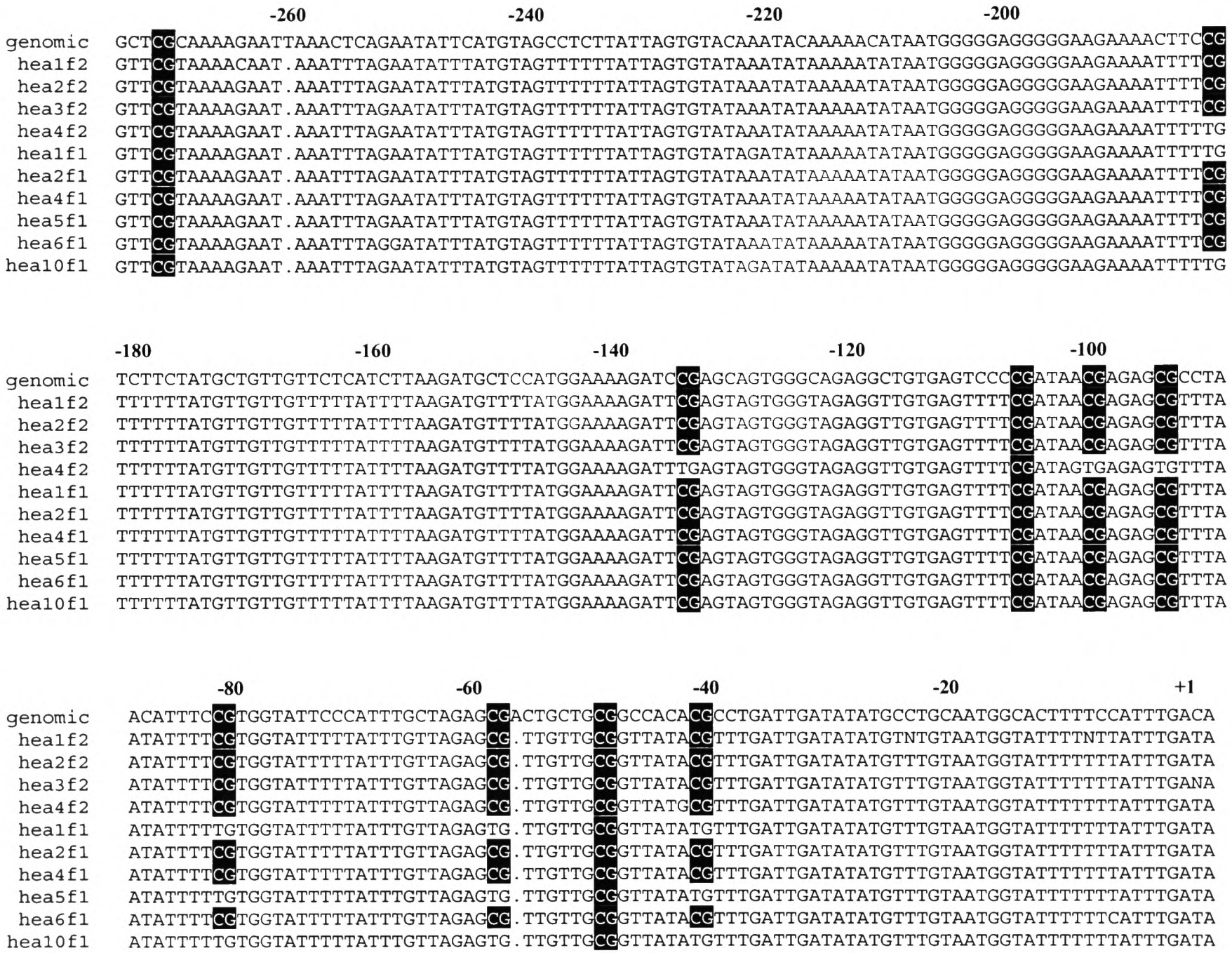


Figure 4.19: Alignment of bisulphite sequences from the genomic DNA of female heart tissue.
The female and male heart tissue bisulphite sequence alignments are explained together in the legend of the male sequence alignment (see Figure 4.20).

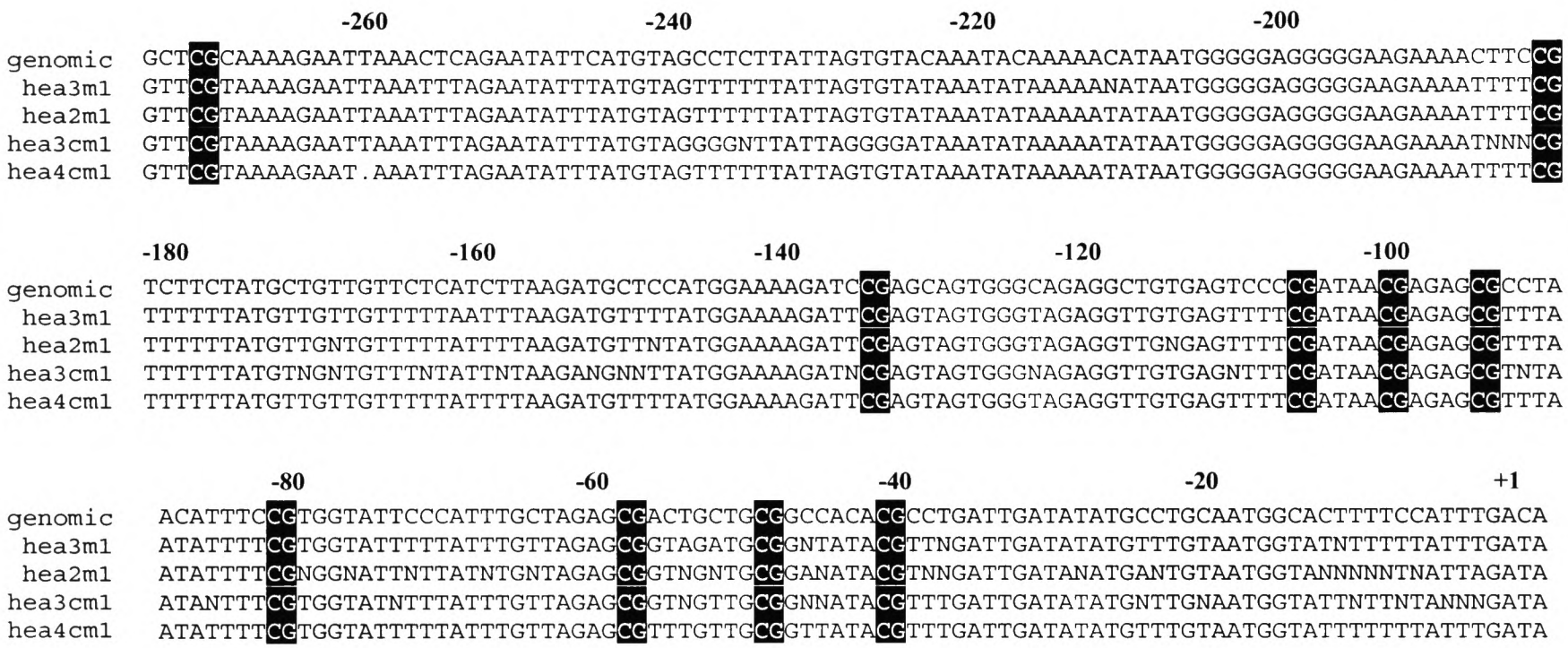


Figure 4.20: Alignment of bisulphite sequences from the genomic DNA of male heart tissue.
Bisulphite sequences from 10 independent female clones (see Figure 4.19) and 4 male clones are aligned against the genomic target sequence with CpG sites highlighted in black. The first CpG site is part of the PCR primer and is highlighted for reference only. All male and all but 3 female sequences are methylated at all CpG sites. Three female sequences (1f1, 5f1 and 10f1) have only 5 of 9 CpG sites methylated with the unmethylated sites concentrated in the last 4 CpG sites closest to the transcription initiation site..

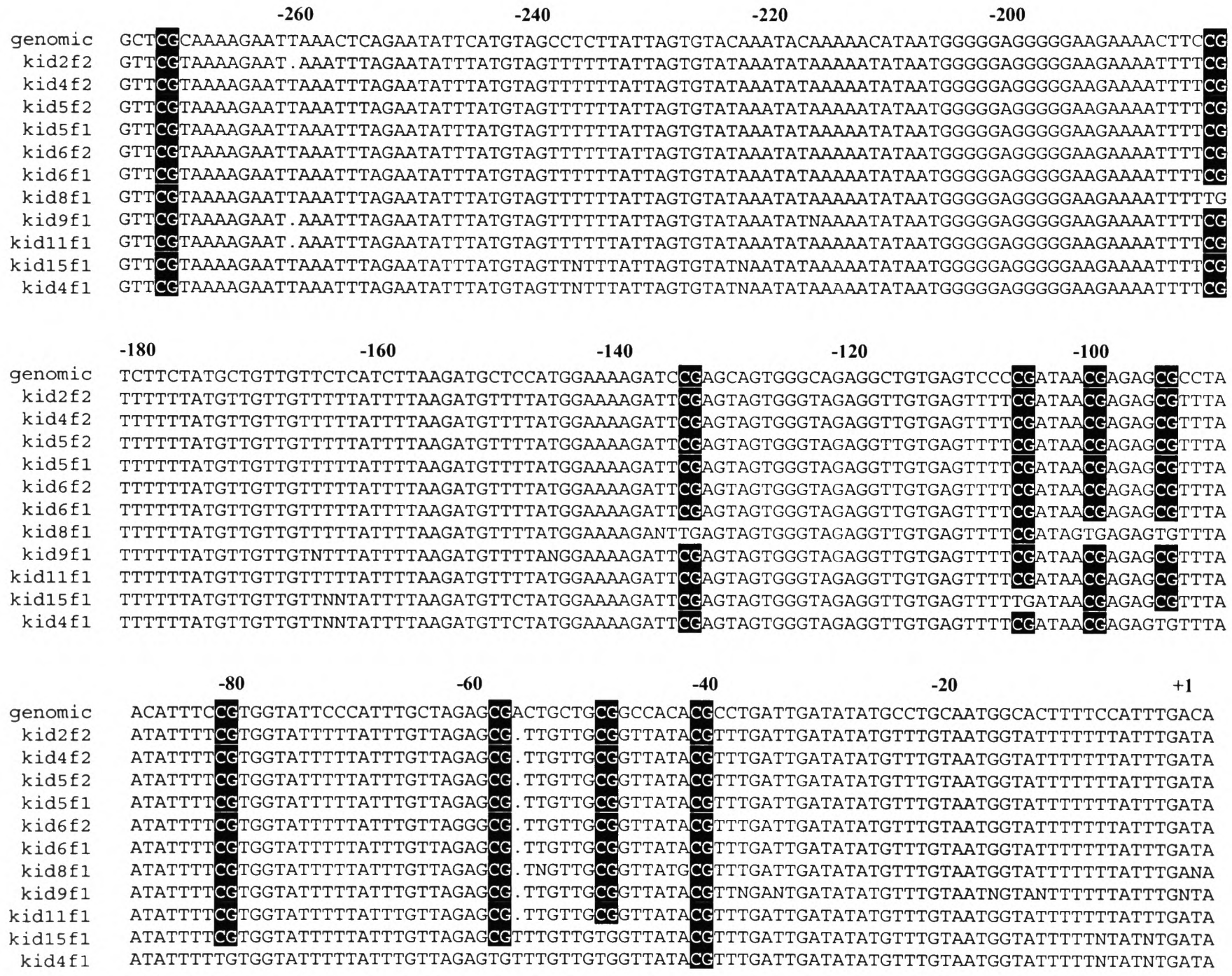


Figure 4.21: Alignment of bisulphite sequences from the genomic DNA of female kidney tissue.
The female and male kidney tissue bisulphite sequence alignments are explained together in the legend of the male sequence alignment (see Figure 4.22).

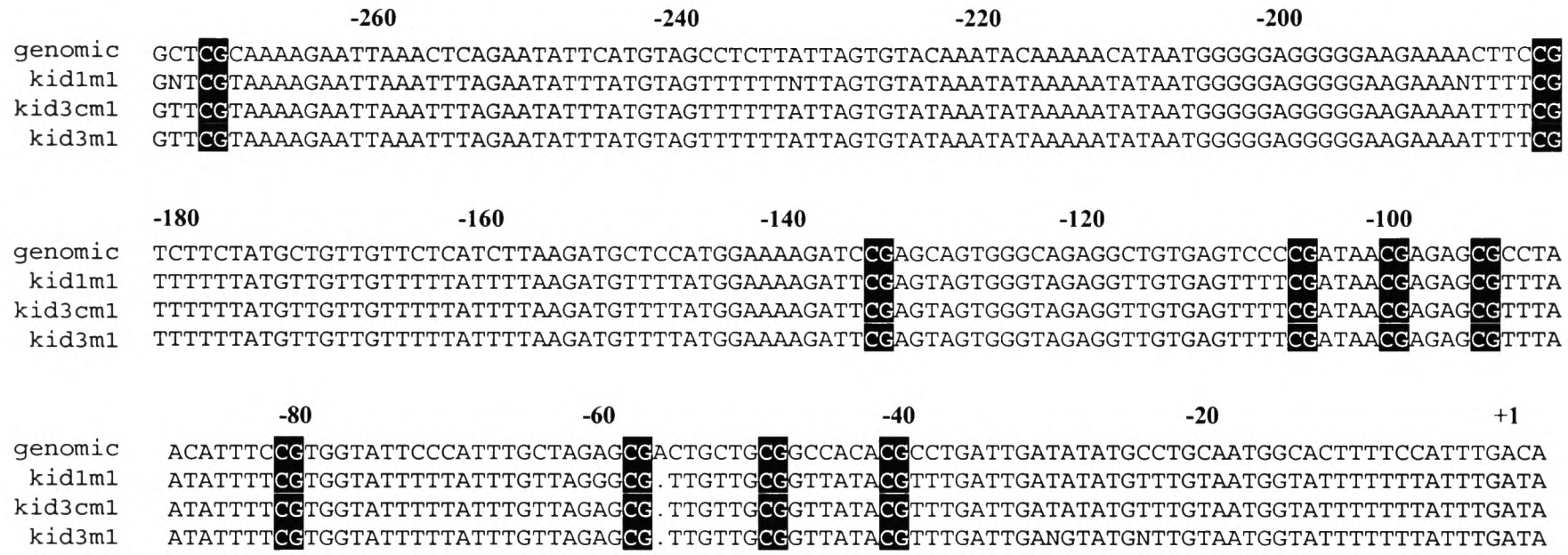


Figure 4.22: Alignment of bisulphite sequences from the genomic DNA of male kidney tissue. Bisulphite sequences from 11 independent female clones (see Figure 4.21) and 3 male clones are aligned against the genomic target sequence with CpG sites highlighted in black. The first CpG site is part of the PCR primer and is highlighted for reference only. All male and all but 3 female sequences are methylated at all CpG sites. Three female sequences (8f1, 15f1 and 4f1) are methylated at only 5, 6 and 5 CpG sites respectively. The unmethylated sites are in the last 4 CpG sites closest to the transcription start site and at CpG sites 3, 4, 5 and 6.

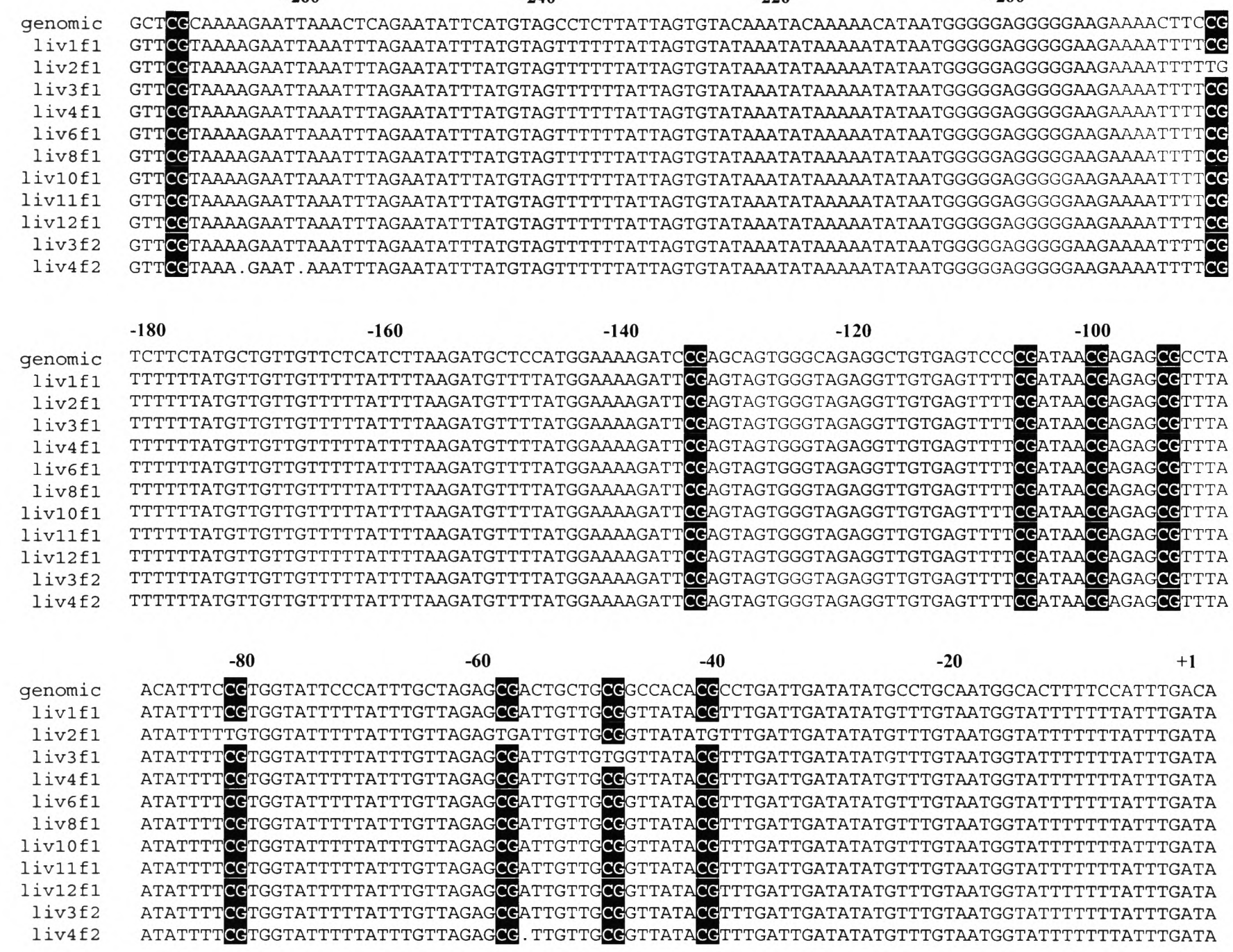


Figure 4.23: Alignment of bisulphite sequences from the genomic DNA of female liver tissue.
Bisulphite sequences from 11 independent female clones are aligned against the genomic target sequence with CpG sites highlighted in black. The first CpG site is part of the PCR primer and is highlighted for reference only. All but two sequences are methylated at all CpG sites. The sequences of clones 2f1 and 3f1 exhibit partial methylation concentrated in the last 4 CpG sites closest to the transcription initiation site.

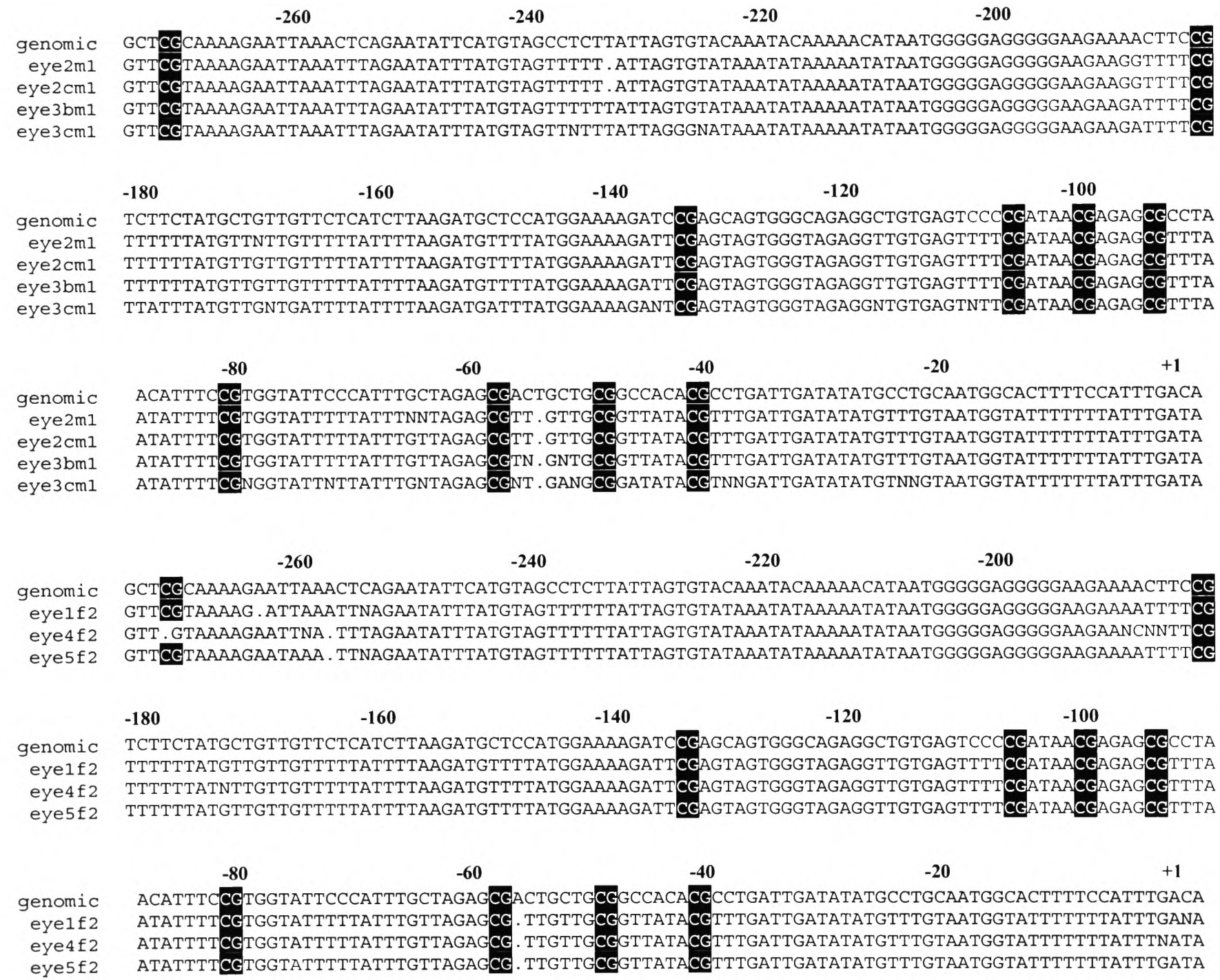


Figure 4.24: Alignment of bisulphite sequences from the genomic DNA of male and female whole eye.
Bisulphite sequences from 4 male and 3 female independent clones are aligned against the genomic target sequence with CpG sites highlighted in black. The first CpG site is part of the PCR primer and is highlighted for reference only. All male and female sequences are methylated at all CpG sites.

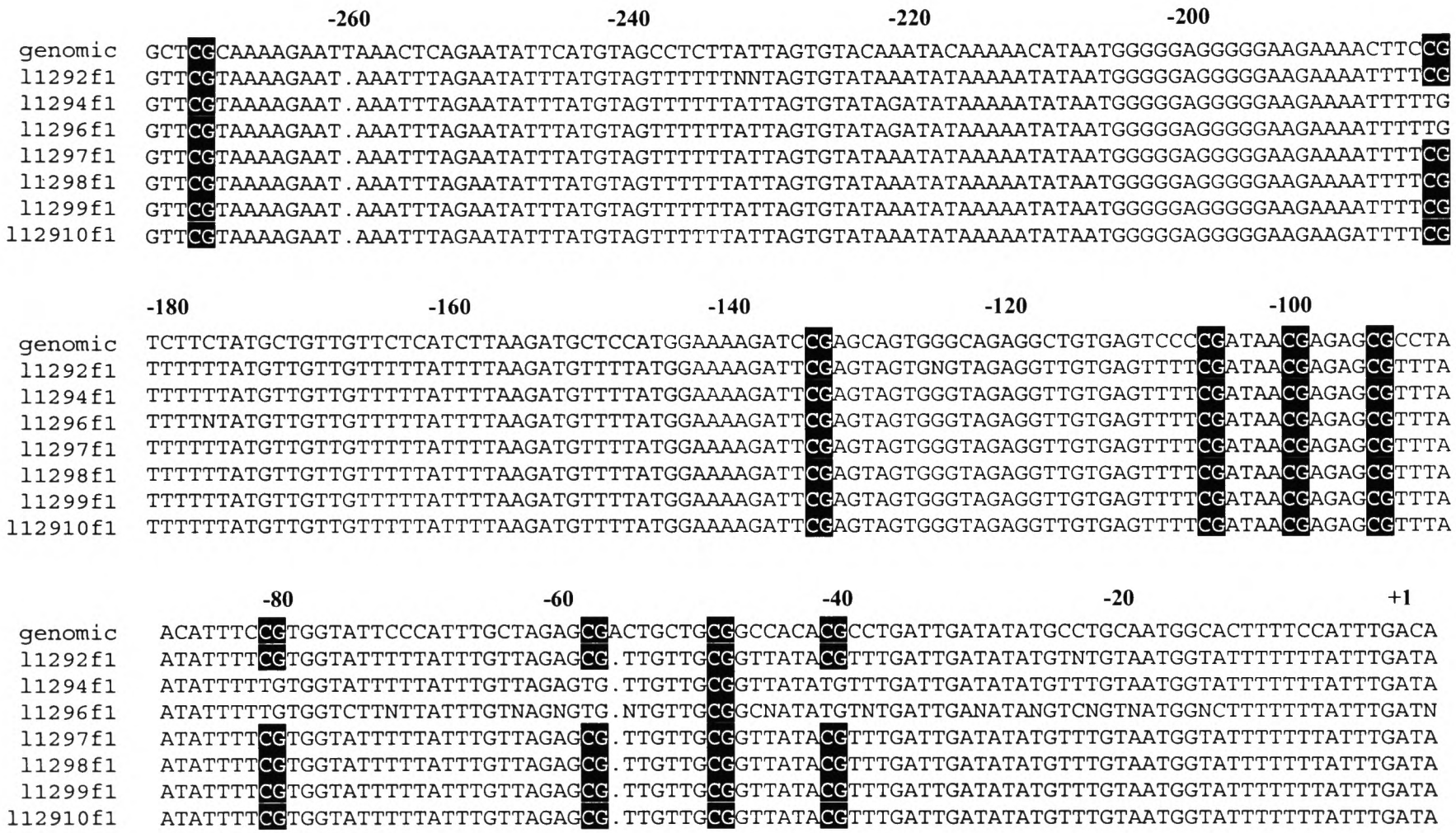


Figure 4.25: Alignment of bisulphite sequences from the genomic DNA of female peripheral blood. Bisulphite sequences from 7 independent female clones are aligned against the genomic target sequence with CpG sites highlighted in black. The first CpG site is part of the PCR primer and is highlighted for reference only. All but 2 sequences are methylated at all CpG sites. The clones 4f1 and 6f1 are methylated at only 5 of 9 CpG sites with the unmethylated sites concentrated in the last 4 CpG sites closest to the transcription initiation site.

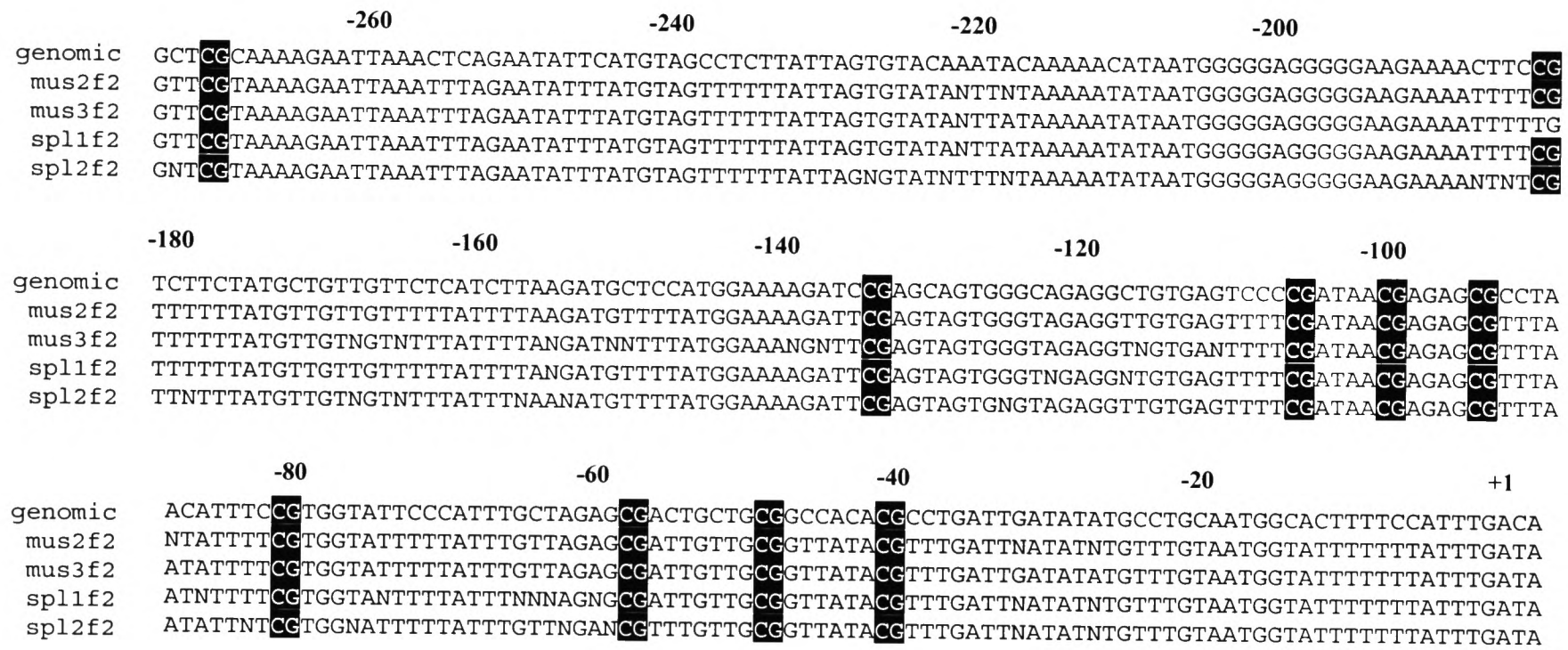


Figure 4.26: Alignment of bisulphite sequences from the genomic DNA of female muscle and spleen tissues. Bisulphite sequences from 4 independent female clones are aligned against the genomic target sequence with CpG sites highlighted in black. The first CpG site is part of the PCR primer and is highlighted for reference only. All but one of the sequences are methylated at all CpG sites.

and eye revealed no unmethylated sequences in a total of 11 sequenced clones. This suggests that the active X-chromosome is methylated in these male tissues. The absence of totally unmethylated sequences from female tissue would further imply that both active and inactive X-chromosomes are methylated in the female tissues studied. In total, the sequencing of 54 independent clones from female tissues of the brain, heart, kidney, liver, muscle, spleen, eye and peripheral blood identified no totally unmethylated clones.

There are, however, variable levels of methylation present in many of the sequences. The schematic of figure 4.27 shows the overall methylation patterns of the different tissues and positions them on a schematic of the promoter. It was important to determine whether sequences with unmethylated CpG sites were restricted to tissues that express the *Ndp* gene. To correlate the expression status of the gene to methylation profiles, RT-PCR was performed on total RNA isolated from the tissues to establish which of them expressed the *Ndp* gene. Total RNA was isolated from each tissue as previously described for human retinal tissue (see section 4.2.2(ii)) and this was used as substrate in first strand cDNA synthesis with random hexamer primers (see Methods 2.45). PCR reactions were performed using the primers MNDPCDNAF1 (5'-TTGGAGGAGTTTCCCTTTACAAC-3') and MNDPCDNAR1 (5'-GAAATTCTTCCAGGGAAAGGAG-3') which were designed to amplify a 629bp fragment of the murine *Ndp* cDNA from the tissue first strand cDNAs. Optimised conditions for the PCR are shown in table 4.3 (below).

Ndp cDNA PCR (629bp product) MNDPCDNAF1 : MNDPCDNAR1 Qiagen buffer	96°C for 2'00'' 96°C for 0'30'' 58°C for 0'30'' 72°C for 0'30'' } 35 cycles

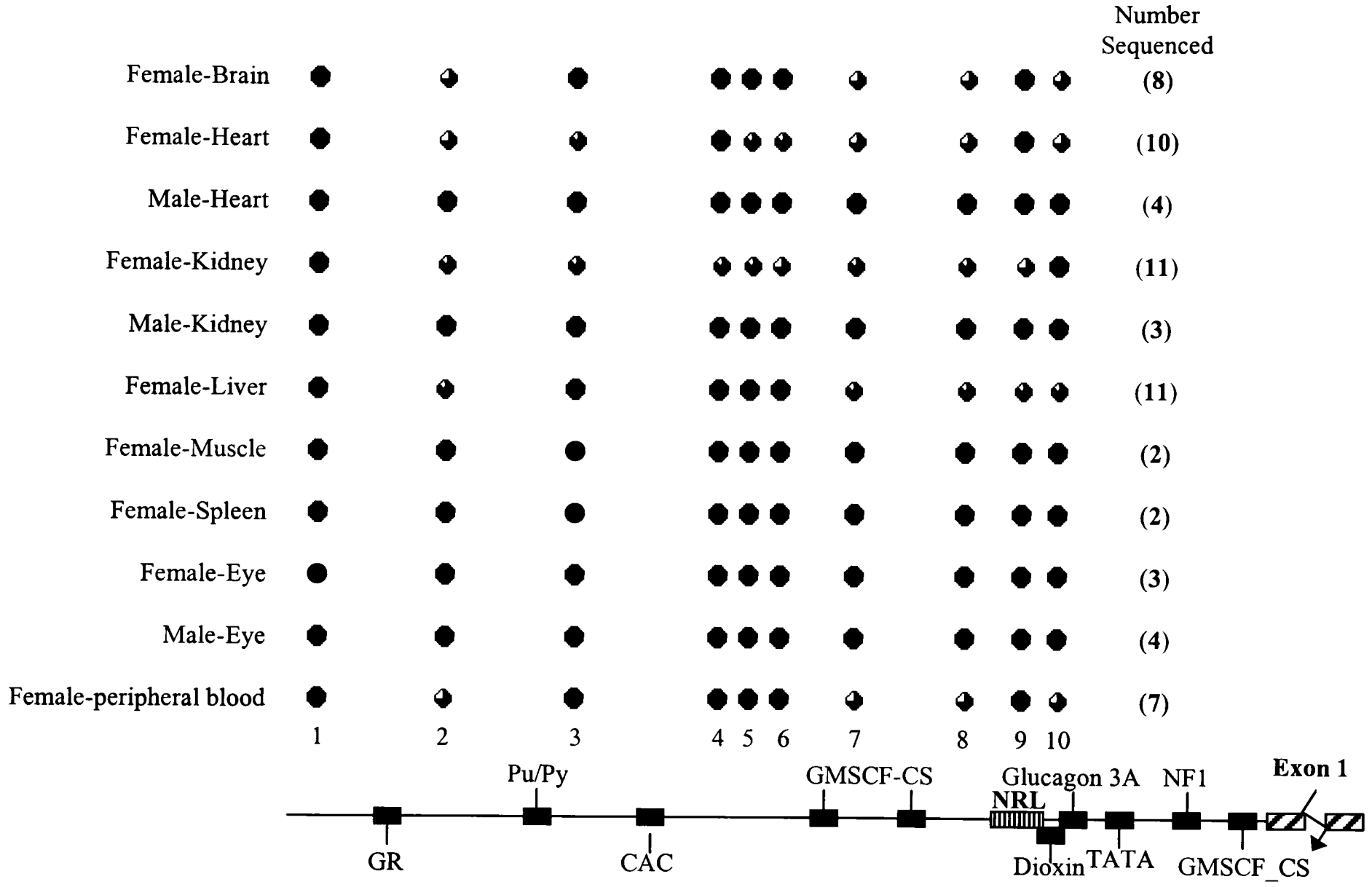


Figure 4.27: Schematic of mouse tissue bisulphite sequence profiles
The composite methylation status of each CpG site derived from bisulphite sequences is shown for each tissue. The individual sites are represented by empty, filled or partially filled circles representing unmethylated, methylated and partially methylated sites. The extent of filling is equal to the fraction of sequences methylated at the site. The first CpG is from the forward primer itself and is shown for reference only. All male sequences are methylated at all CpG sites. Female sequences exhibit variable, but high levels of methylation throughout. A schematic of the Ndp gene promoter showing the positions of transcription factor sites is aligned for ease of reference.

Table 4.3 Conditions for the murine *Ndp* cDNA PCR reaction.

Figure 4.28 shows RT-PCR results on mouse tissues of the brain, heart, kidney, liver, muscle, spleen and eye along with a negative control. The figure also includes a control series of RT-PCR reactions for the expression of the ubiquitously expressed *Pdha1* gene as detailed in chapter 3 (section 3.2.4). Expression of the *Ndp* gene can be seen in whole brain, heart, kidney and whole eye tissues, but no detectable expression is observed in the liver, muscle or spleen. Previous reports on the expression of the *Ndp* gene have found it to be expressed in the brain and eye (retina) (Chen et al., 1995), there is no mention of expression in the heart, but there is a report of expression in the fetal kidney (Meitinger et al., 1993).

The methylation profiles of the female brain, heart and kidney contain unmethylated sites in the proximal promoter correlating with the observed expression in these tissues by RT-PCR. Totally methylated clones in these tissues are presumably derived from the inactive X-chromosome. The profile of the female liver has 2 sequences (2f1 and 3f1) which also exhibit some unmethylated sites in the same region as the expressing tissues, but no expression was observed by RT-PCR in the liver sample. The sequences derived from whole eye samples and tissues of the muscles and spleen exhibited no unmethylated sites, correlating with the lack of expression of the *Ndp* gene in the muscle and heart. The lack of unmethylated sequences in samples from whole mouse eyes is most likely a result of the sequence contributions by other tissues in the eye which do not express the *Ndp* gene. None of the 11 sequences from male tissues of the Eye, Kidney or Heart exhibit unmethylated sites. These results are considered in detail in the discussion (see section 4.3).

4.2.4 Expression analysis.

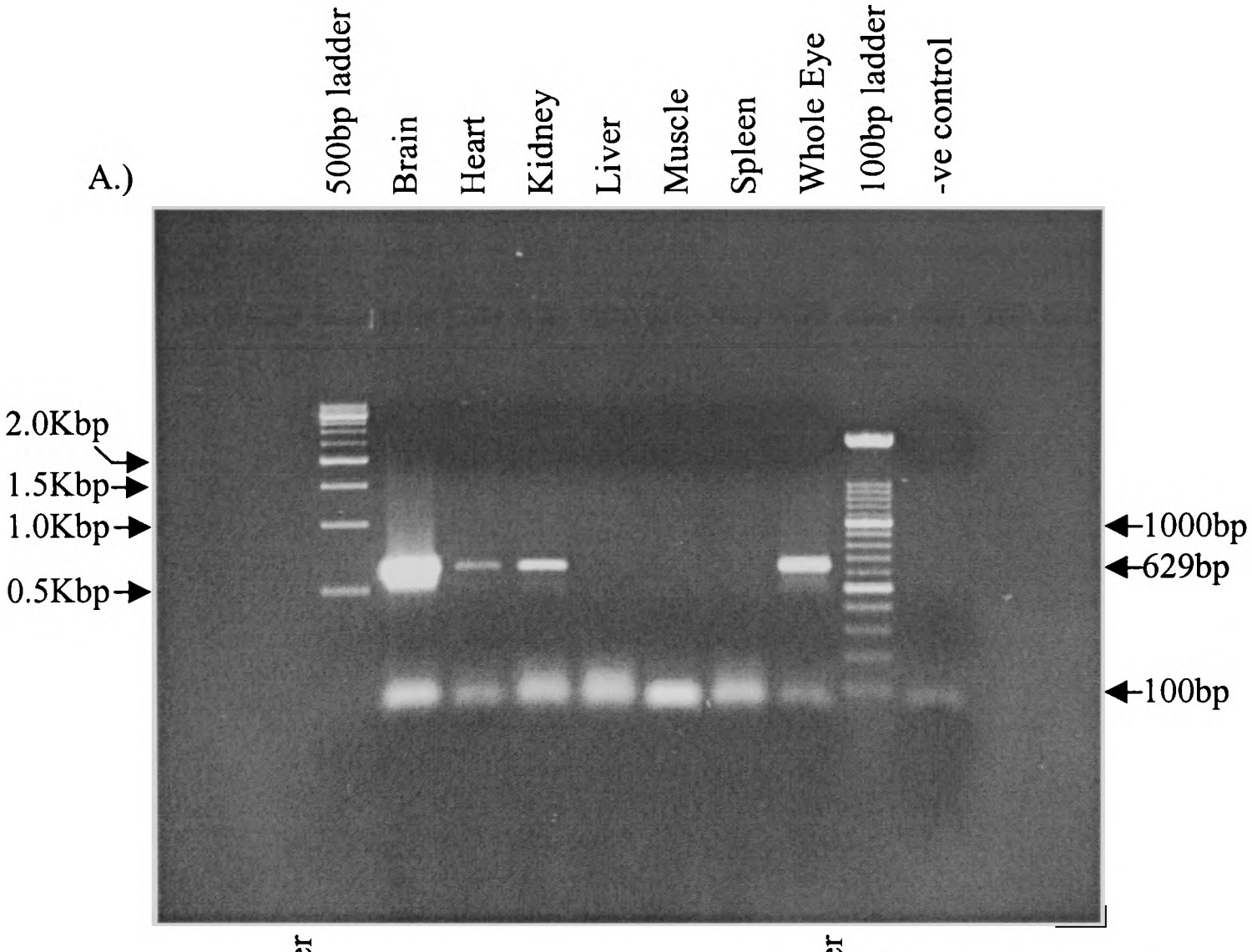


Figure 4.28: Expression analysis of the murine Ndp gene in multiple tissue samples. Total RNA was prepared from samples of freshly isolated murine tissues and these used in first strand cDNA synthesis reactions with random hexamer primers. The cDNAs prepared from each tissue were used as substrates in PCR reactions with the primers MNDPCDNAF and MNDPCDNAR designed to amplify a 629bp fragment of the Ndp cDNA. The Ndp gene was found to be expressed primarily in the brain and eye with weaker products observable in the kidney and heart (A). To demonstrate that cDNA had been successfully synthesised from all tissues, the primers MPDHCDNAF1 and MPDHCDNAR1 were used to amplify a 756bp fragment of the Pdha1 cDNA (B).

The section of the human *NDP* gene promoter chosen for methylation analysis by bisulphite sequencing had previously been shown to act as a promoter in expression constructs driving the luciferase reporter gene (Kenyon and Craig, 1999). In order to confirm that this was also the case for the homologous section of the mouse promoter, expression constructs were generated containing the promoter in forward and reverse orientations driving the EGFP reporter gene. Equivalent expression constructs were also made for the human *NDP* gene promoter so that the effect of methylation on the promoters of both human and murine *NDP/Ndp* genes could be assessed.

i) Generation of expression constructs.

Expression constructs were made in essentially the same way as for the PDHA1-EGFP expression constructs in chapter 3 (see section 3.2.4 (ii)). Briefly, the human expression constructs were made by amplifying a 752bp fragment of the *NDP* gene (–594 to +158bp) using the primers HNDPEGFPF (5'-AGATGAAGGGGCTTATTAAAG-3') and H&MNDPEGFPR (5'-ACTTCCAAATGCTTTTGTTC-3') using the optimised conditions shown in table 4.4. The PCR product (Figure 4.29A) was *PfU* "polished", purified and then blunt cloned into the *Eco RV* site of the vector pZErO2. Clones carrying the correct insert in forward and reverse orientations were identified and their inserts excised with *Bam HI* and *Xho I* and cloned into the vector pEGFP-1. Clones containing the human promoter in the EGFP-1 vector in forward and reverse orientations were identified by their different restriction patterns when digested with *Pvu II* (Figure 4.29B).

Murine forward and reverse expression constructs were made in an identical manner to those of the human constructs (above). The murine insert was generated by amplifying a 751bp fragment of the *Ndp* gene (-594 to +158bp) with the primers MNDPEGFP1F (5'-

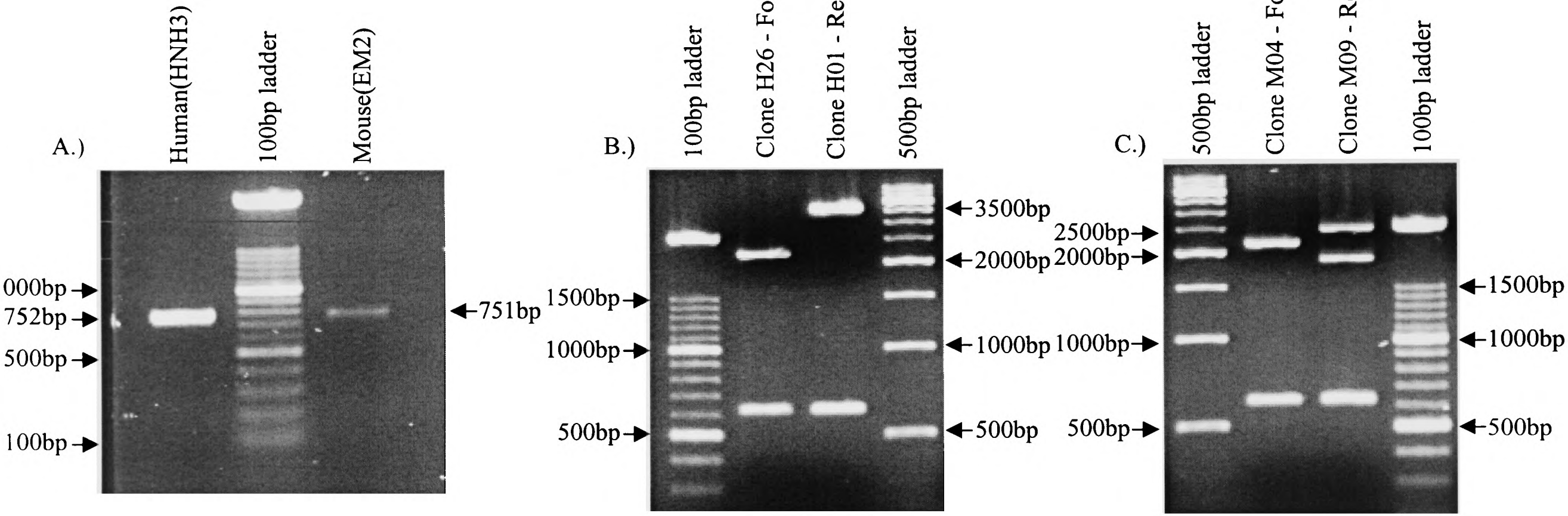


Figure 4.29: Generation of the human and murine NDP-EGFP/Ndp-EGFP expression constructs.
Expression constructs containing the human and murine NDP/Ndp gene promoters in both forward and reverse orientations were made in the vector pEGFP1. The human promoter insert was made by amplifying a 752bp fragment from the plasmid clone HNH3 (J. R. Kenyon, D. Phil. thesis) using the primers HNDPEGFPF and H&MNDPEGFPR (A). The plasmid HNH3 contains a 1273bp Hind III insert containing the human gene promoter, exon 1 and part of intron 1.The murine insert was generated by amplifying a 751bp fragment from the plasmid sub-cloneEM2 (J. R. Kenyon, D. Phil. thesis) using the primers MNDPEGFPF and H&MNDPEGFPR (A). The EM2 clone contains a 2488bp Eco RI insert containing the murine gene promoter, exon 1 and part of intron 1. The PCR products were separated from primers on an agarose gel and then *PfU* "polished" prior to blunt-cloning into the Eco RV site of the vector pZErO2. Clones containing the correct sized inserts in both forward and reverse orientations were identified by colony PCR (not shown). Inserts from these pZErO2 clones were excised with Bam HI and Xho I and cloned into the MCS of the expression vector pEGFP1. Clones containing the promoter inserts in forward and reverse orientations were identified by unique restriction patterns when digested with the enzyme Pvu II (B and C).

CAGTGAAGGCACTTATTAAAG-3') and H&MNDPEGFPR (5'-ACTTCCAAATGCTTTTGTTC-3') using the optimised conditions shown in table 4.4. The product was blunt-cloned into the *Eco RV* site of the vector pZErO2 and then into the expression vector pEGFP1. Clones containing the murine promoter in the EGFP-1 vector in forward and reverse orientations were identified by their different restriction patterns when digested with *Pvu II* (Figure 4.29C).

Human *NDP* promoter PCR (752bp product) HNDPEGFPF : H&MNDPEGFPR Qiagen buffer	96°C for 2'00'' 96°C for 0'30'' 57°C for 0'30'' 72°C for 1'00'' } 35 cycles
Murine *Ndp* promoter PCR (751bp product) MNDPEGFPF : MNDPEGFPR Qiagen buffer	96°C for 2'00'' 96°C for 0'30'' 55°C for 0'30'' 72°C for 1'00'' } 35 cycles

Table 4.4 Conditions for amplifying human and murine promoter fragments for the generation of expression constructs.

ii) Functional analysis of the human *NDP* promoter-reporter gene constructs.

The effect of artificial methylation on the ability of the promoter fragment to drive EGFP expression was examined by methylating the forward and reverse constructs with the enzyme *Sss I* methylase (see Methods 2.48). "Dummy" methylation experiments were also set up to demonstrate that the effects of methylation were not derived from anything other than actual methylation itself (see Chapter 3 section 3.2.4(iii)). Complete methylation of constructs was confirmed by their total resistance to digestion with *Hpa II* (data not shown).

The choice of cell lines to transfect the human *NDP* expression constructs into was based on the known expression pattern of the *NDP* gene *in vivo*. The main site of Norrie expression is in the retina so the human retinal cell lines Weri-Rb-1 and ARPE-19 (see Materials 2.14) were selected as possible cell lines for transfection. RT-PCR experiments have previously

demonstrated that the *NDP* gene is expressed in these cell lines (Kenyon and Craig, 1999). Transfection of the *NDP* forward EGFP construct into ARPE-19 and Weri-Rb-1 cells resulted in no EGFP expression (data not shown). This section of promoter had previously been shown to drive luciferase reporter gene expression in Weri-Rb-1 cells (Kenyon and Craig, 1999) so the inactivity in both ARPE-19 and Weri-Rb-1 cells was unexpected. The lack of EGFP product in these cells was most likely a result of the greater sensitivity of luciferase assays as opposed to the qualitative visualisation of EGFP fluorescence.

It was possible that other tissue culture cell lines might express the *NDP* gene at a level that would allow visual detection of EGFP product from the expression construct. The RT-PCR studies in mouse tissues indicated that kidney cells expressed the *Ndp* gene so one readily available kidney cell line named 293 (ATCC No. 45504 and Materials 2.14) was tested in transfection experiments with the forward *NDP* expression construct.

Aliquots of human kidney 293 cells were plated out the evening before transfections into the wells of an 8-well multi-plate containing sterile cover-slips and culture medium (see Methods 2.50). Cells were cultured overnight and the next day 2μg aliquots of either unmethylated, "dummy" methylated or methylated forward and reverse expression constructs or the mitochondrial targeted mtEGFP construct were transfected into the cells. Transfection was performed using the lipid-based Fugene™6 reagent as described previously (see Methods 2.51). Transfected cells were cultured for 48 hours then fixed in paraformaldehyde and mounted with a propidium iodide counter stain (see Methods 2.52-54). Slides were visualised with a Zeiss (Axioskop) photo microscope under UV illumination using filters designed for the viewing of either FITC (green) or Texas Red (red) fluorescent signals. Images were captured digitally by a mounted CCD camera connected to a MacRed™ computer running the

IPlab™ and IPlab extensions™ software packages. Images were taken at the same exposure and magnification so as to allow direct comparison of results and were stored as TIF compressed image files.

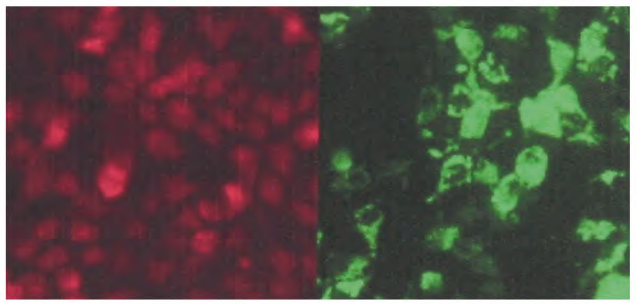
Figure 4.30 shows the results of transfecting the mtEGFP and forward *NDP* promoter-EGFP constructs into human kidney 293 cells. The mtEGFP construct directs expression of the EGFP protein targeted to the mitochondria of the cells under the control of the strong SV40 promoter (Figure 4.30A). The mtEGFP transfection was used as a qualitative control for transfection between experiments and to assess the quality of fixation by batch. The unmethylated and "dummy" methylated NDP-EGFP forward constructs are seen to direct expression of the EGFP reporter gene in 293 cells (Figure 4.30B and C). The methylated NDP-EGFP construct does not direct EGFP expression in 293 cells (Figure 4.30D).

The conclusion from these experiments is that human embryonic kidney cells are capable of using the *NDP* gene promoter to drive reporter gene expression, but lose that ability when the construct is methylated *ex-vivo*. Figure 4.31 demonstrates that the NDP-EGFP reverse constructs do not direct EGFP expression in agreement with the equivalent experiments with NDP-luciferase reverse constructs in Weri-Rb-1 cells (Kenyon and Craig, 1999). This confirms that the *NDP* gene promoter contains directional transcription control elements typical of many genes with limited expression profiles.

iii) Functional analysis of the murine *Ndp* promoter-reporter gene constructs.

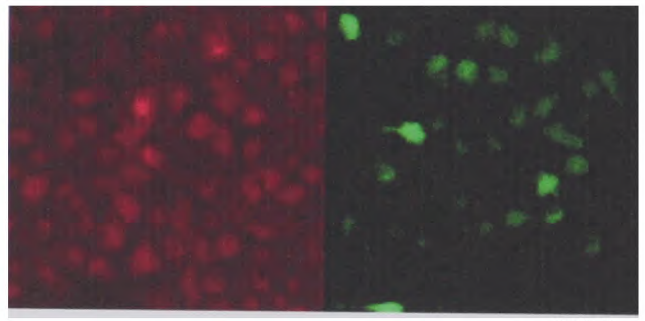
The murine *Ndp*-EGFP constructs were used in identical transfection experiments to those described above. The cell line chosen for the transfections was the 661W retinal derived cell line (kind gift of Drs A. Wright and Al-Ubaidi) which is known to express the Norrie gene

A.)



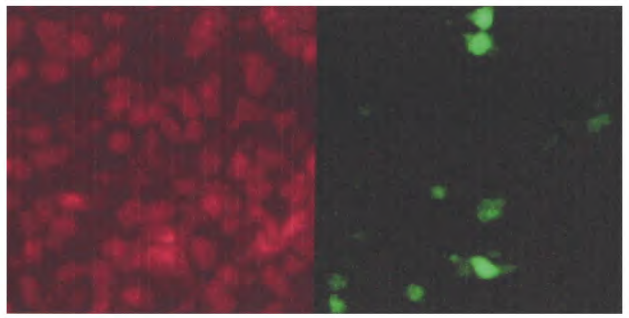
x200

B.)



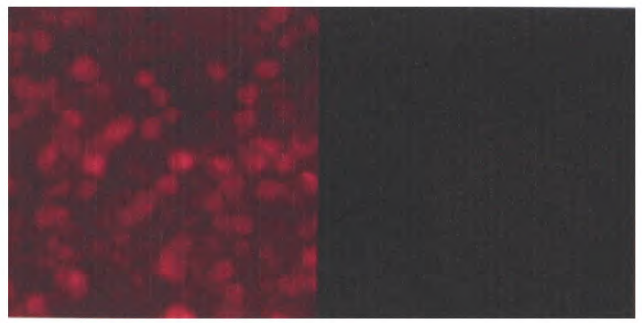
x200

C.)



x200

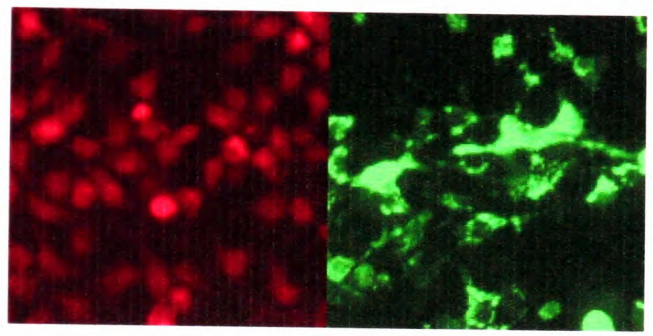
D.)



x200

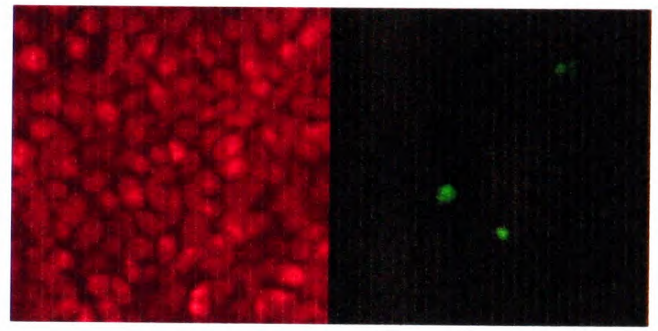
Figure 4.30: Transfection of forward NDP gene promoter-EGFP expression constructs into human kidney 293 cells.
Human kidney 293 cells were transfected with a variety of vector constructs using the transfection reagent Fugene™6 (see text). The expression vector mtEGFP expresses a fusion protein of EGFP targeted to the mitochondria of cells under the control of an SV40 promoter. This vector was transfected into 293 cells to test conditions for transfection and to confirm the quality of cell fixation. The resulting transfected cells expressed the mtEGFP-fusion presenting a typical fibrous network of fluorescent signal (A). The identical transfection conditions were used in parallel for transfection experiments with unmethylated, "dummy" methylated and methylated expression constructs containing the forward orientation NDP gene promoter upstream of the EGFP coding sequence (see text). The unmethylated construct expresses the EGFP gene in 293 cells presenting a diffuse cytoplasmic signal (B) as does the "dummy" methylated construct (C). Transfections with the totally methylated construct did not result in the production of any signal in cells (D). These results demonstrate that the section of the human NDP gene promoter cloned into the pEGFP expression vector is capable of acting as a functional promoter in 293 cells and that this ability is removed when the promoter is methylated.

A.)



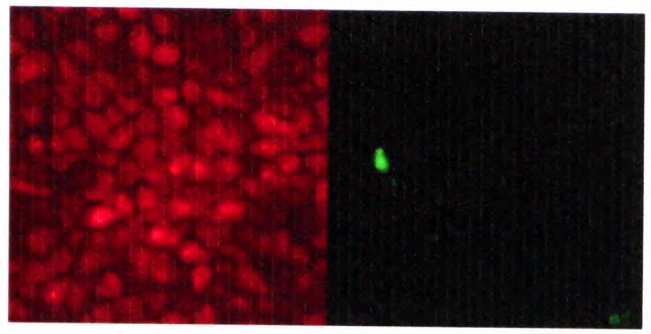
x200

B.)



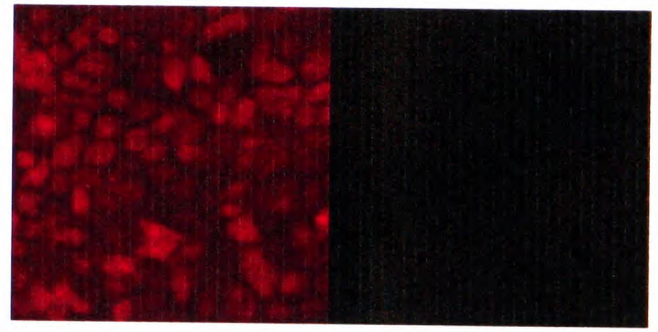
x200

C.)



x200

D.)



x200

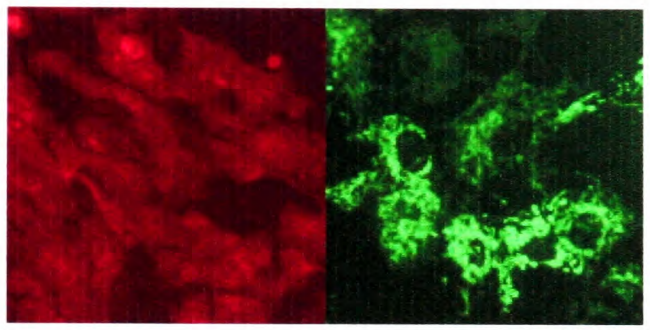
Figure 4.31: Transfection of reverse NDP gene promoter-EGFP expression constructs into human kidney 293 cells.
Human kidney 293 cells were transfected with a variety of vector constructs using the transfection reagent Fugene™6 (see text). The expression vector mtEGFP expresses a fusion protein of EGFP targeted to the mitochondria of cells under the control of an SV40 promoter. This vector was transfected into 293 cells to test conditions for transfection and to confirm the quality of cell fixation. The resulting transfected cells expressed the mtEGFP-fusion presenting a typical fibrous network of fluorescent signal (A). The identical transfection conditions were used in parallel for transfection experiments with unmethylated, "dummy" methylated and methylated expression constructs containing the reverse orientation NDP gene promoter upstream of the EGFP coding sequence (see text). Neither the unmethylated (B), "dummy" methylated (C) or totally methylated construct (D) expressed the EGFP reporter gene in 293 cells. These results demonstrate that the promoter of the NDP gene directs gene expression in an orientation dependent manner as described previously (J. R. Kenyon, D. Phil. thesis).

and exhibits many of the properties of retinal cells (Kido et al., 1996). Transfection experiments were carried out as detailed above using the transfection reagent Fugene™6 with unmethylated, "dummy" methylated and methylated forward and reverse *Ndp*-EGFP constructs as well as the mitochondrial targeted mtEGFP construct. Cells were fixed, mounted in Moviol mountant containing the counter-stain propidium iodide and visualised as described above.

Figure 4.32 shows the results of transfecting the mtEGFP and forward *Ndp* promoter-EGFP constructs into murine 661W retinal cells. The mtEGFP construct directs expression of the EGFP protein giving rise to a typical fibrous network of green fluorescence (Figure 4.32A). The intensity of EGFP signal over background fluorescence in the unmethylated and "dummy" methylated construct transfection experiments was low. The fluorescence digital images acquired from these were not representative of the EGFP expression observable with the dark-adapted eye. This was primarily a result of the image acquisition software being unable to differentiate low level fluorescence signals from the natural auto-fluorescence of the cells. This auto-fluorescence could be caused by the pigment bodies that are present in retinal and photoreceptor derived cell lines. When the fluorescence on the slides was analysed by eye the forward unmethylated and "dummy" methylated *Ndp*-EGFP constructs expressed the EGFP protein in 661W cells (Figure 4.32B and C), but the methylated construct did not directly paralleling the results in human kidney 293 cells.

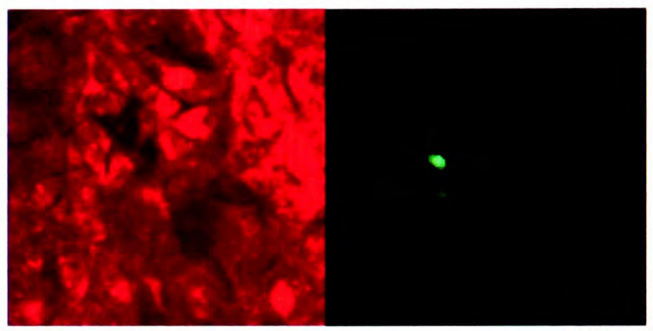
Transfection experiments were repeated with the unmethylated, "dummy" methylated and methylated *Ndp*-EGFP reverse constructs and with the mitochondrial targeted EGFP construct in 661W cells (Figure 4.33A-D). There was no detectable expression of the EGFP

A.)



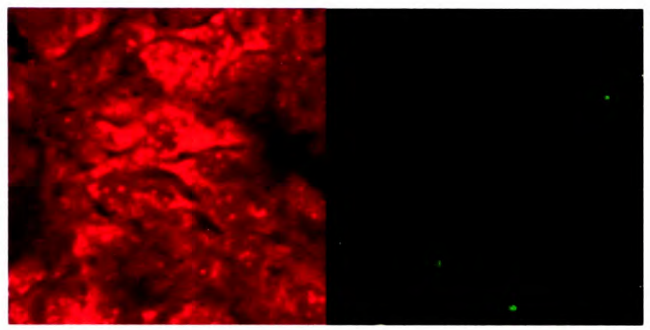
x200

B.)



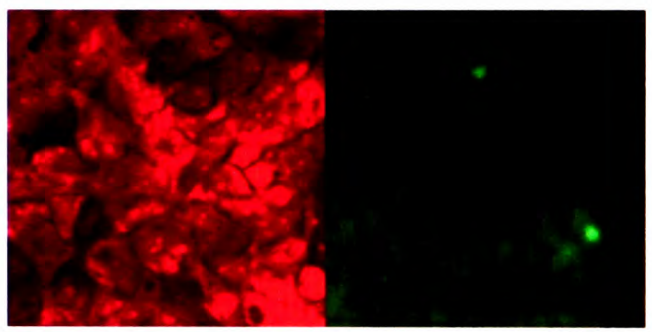
x200

C.)



x200

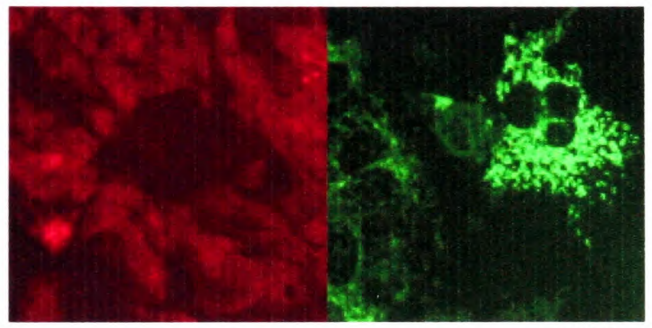
D.)



x200

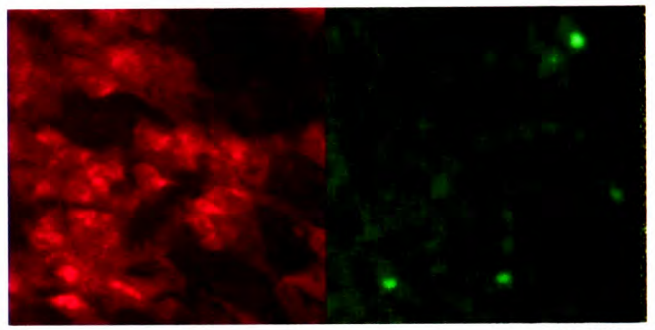
Figure 4.32: Transfection of forward Ndp gene promoter-EGFP expression constructs into murine retinal 661W cells.
Mouse retinal 661W cells were transfected with a variety of vector constructs using the transfection reagent Fugene™6 (see text). The expression vector mtEGFP expresses a fusion protein of EGFP targeted to the mitochondria of cells under the control of an SV40 promoter. This vector was transfected into 661W cells to test conditions for transfection and to confirm the quality of cell fixation. The resulting transfected cells expressed the mtEGFP-fusion presenting a crisp fibrous network of fluorescent signal (A). The identical transfection conditions were used in parallel for transfection experiments with unmethylated, "dummy" methylated and methylated expression constructs containing the forward orientation Ndp gene promoter upstream of the EGFP coding sequence (see text). Low levels of signal were observed in the 661W cells with the unmethylated (B) and "dummy" methylated (C) forward construct, but the signals were obscured by auto-fluorescence in the retinal 661W cells. The CCD camera used to grab images was unable to acquire signal over the auto-fluorescence although EGFP expression in the cells was clearly observable with the dark-adapted eye. The methylated construct did not produce any signal over auto-fluorescence either as acquired by the CCD camera or by the dark-adapted eye (D). These results suggest that the region of the Ndp gene promoter cloned into these constructs can act as a functional promoter in 661W cells and that methylation suppresses this activity, but transfection experiments using other Ndp expressing cell lines would be needed to confirm this.

A.)



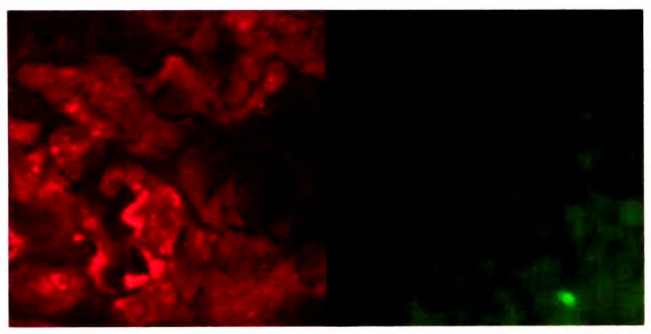
x200

B.)



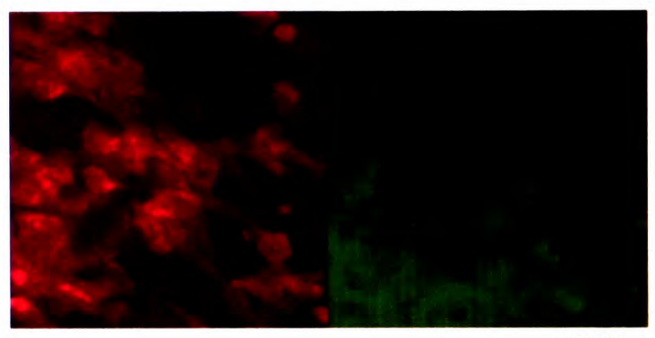
x200

C.)



x200

D.)



x200

Figure 4.33: Transfection of reverse Ndp gene promoter-EGFP expression constructs into murine retinal 661W cells.
Mouse retinal 661W cells were transfected with a variety of vector constructs using the transfection reagent Fugene™6 (see text). The expression vector mtEGFP expresses a fusion protein of EGFP targeted to the mitochondria of cells under the control of an SV40 promoter. This vector was transfected into 661W cells to test conditions for transfection and to confirm the quality of cell fixation. The resulting transfected cells expressed the mtEGFP-fusion presenting a crisp fibrous network of fluorescent signal (A). The identical transfection conditions were used in parallel for transfection experiments with unmethylated, "dummy" methylated and methylated expression constructs containing the reverse orientation Ndp gene promoter upstream of the EGFP coding sequence (see text). Neither the unmethylated (B), "dummy" methylated (C) or methylated (D) constructs expressed the EGFP reporter gene in the 661W cells as assessed by the CCD camera used to acquire the images and by the dark-adapted eye. These results suggest that the promoter of the murine Ndp gene directs gene expression in an orientation dependent manner as does the human NDP promoter. Further transfection experiments using other Ndp expressing cell lines with lower auto-fluorescence would be needed to confirm these results.

product in 661W cells transfected with the *Ndp*-EGFP reverse constructs either digitally or by eye.

In conclusion, the homologous sections of the human and murine *NDP/Ndp* gene promoters cloned into the EGFP-1 expression vector express the reporter gene in human kidney 293 and mouse retinal 661W tissue culture cells. These cell types represent tissues which are known to express the *NDP/Ndp* gene *in vivo*. Methylation of the human and murine promoters in these expression constructs prohibits the expression of the reporter gene suggesting that methylation of the promoter might be capable of down-regulating the gene *in-vivo*.

4.3 Discussion.

The detailed analysis of methylation at the promoters of X-linked genes has so far been limited to either genes which escape X-chromosome inactivation such as *PCTK-1* and *UBE1* or to housekeeping genes subject to X-chromosome inactivation such as *PGK-1*, *HPRT* and in this thesis *PDHA1*. To date the only studies to have used the bisulphite genomic sequencing technique to determine the methylation status of the promoter of an X-linked tissue specific gene are those on the Fragile-X (*FMR1*) gene promoter. The *FMR1* studies concentrated on the methylation status of an unstable triplet repeat in the promoter rather than on detailed methylation analysis of the entire promoter.

The study of a pair of homologous gene promoters provides an opportunity to identify functionally essential elements by their conservation in both promoter sequences. By the same measure it is possible that conservation of methylatable sites between homologues

could identify sites that are needed as part of the gene's control mechanisms. The human and murine *NDP/Ndp* gene promoters share 90% nucleotide identity in the proximal 160bp of upstream sequence. This part of the human *NDP* promoter has been shown to contain the core promoter activity (Kenyon and Craig, 1999) so the high level of sequence conservation suggests that functional promoter elements might be common between these two genes. The close sequence identity between the core promoters of the two genes increases the confidence that studies in the mouse could be used to make predictions about the human gene. These predictions could then be tested on human material in a much more efficient manner which is especially important when sources are difficult to obtain.

In designing experiments for the study of methylation it is necessary to characterise the number and distribution of methylatable sites within the region(s) of interest. In this chapter, analysis of the human and murine gene promoters (Figures 4.1, 4.2, 4.3 and 4.12) reveals that they are very similar with respect to CpG site number, distribution and (C+G)% content. Neither gene promoter when analysed by the CpG-plot program was found to contain a CpG-island. It was not known at this point how likely it was that any methylation would be present at these promoters because up until then studies of methylation had been restricted to genes on the X-chromosome that had CpG-islands at their 5'-ends. It is important to state that there is no *a priori* reason to assume that the absence of a CpG-island precludes the possibility that a section of DNA is methylated. It is simply the presence of a potentially methylatable site i.e. a CpG di-nucleotide (although rarely other cytosines can be methylated in gene promoters) that provides this potential.

The comparative studies of the composition of the human and murine *NDP/Ndp* gene promoters revealed that the murine sequence had lost approximately 5% of its (C+G)%

content and 5 of the 13 CpG sites present in the human sequence. This loss of CG content is consistent with the observation that the murine genome contains approximately 70% of the number of CpG-islands of the human (Cross et al., 1997). The cause of this loss of CG content is thought to be a combination of the increased rate of mutation at methylated cytosines, which mutate 9x more frequently than do unmethylated ones (Bird, 1980); and of the fact that the murine generation time is much shorter than the human one resulting in a differential rate of loss between man and mouse over evolutionary time. The loss of methylatable sites in the murine promoter provides an opportunity to identify those sites which could be conserved for functional reasons. Figure 4.34 shows an alignment of the human and murine promoter sequences in the region –273 to +3bp which includes all of the CpG sites whose methylation status were characterised. In the core promoter section –160 to +3bp the murine sequence has three less CpG sites than the homologous part of the human promoter. These three changes could have been caused by mutations in the minus strand cytosines of CpG pairs. Two of the changes are consistent with the cytosine to thymine transitions that result from methylation induced deamination. The other change may simply represent a natural transversion mutation of the cytosine to an adenine base.

The result of this analysis is that in the very highly conserved core promoter region of the two genes a quarter of the CpG sites have been lost in the mouse implying that these sites were not essential for the control of expression. In the core promoter, 9 of 12 CpG sites are conserved between man and mouse. The four CpG sites at –58, -49, -45 and –40bp saturate the putative NRL binding site, but the other five do not lie in the consensus binding sites of any known transcription factors.

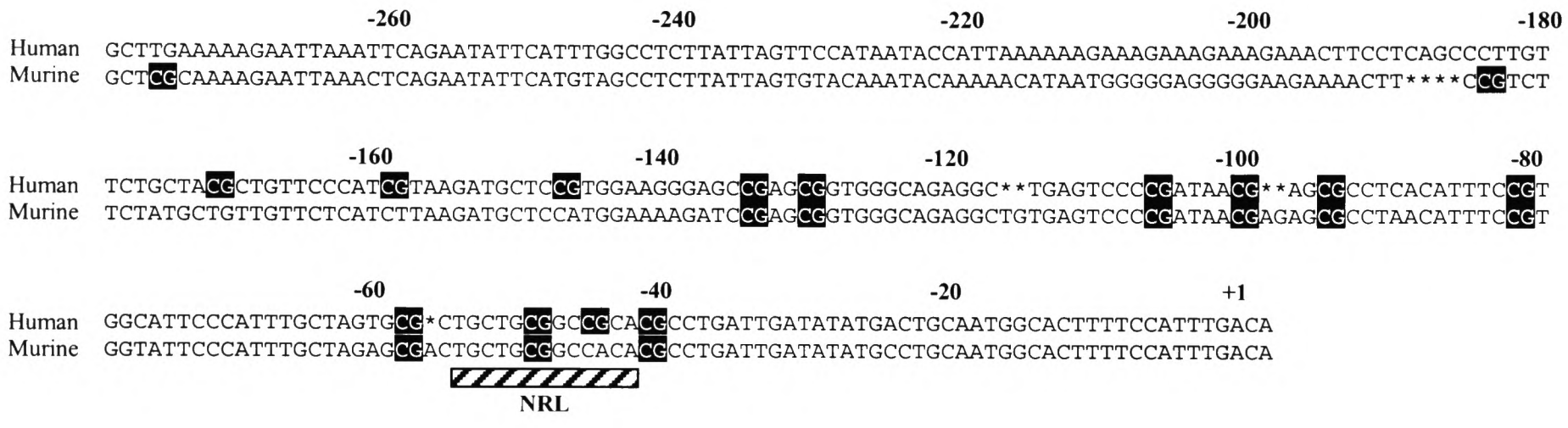


Figure 4.34: Gapped alignment of the human and murine gene promoters in the bisulphite target region.
Alignment of the human and murine NDP/Ndp gene promoters in the bisulphite target regions up to and including the transcription start site. The positions of CpG sites are highlighted in black and the position of the NRL consensus binding site is indicated for reference. The murine sequence has lost 3 CpG sites in the core promoter region when compared to that of the human. In each case the loss is accounted for by a mutation in the minus strand cytosine of what is presumed to have been a CpG di-nucleotide. Two of the mutations are cytosine to thymine transitions typical of CpG methylation induced loss. The positioning of 4 CpG sites around the putative NRL binding site is of particular interest (see text).

The comparative sequence analysis of the two promoters and the knowledge that the core promoter activity of the gene lay within 160bp of the transcription start site determined the region to be studied by bisulphite sequencing.

Methylation at the human *NDP* gene promoter.

The studies of methylation in the Weri-Rb-1 cell line were particularly interesting because the sequenced section contained a polymorphism allowing the two X-chromosomes to be distinguished. Bisulphite sequencing produced some sequences which contained only totally methylated CpG sites and these were always associated with the same allele of the polymorphism. The other sequences were either totally unmethylated or very lightly methylated and were always associated with the other allele of the polymorphism. The absolute association of the same allele with totally methylated sequences implied that the cell line was derived from a clonal cancer with the same X-chromosome inactive in each cell. The puzzling observation was that some of the sequences containing the active X-chromosome allele were lightly methylated at the promoter. A possible explanation for this could be that methylation at the *NDP* promoter was re-programmed during the transformation process. Transformation often involves de-differentiation, reprogramming of methylation and the down-regulation of many tissue-specific genes (Bird, 1996). Previous studies of the retinoblastoma cell line Weri-Rb-1 have demonstrated that it retains some, but not all the features of retinal cells, implying that it is at least part way down a de-differentiation route. The sequences that show low-levels of methylation on the presumed active X-chromosome do so in the exact part of the promoter that corresponds to the position of the NRL binding site. It is tempting to speculate that the cells which have the active X-chromosome promoter methylated in these few sites either do not express the *NDP* gene or express it at a much

reduced level. This could be directly tested by in-situ hybridisation of *NDP*-riboprobes to fixed Weri-Rb-1 cells to see if there is mosaic expression of *NDP* mRNA between the cells.

The bisulphite sequences obtained from human retinal tissue confirmed the results from the Weri-Rb-1 cells. In both, the active X-chromosome was totally unmethylated and the inactive X-chromosome methylated. What was most striking about the results from the human retinas was that the level of methylation on the inactive X-chromosomes was much lighter than in Weri-Rb-1 cells. The restriction of methylation to the CpG sites saturating the NRL binding site in female retinal samples implies that methylation in this region could be correlated to repression of the gene.

There are a number of possible ways in which CpG methylation of a gene promoter could affect the expression of the gene and these are described in detail in chapters 1 and 6. From the perspective of the *NDP* promoter on the inactive X-chromosome of the human retinal samples it seems unlikely that such a low level of methylation could induce the changes in chromatin required to inactivate the gene although this cannot be ruled out. It is perhaps more reasonable to suggest that either methylation itself or the recruitment of proteins such as MeCP2 to bind to the methylated target prevents the transcription factor(s) from binding to their cis-acting site and inducing gene expression. There is some evidence to support this idea. The NRL transcription factor is a member of the leucine-zipper family and binds to its target site as either an NRL-NRL homodimer or NRL-fos or NRL-jun heterodimers (Kerppola and Curran, 1994). As described in chapter 3, methylation of c-fos or c-jun binding sites ablates protein binding in-vitro. As the NRL is closely related to c-fos and c-jun it is possible that methylation of its binding site would also prevent its binding.

The data from the human bisulphite sequencing has provided a wealth of information which has established clear ideas about how the gene could be inactivated on the inactive X-chromosome. These ideas could be tested experimentally by assessing the ability of recombinant NRL protein to bind to its putative binding site in the human *NDP* promoter *in-vitro* with both methylated and unmethylated DNA probes. In addition DNase I footprinting studies using nuclear protein from human retinal samples or Weri-Rb-1 cells could identify any regions of protein binding in methylated and unmethylated *NDP* promoter probes. The involvement of the NRL protein in the control of the *NDP* gene is dependent upon its expression in the human retinal samples used in this chapter and in Weri-Rb-1 cells. The NRL gene is expressed in the brain, spinal cord and retina during development (Liu et al., 1996). Its expression becomes limited solely to the retina in the adult implying that it should be expressed in the samples used in this chapter. This could be confirmed experimentally by RT-PCR.

Methylation at the mouse *Ndp* gene promoter.

The study of the human *NDP* gene promoter provided data that could be used to help interpret the results of the studies on the murine *Ndp* promoter. If methylation of the human promoter was indeed correlated with inactivity of the gene *in-cis* then the first prediction that could be made would be that unmethylated sequences should only be seen in tissues that express the *Ndp* gene in the mouse.

The RT-PCR analysis of total RNA isolated from the mouse tissues demonstrated that the brain, heart, kidney and whole eye expressed the *Ndp* gene (see Figure 4.28A). These results agree with previous reports of the expression profile of the gene with the exception of heart tissue which has not previously been reported to express the gene. The intensity of the

product cannot be accurately quantitated in this system, but the indication from these experiments in which identical amounts of cDNA were used in each reaction (as confirmed in part by the *Pdha1* RT-PCR results in Figure 4.28B) is that the amount of product in the kidney is small and even less is present in the heart.

If methylation is correlated to the repression of the *Ndp* gene then it would be expected that unmethylated sequences would only be found in tissues expressing the *Ndp* gene. A schematic of the methylation profiles in each tissue is shown in Figure 4.27. The schematic shows that unmethylated sequences were observed in the brain, heart, kidney, liver and in peripheral blood. These results demonstrate that the expressing tissues of the brain, heart and kidney have unmethylated sequences within the core promoter section in some of the bisulphite sequences. However some bisulphite sequences from DNA of the liver and peripheral blood which do not express the *Ndp* gene also contained unmethylated sites in the core promoter. This implies, quite logically, that it is not simply the methylation status of the core promoter that determines the activity of the gene, but also the required presence of tissue specific factors such as the NRL protein. Muscle and spleen samples were not found to express the *Ndp* gene and produced only totally methylated clones although only a small number were sequenced. Bisulphite sequences of DNA from male or female whole eye samples and from male heart and kidney samples contained no unmethylated sites.

The fact that none of the bisulphite sequences derived from murine whole eye samples were found to have unmethylated sites was likely to be due to contamination of *Ndp* expressing cells with non-expressing cells. The bisulphite sequences from human eye samples were derived from dissected pure retina, the exact part of the eye known to express the *NDP* gene. In contrast, the murine whole eye samples were derived from homogenised whole eyes

containing a mixture of cell types many of which do not express the *Ndp* gene. The yield of genomic DNA from expressing retinal cells from the very small amount of retina present in a single mouse eye is likely to be small. As a result the bisulphite PCR products derived from the DNA of expressing retinal cells are out-numbered by those from non-expressing cells because every cell has a perfect template for the *Ndp* bisulphite PCR reaction. The sequencing of mouse eye bisulphite clones was limited to seven because every one was totally methylated in male and female samples. These arguments are not inconsistent with the *Ndp* RT-PCR result performed on murine whole eye RNA which produced a strong product band because the RT-PCR simply shows that some cells in the eye express the *Ndp* gene it does not indicate what proportion of cells express the gene.

The observation that no male bisulphite clones of heart and kidney tissues contained unmethylated sites was surprising. Some bisulphite sequences from female heart and kidney samples contained unmethylated sites, 40% and 27% of sequences respectively. This is notably less than 50% in both cases implying that the active X-chromosome is methylated in some female cells. Assuming that the unmethylated sequences of female samples are derived from the active X-chromosome the male samples would also be expected to contain clones with unmethylated sequences. It is therefore most likely that bisulphite clones derived from expressing DNA molecules from male heart and kidney samples were not sequenced. As only a total of seven bisulphite clones were sequenced from the male heart and kidney it is possible that further sequencing might have revealed unmethylated clones.

The distribution of methylation in murine bisulphite clones containing unmethylated sites is different to that present on the active X-chromosomes of either human Weri-Rb-1 cells or human retinal tissue. When the sites are unmethylated they are, like the human samples,

predominantly in the region of the NRL binding site, but unusually the sites upstream of this are methylated. This supports the idea that was established from the human results, namely that only the proximal 160bp of upstream sequence or less is required in an unmethylated or partially unmethylated form to allow expression of the

gene. These observations in murine expressing tissues are strikingly reminiscent of the observed methylation patterns at the promoter of the autosomal galectin-1 gene (Salvatore et al., 1998). At this promoter, murine expressing tissues were found to have promoters that were methylated upstream and downstream of the core promoter section and were unmethylated only in the core promoter. Interestingly, they also found that the promoters in expressing human tissues and tissue culture cells were devoid of methylation as shown here for the Weri-Rb-1 and human retinal cell active promoters. This raises the possibility that these kinds of methylation profiles are common to many genes expressed only in specific tissues.

The methylation studies at the murine *Ndp* gene promoter have taken forward the ideas proposed for the human promoter. The murine results imply as in the human that the active X-chromosome of both male and female samples are by and large methylated in non-expressing tissues. It is essential to investigate whether methylation can in fact down-regulate the promoter activity of this gene. In order to begin these investigations experiments were carried out with expression constructs containing the promoters of the human and murine *NDP/Ndp* genes driving the EGFP reporter gene.

Expression studies with the human and murine *NDP/Ndp* promoter-reporter gene constructs.

The results from the human *NDP* gene promoter expression constructs are consistent with the proposal that methylation is involved in the down-regulation of the gene's expression. The

results show that the construct containing the human promoter in forward orientation expresses the EGFP reporter gene when transfected into human kidney 293 cells. When the construct is methylated by the enzyme *Sss I* methylase, the expression of EGFP is lost. The interpretation of results such as these, where the entire vector backbone is also methylated must include the consideration that backbone methylation rather than the methylation at the gene promoter causes the down-regulation of the reporter gene.

Identical methods of assessing the effects of methylation on promoter-reporter gene constructs have been performed by a number of groups and in each case vector backbone methylation effects have been discounted (Lu and Davies, 1997; Salvatore et al., 1998; Simmler et al., 1997; Weiss et al., 1996). All of these experiments have been performed with plasmids that are transiently transfected and which cannot replicate ectopically in mammalian cells. Comparable studies using plasmids which can replicate ectopically have clearly demonstrated that vector backbone methylation does repress promoter-reporter gene activity in the absence of promoter methylation (Hsieh, 1994; Hsieh, 1997). The only apparent difference between these two sets of experiments is the ability of the plasmids to replicate in mammalian cells. As the pEGFP-1 construct used in this thesis cannot replicate ectopically in mammalian cells it has been assumed that vector backbone methylation does not significantly affect the promoters in these constructs.

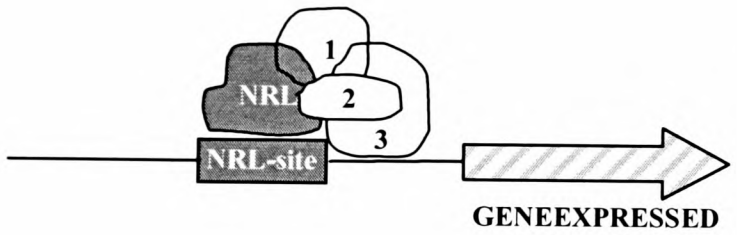
Transfection experiments into murine retinal 661W cells did not produce results which could be easily recorded because the cell line auto-fluorescence hides the EGFP signal in the cytoplasm of expressing cells. Observations by eye confirmed that the unmethylated murine forward construct expressed the reporter gene and that this expression was lost when it was methylated with the enzyme *Sss I* methylase. In order to convincingly demonstrate that these

results are genuine, the experiments would have to be repeated in another murine retinal cell line, or another reporter gene could be used whose activity would not be hidden by the 661W cell auto-fluorescence.

It has been reported that the amount and density of CpG methylation at a gene promoter determines the extent of the effect on gene expression (Hsieh, 1994). Other experiments have suggested that even dense and highly methylated CpG sites in a promoter section as large as 500bp are not necessarily sufficient to down-regulate gene activity (Hsieh, 1997). Comparable sections of the *NDP/Ndp* promoters used in these experiments contain 40% and 25% respectively of the CpG density of the *PDHA1* promoter. The low number and density of methylatable sites at the promoters of many tissue-specific genes may mean that repression mediated via methylation may not solely be achieved by the same mechanisms as for genes containing promoters with dense and highly methylated CpG sites. The *NDP* gene promoter has been shown to have a very restricted methylation profile in Weri-Rb-1 cells and human retinal tissue and may be down-regulated by methylation blocking the NRL transcription factor binding site or recruiting other proteins which bind to the methylated sequences and block access. The *Ndp* promoter has been shown to have a much higher level of methylation in general and may use both specific repression at the Nrl site and the non-specific mechanisms mediated through MBD proteins that are used ubiquitously to repress gene transcription throughout the genome. These mechanisms are summarised in figure 4.35 where three simple models for repressing *NDP/Ndp* expression via methylation are shown.

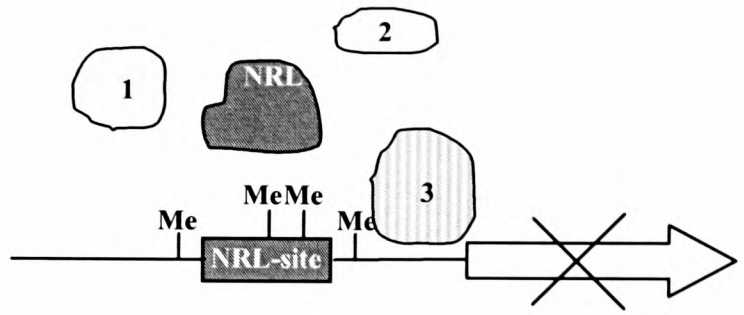
In conclusion the methylation experiments on the human and murine *NDP/Ndp* gene promoters have demonstrated that the methylation profiles *in-vivo* and in tissue culture cells are consistent with developmentally established tissue-specific methylation of the promoter

A.)



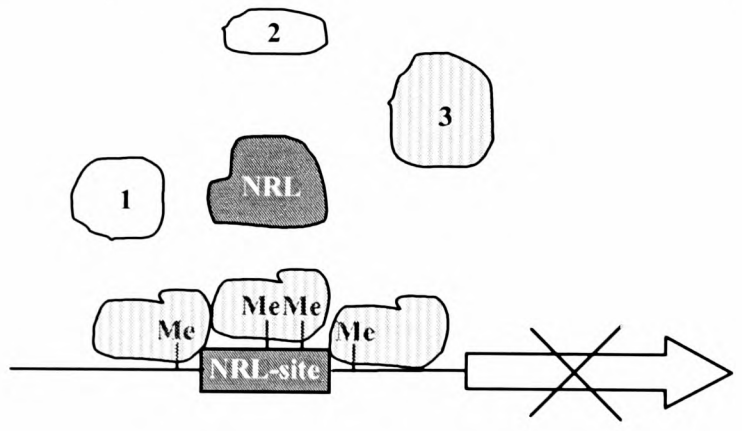
UNMETHYLATED PROMOTER

B.)



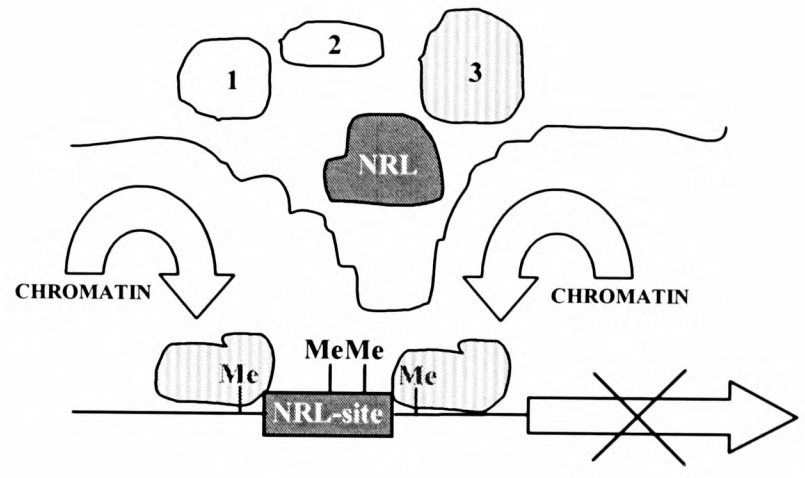
METHYLTION BLOCKS BINDING

C.)



PROTEINS BLOCK BINDING

D.)



CHROMATIN CONFIGURATION BLOCKS BINDING

Figure 4.35: Simple models to explain the possible mechanisms by which methylation affects the promoter of the NDP gene. The promoter of the NDP gene is methylated in tissues and cell lines that do not express the NDP gene. There are several possible mechanisms by which methylation could direct this effect *in-vivo*. The unmethylated promoter binds the NRL and associated transcription factors including the TATA binding protein (**1**, **2**, and **3**) initiating transcription of the gene (A). Methylation could directly affect transcription by altering the structure of the NRL DNA binding site so that the NRL protein can no longer bind to it (B). Alternatively proteins which bind in a non-specific manner to methylated sections of DNA such as the protein MeCP2 (see text) may block access of the NRL protein to its binding site (C). Finally, methylation at the NRL site and at flanking regions could seed the generation of an inactive chromatin configuration at the promoter site preventing access of the transcription factors (D).

marking non-expressed copies of the gene. The NRL protein has been implicated in the expression of the *NDP* gene in this chapter and previously (Kenyon and Craig, 1999). The consensus NRL binding site is conserved in the promoters of human and murine genes, it is contained in the core promoter region and it is the only site consistently unmethylated in bisulphite sequences of cells and tissues which express the *NDP/Ndp* gene. Transfection experiments with constructs containing the human and murine *NDP/Ndp* promoters driving the EGFP reporter gene show that *ex-vivo* methylation ablates reporter gene expression. This demonstrates that methylation is capable of directing the down-regulation of *NDP/Ndp* promoter activity in tissue culture cells and implies that it could be part of the mechanism of inactivation of the genes *in-vivo*.

The human and murine *NDP/Ndp* genes are markedly similar in the core promoter region. The results of the methylation and expression studies in this chapter show that the murine *Ndp* promoter is a good model for understanding the human gene promoter and illustrates the benefit of studying gene promoters in a comparative manner.

Chapter 5 - Isolation of the human and murine *SMCX/Smcx* gene promoters

5.1 Introduction.

This chapter describes the cloning and characterisation of the proximal upstream regions of the human *SMCX* gene and its murine homologue, *Smcx*. Cosmid clones of the human *SMCX* gene were isolated by screening the LLNL X-only cosmid library with a probe comprising approximately the first 300bp of the published human *SMCX* cDNA sequence (Wu et al., 1994)(Genbank accession number L25270). The murine *Smcx* gene was isolated by screening the RPCI21 PAC library with the identical human cDNA probe. Sub-clones of both genes' upstream regions were made and sequenced. The two sequences were compared to identify sections of nucleotide identity, known transcription factor consensus binding sites and conserved potential methylation sites.

Expression constructs were made of the *SMCX* gene upstream sequence in the expression vector pEGFP-1 to determine whether it was capable of acting as a functional promoter in tissue-culture cells. *Ex-vivo* methylation of the resulting SMCX-EGFP1 constructs with the enzyme *Sss I* methylase was used to assess the potential of methylation to regulate the promoters activity as visualised by reporter gene expression.

The isolation and study of the promoters and upstream elements of the human and murine *SMCX/Smcx* genes will contribute towards the identification of features common to genes that escape X-chromosome inactivation. This is of particular interest for the *Smcx* gene which has been shown to escape X-inactivation to different extents by tissue (Sheardown et al., 1996). An intriguing possibility is that the degree of methylation at the promoter correlates

with the degree of escape in each tissue. The isolation of the human and murine gene promoters will allow this theory to be tested directly by bisulphite genomic sequencing.

5.1.1 Genomic DNA libraries.

Genomic DNA libraries are constructed to facilitate the propagation and isolation of genome fragments for analysis. Many kinds of genomic DNA library exist and are characterised by the kind of vector used in their construction. In this chapter genomic Cosmid and PAC (P1 artificial chromosome) libraries provided by the HGMP-RC have been used to isolate clones containing the human and murine *SMCX/Smcx* genes.

i.) Cosmid libraries.

Cosmids are a variety of vector that use the bacteriophage lambda packaging system as a vehicle for transfer into the host organism (Collins and Hohn, 1978). A cosmid vector contains the *λ-cos* (cohesive end site) site which is required to package itself into the *λ-capsid* as well as selection cassettes for antibiotic resistance and unique restriction sites for cloning (Collins and Bruning, 1978). Cosmid vectors are capable of carrying larger insert fragments than plasmid (<15kb) and phage (<20kb) vectors with cosmid vectors typically able to carry inserts up to 35-45kb in size (Hohn and Collins, 1980). The schematic of figure 5.1 shows the principle behind the use of cosmids as cloning vectors. The ligation of large genomic restriction fragments into the pre-cut multiple cloning site of the cosmid vector yields long concatenated molecules which are ideal substrates for phage head packaging. These concatenated molecules are cleaved at *λ-cos* sites and fragments of 35-45kb are packaged then transformed into *λ*-sensitive *E.coli*. Once transformed into bacteria, the cosmid vectors with their genomic inserts replicate as any other plasmids and are selected for by the antibiotic resistance cassette that they carry.

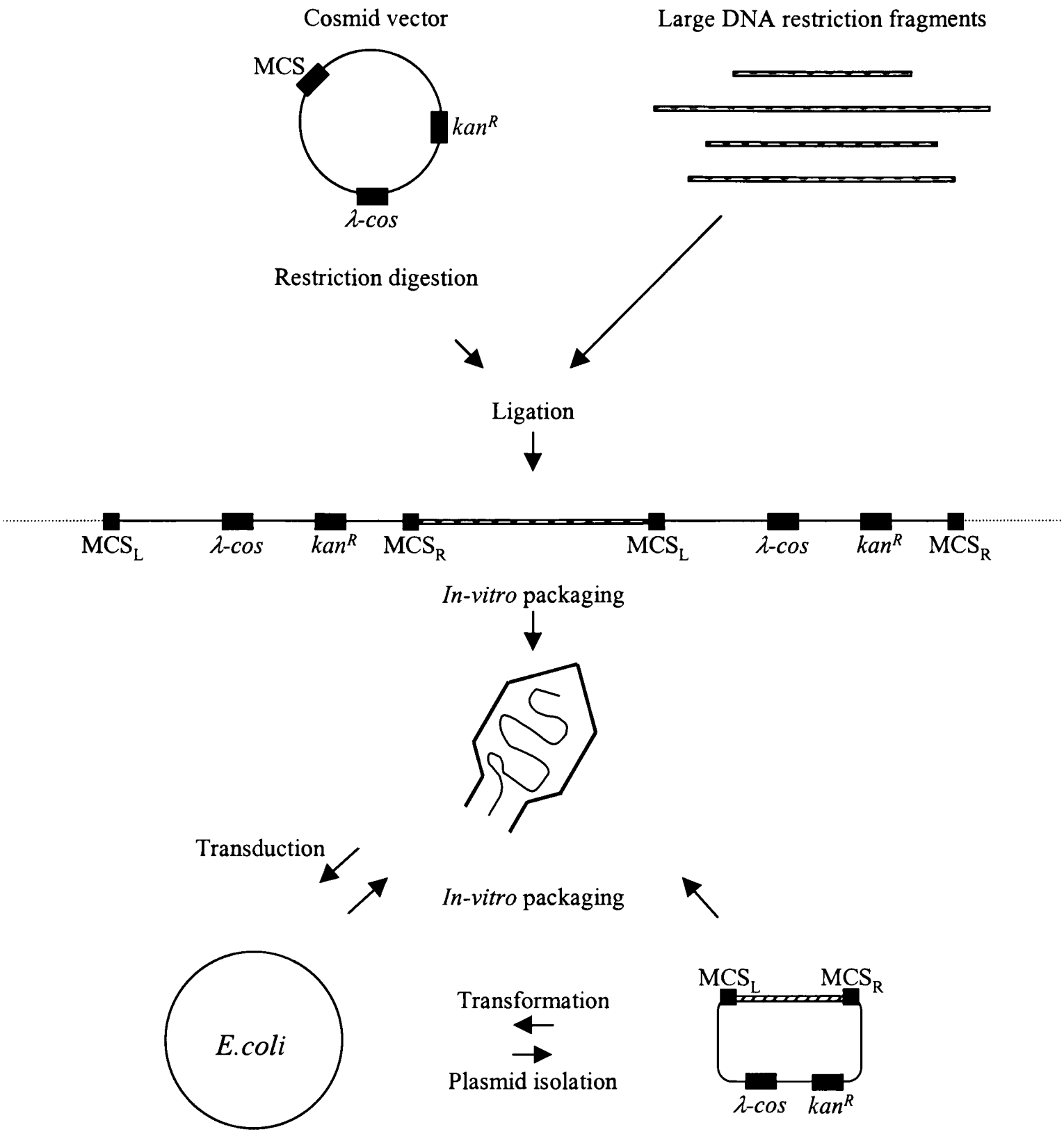


Figure 5.1 Cloning using Cosmid Vectors

The schematic shows the stages involved in the use of cosmid vectors to make DNA libraries. The main features of a cosmid vector are a λcos cleavage site, an antibiotic resistance cassette and a multiple cloning site (MCS) with unique restriction enzyme sites. In the construction of a library the vector is typically linearised by cutting in the MCS, but not de-phosphorylated. The DNA source from which inserts are to be generated is partially digested with an enzyme that produces fragments with compatible cohesive ends to the linearised vector. Ligation reactions are then carried out with the linearised vector and insert fragments under conditions which favour concatenation. The long fragments are then used as substrates in *in-vitro* packaging reactions in which λcos sites are recognised and cleaved to generate compatible cohesive ends. Fragments in the size range 30-45Kbp are packaged into the phage heads and these are then transduced into susceptible bacteria. The cosmid then propagates as a normal plasmid in the bacteria and can be purified by standard alkaline-lysis protocols.

ii.) PAC (P1 artificial chromosome libraries).

Large scale physical mapping of genomes was greatly facilitated by the development of Yeast artificial chromosomes (YACs) as a means of cloning and propagating large fragments of DNA (Burke et al., 1987; Larin et al., 1991). YACs are capable of carrying inserts up to approximately 1Mb in size but are prone to a number of problems which complicate their use. YACs are propagated in yeast which are not as easy to handle or culture as bacteria and purifying YAC DNA from yeast genomic DNA can be difficult. YACs from genomic DNA libraries can be chimeras containing two or more inserts of non-contiguous DNA which greatly complicates their use. The large size of YAC inserts and the frequency of DNA recombination in yeast mean that inserts can often be rearranged or contain deletions.

In response to the demand for more easily handled and stable systems for the cloning of large fragments of DNA the BAC (Bacterial Artificial Chromosome) (Shizuya et al., 1992) and P1 cloning system were developed (Sternberg, 1990). The BAC system uses derivatives of the E.coli fertility plasmid (F-factor) as a vector carrier for large DNA inserts. The F-plasmid copy number in bacterial cells is very tightly regulated at a low level reducing the probability of rearrangement by recombination. BACs are easily handled and isolated from bacterial genomic DNA and can be transferred directly into bacteria by electroporation. The main problem associated with BACs is the small quantity of BAC DNA isolated from bacterial cultures as a result of their low-copy number.

The P1-cloning system uses a very similar principle to classic bacteriophage vectors with the exception that an *in-vitro* packaging system is used to pack the vector DNA into phage heads allowing an increase in the amount of insert DNA that can be carried by the vector (70-100kb). These vectors have both high and low copy number P1 replicons allowing induction

of high-copy number for preparative isolation of the vector and low-copy number at other times to reduce the probability of recombination. Another advantage of the P1 system is that during the construction of libraries insert containing clones can be selected for over non-recombinants as the cloning site of the vector is positioned so that inserts ablate the expression of the lethal sacB gene (Pierce et al., 1992). The P1-system suffers in that the size of insert is limited by phage head size at approximately 100kb.

In order to take advantage of the best features of BAC and P1 cloning vectors the PAC (P1 Artificial Chromosome) cloning system was developed (Iaonnou et al., 1994). PAC vectors contain the high and low copy number replicons of previous P1-systems and the sacB gene driven by an *E.coli* promoter. The multiple cloning site of the vector is positioned between the promoter and the sacB gene to allow recombinants to be positively selected for by ablation of sacB activity (see Figure 5.2). PACs are transformed into bacteria by electroporation in the same way as developed for BAC vectors thereby simplifying library construction by eliminating the need to use *in-vitro* packaging systems.

PAC vectors are capable of carrying inserts of up to 300kb, but libraries are typically constructed with average insert sizes of approximately 150kb. They contain very small numbers of non-recombinants (typically 1-2%) and are highly stable. PACs are now widely used in large scale genomic mapping and are easily managed in small laboratories as the same culture and isolation protocols can be used as for traditional plasmid cloning vectors.

5.1.2 Methods for studying X-chromosome inactivation in the mouse.

Mice which carry Searle's translocation have been widely used to determine whether a gene escapes X-inactivation *in-vivo* (Brown et al., 1997). These mice have a balanced translocation

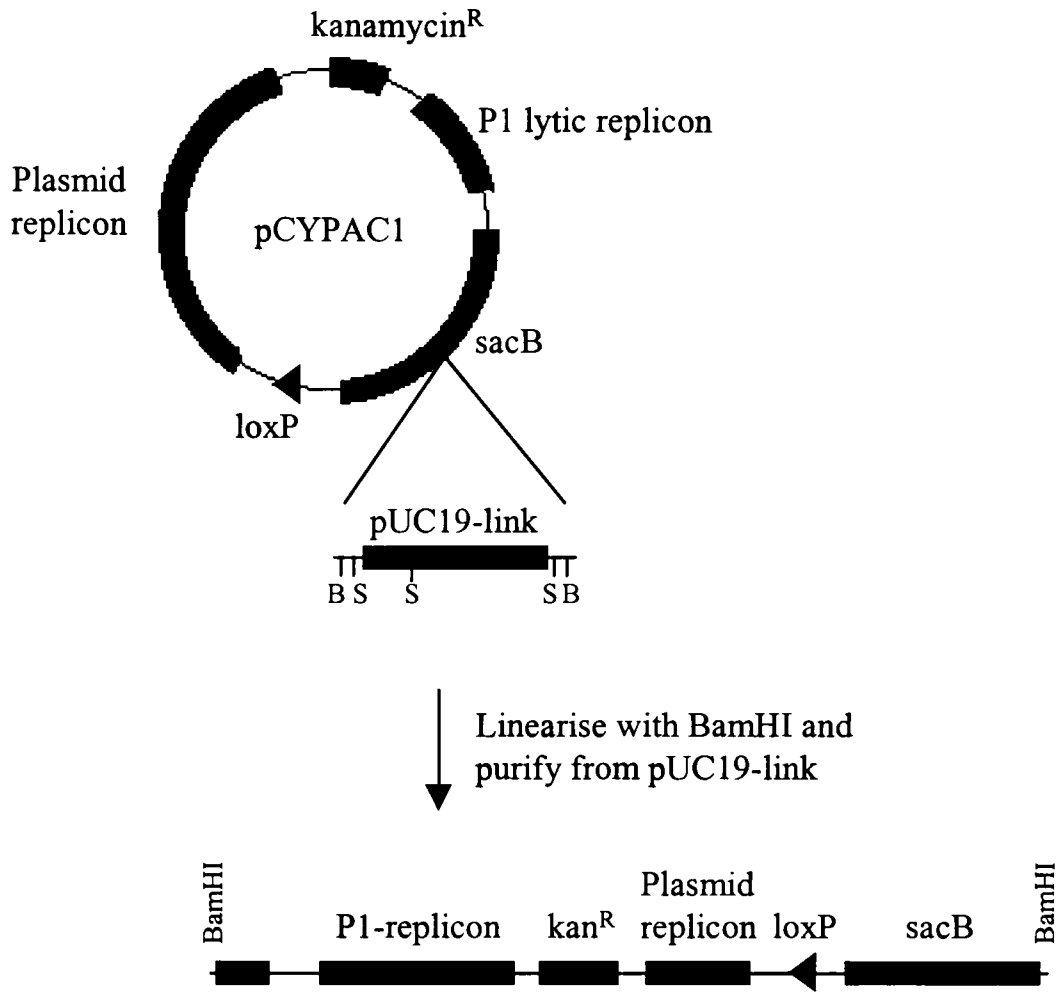


Figure 5.2 The PAC (P1 artificial chromosome) cloning system (after Ioannou et.al. 1994)
The PAC cloning system was developed to exploit the best features of both the BAC and P1-cloning systems (***see text***). The PAC system allows non-recombinants to be selected against because they express an uninterrupted lethal sacB gene. The presence of low copy number and inducible high copy number origins of replication mean that the PAC can be present in two replication modes. At low copy number recombination frequency is very low allowing stable maintenance of the PAC. The PAC is only induced to replicate to high copy number when DNA needs to be isolated for analysis.The other advantage is that PAC libraries are made by electroporation thereby avoiding complicated and expensive *in-vitro* packaging systems.

between chromosome 16 and the X-chromosome {T(16;X)16H} which results in selection so that only cells that have the intact X-chromosome inactive can survive (Disteche et al., 1981; Takagi, 1980). If female mice from one sub-species carrying Searle's translocation such as *Mus musculus* are bred onto a divergent sub-strain such as *Mus castaneus* then some of the F1-female mice will inherit the translocated chromosomes from *M. musculus* and an intact chromsome 16 and X-chromosome from *M. castaneus* (see Figure 5.3). The divergence between these two sub-species has resulted in the presence of many variants between their chromosomes. This facilitates the identification of polymorphisms in any gene and allows the quantitation of the expression of both alleles separately by allele specific RT-PCR.

Inter sub-species crosses generating F1-females with the Searle's translocated X-chromosome allows the quantitation of expression of both alleles of an X-linked gene in which a polymorphism has been found. Further, the knowledge that the paternally derived allele is always on the inactive X-chromosome allows the assignation of the expression level to either the active or inactive X-chromosome respectively. This system has been widely used to determine whether a gene escapes X-inactivation *in-vivo* and has also been used to quantitate the variation in the degree of escape in different tissues at different times during development (Disteche, 1995; Heard et al., 1997) (and references therein).

5.1.3 Methods for studying escape from X-chromosome inactivation in the human.

Determining whether a human gene escapes inactivation is very difficult as there is no *in-vivo* system available. Evidence that a human gene escapes inactivation is most frequently taken as the ability to generate an RT-PCR product for the gene using RNA isolated from human-rodent somatic cell hybrids in which the human inactive X-chromosome is retained (Brown et al., 1997).

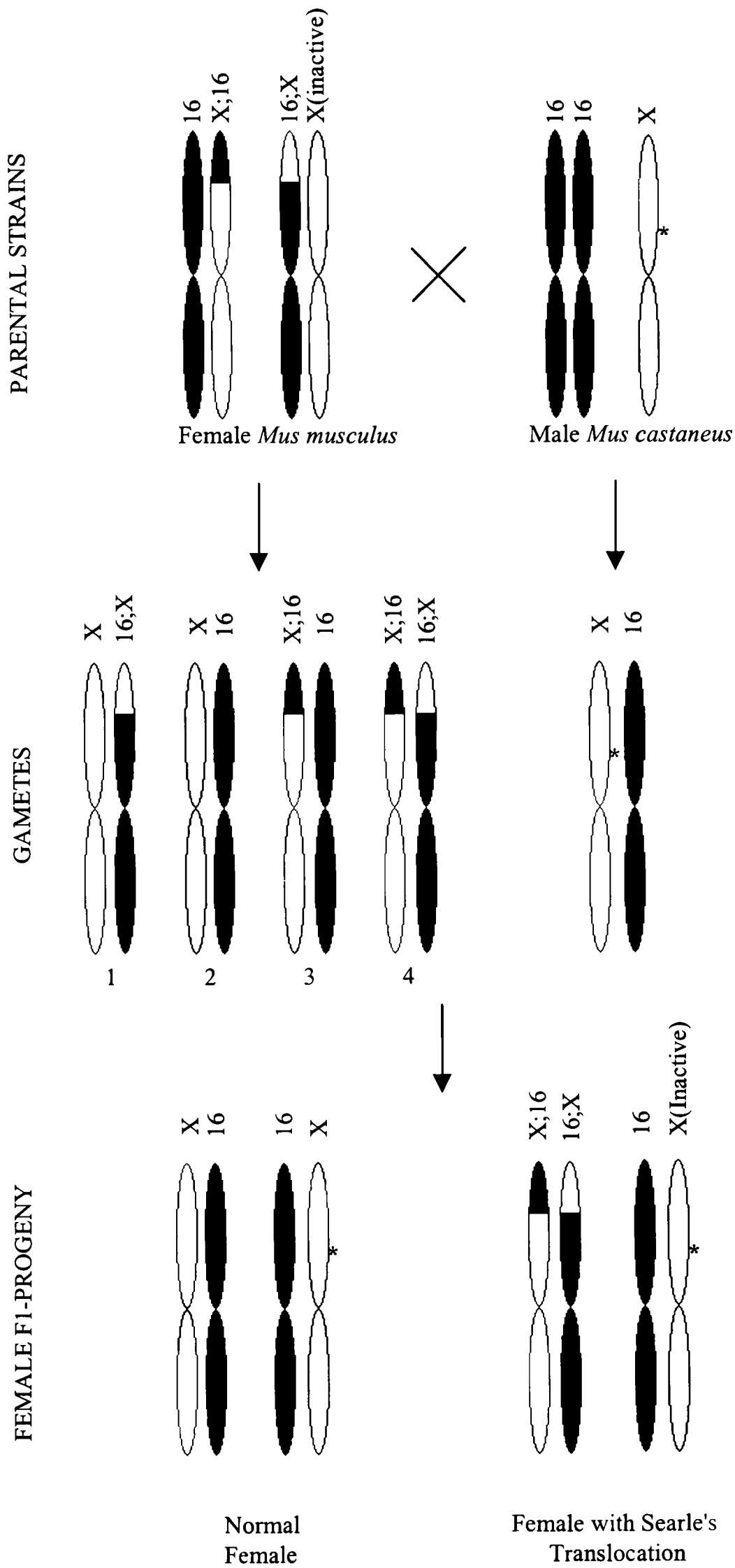


Figure 5.3: Generating mice to characterise genes escaping X-inactivation
Schematic illustrating how mice with many polymorphisms between parental chromosomes can be made carrying Searle's translocation to characterise a gene escaping X-inactivation *in-vivo* (***see text for explanation***).

Human-rodent somatic cell hybrids and radiation treated hybrids have been used widely in physical mapping approaches to localise genes to chromosomal and sub-chromosomal regions. Hybrids can be made by a variety of methods using either electro-fusion, infection with Sendai virus (Chang, 1989; Okada, 1993) or more commonly by treatment with polyethylene glycol (PEG) (McCullogh and Spier, 1990). The resulting heterokaryons are sub-cloned and cultured each clone losing different human chromosomes while retaining the full complement of rodent chromosomes. The same methods of generating somatic cell hybrids have been used by several groups with the aim of obtaining clones which retain either the human active or inactive X-chromosome. The resulting hybrids are inherently unstable and continue to lose chromosomes when cultured.

The use of somatic cell hybrids in the study of X-inactivation was greatly advanced by the development of techniques that selected for the retention of either the active or inactive X-chromosome. Murine cells deficient in the essential enzyme *HPRT* (hypoxanthine phosphoribosyl transferase) can be rescued by forming somatic cell hybrids with human cells expressing the *HPRT* gene. As the *HPRT* gene is X-linked and only expressed from the active X-chromosome hybrids can easily be generated which retain only the human active X-chromosome (Brown et al., 1997).

The more difficult problem was to find a method of selecting for somatic cell hybrids retaining the inactive X-chromosome. The answer came in the form of a mutant mouse cell line called tsA1S9 which had a temperature sensitive cell cycle defect (Brown et al., 1989). In an attempt to find the gene responsible for the phenotype human-mouse somatic cell hybrids were made to find out which chromosome(s) complemented the mutation. Surprisingly both

the active and inactive X-chromosomes were capable of rescuing the temperature sensitive mutation. This discovery led to the cloning of the human X-linked A1S9T gene which turned out to be identical to the ubiquitin activating enzyme E1 (*UBE1*). As the inactive X-chromosome was capable of complementing the tsA1S9 cell line the *UBE1* gene was predicted to escape X-inactivation. Several lines of evidence confirm that *UBE1* escapes X-inactivation including RNA *in-situ*, somatic cell-hybrid and methylation studies (Carrel et al., 1996).

This *UBE1* complementation system is now routinely used to generate and maintain human-rodent somatic cell hybrids retaining the human inactive X-chromosome as the only human material. The ability to generate human-rodent somatic cell hybrids retaining either the active or inactive X-chromosome has allowed several groups to develop subtraction based approaches to identify novel expressed sequences as candidates for genes that escape X-inactivation (Ellison et al., 1992; Esposito et al., 1997; Sultana et al., 1995). These two kinds of somatic cell hybrid are also routinely used to test the inactivation status of any novel cloned genes from the X-chromosome by RT-PCR. The human genome sequencing project will eventually allow systematic screening of the X-inactivation status of all genes by RT-PCR using this system.

5.1.4 The *SMCX/Smcx* gene.

The *SMCX/Smcx* gene is one of only three non pseudo-autosomal region (PAR) genes known to escape X-chromosome inactivation in man and mouse (Disteche, 1995; Distesch, 1997). As such it represents a rare target for studying local promoter features that might be required to escape the inactivation process.

The *Smcx* gene was isolated by screening a mouse testis cDNA library with a probe derived from the *Smcy* gene based on the assumption that the widely expressed Y-linked gene would have an X-linked homologue that escaped X-inactivation (Agulnik et al., 1994). The human *SMCX* gene was isolated as part of a systematic investigation to isolate genes which escaped X-inactivation. Human cDNA clones were isolated from human-rodent somatic cell hybrids whose only human content was the inactive X-chromosome (Wu et al., 1994). The X-inactivation status of the human and murine genes was investigated by methods described below. The gene continues to be known under a variety of names including *SMCX/Smcx*, *XE169/Xe169* and *DXS4274E*. This thesis refers to the gene as *SMCX* or *Smcx* for consistency.

Expression analysis of the *Smcx* gene has been carried out by several groups using the Searle's translocation system (see section 5.1.2) (Carrel et al., 1996; Sheardown et al., 1996). The *Smcx* gene is expressed in many tissues including the brain, heart, kidney, lung, liver and spleen. The level of escape of the *Smcx* gene was found to be variable by tissue type with the greatest escape observed in the heart. Overall, levels of escape from inactivation varied between approximately 30-65% of the level of expression from the active X-chromosome. It has been proposed that the variable levels of escape could reflect differences in the timing of X-inactivation in different cell lineages or tissue specific differences in the regulation of the X-inactivation process (Carrel et al., 1996; Sheardown et al., 1996). Every gene known to escape X-inactivation to date has been show to have a hypo- or un-methylated promoter on the active and inactive X-chromosome (Brown et al., 1997). It is tempting to speculate that the variable level of escape observed for the *Smcx* gene is as a result of differential methylation at its promoter by tissue.

The *SMCX* gene has been shown to escape inactivation in human-rodent somatic cell hybrids retaining the inactive human X-chromosome (see section 5.1.3) (Agulnik et al., 1994; Wu et al., 1994). Like its murine homologue, the *SMCX* gene is expressed in many tissues including brain, heart, kidney, lung, liver, muscle and pancreas.

Screening the Genbank genomic database with the full-length *SMCX* cDNA sequence reveals that it is highly similar to the Retinoblastoma binding protein-2 (RBBP2) gene sharing on average 52% identity at the amino acid level. In addition the protein has identity with zinc-finger (38%) and chromodomain (44%) containing proteins. It has been suggested that *SMCX/Smcx* is a transcription factor although this has not been demonstrated experimentally. *SMCX* and *SMCY* have 87% nucleotide identity, 95% amino acid identity and 100% amino acid similarity suggesting that they are functionally equivalent. If this is the case then the widely expressed nature of *SMCY/Smcy* would imply that *SMCX/Smcx* should also be widely expressed if dosage equivalence is to be achieved between males and females.

5.2 Results.

5.2.1 Generation of the human *SMCX* cDNA probe.

The entire cDNA sequence for the human *SMCX* gene has been isolated and submitted to the Genbank genome database under accession number L25270 (Wu et al., 1994). In order to isolate genomic fragments containing the 5'end of the *SMCX* gene a probe was generated to screen the human LLNL X-only cosmid library (Materials 2.11). Total RNA was isolated from two different fibroblast tissue culture cell lines and first strand cDNA prepared from them as previously described (Methods 2.19 and 2.45). First-strand cDNA was used in PCR reactions with the primers XE169F1 (5'-CCGGGGTGGTACTTTTGAG-3') and XE169R1 (5'-TACTGCTACCGCCTCTTCG-3') to amplify a 311bp fragment of the 5' end of the human cDNA

(+41 to +360bp) using standard PCR conditions (Figure 5.4A). The PCR products were *PfU* "polished" and blunt-cloned into the *Eco RV* site of the vector pZErO2 as described previously. The ligation products were transformed into electro-competent TOP10 *E.coli*, which were then plated onto selective plates and incubated overnight at 37°C. Surviving colonies were screened by colony PCR to identify those with correct sized inserts. One such clone, Xe169c1 was then sequenced to confirm the identity of the insert for use as an XE169 probe (data not shown).

5.2.2 Screening the Human LLNL X-only cosmid library.

The plasmid clone Xe169c1 was digested with the enzymes *Eco RI* and *Xho I* to excise the probe insert (Figure 5.4B). This fragment was separated from the vector backbone by electrophoresis in a 1% (w/v) agarose gel, excised from the gel and purified prior to radioactive multiprime labelling (see Methods 2.30). The radiolabelled probe was denatured and hybridised to the LLNL cosmid library filters overnight at 65°C in Church buffer (see Methods 2.32). Post-hybridisation, the filters were washed to a low stringency and then exposed to X-OMAT film at -70°C overnight (see Methods 2.33). Photographic films were developed manually to a suitable intensity and contrast. A total of 8 strong positives were obtained and 3 of these (U30E12, U34J10 and U41D15) were ordered from the HGMP-RC for further analysis. Figure 5.5 shows a typical auto-radiograph for the cosmid library screen with the positive representing the cosmid clone U34J10 identified by an arrow.

DNA was prepared from the cosmid clones (see Methods 2.20) and used as a template in PCR reactions with the primers XE169F1 and XE169R1 to confirm that the clones contained the 5' coding region of the *SMCX* gene. Figure 5.6 shows the results of the PCR reactions confirming that all three genomic clones contain the desired genomic interval. In addition the

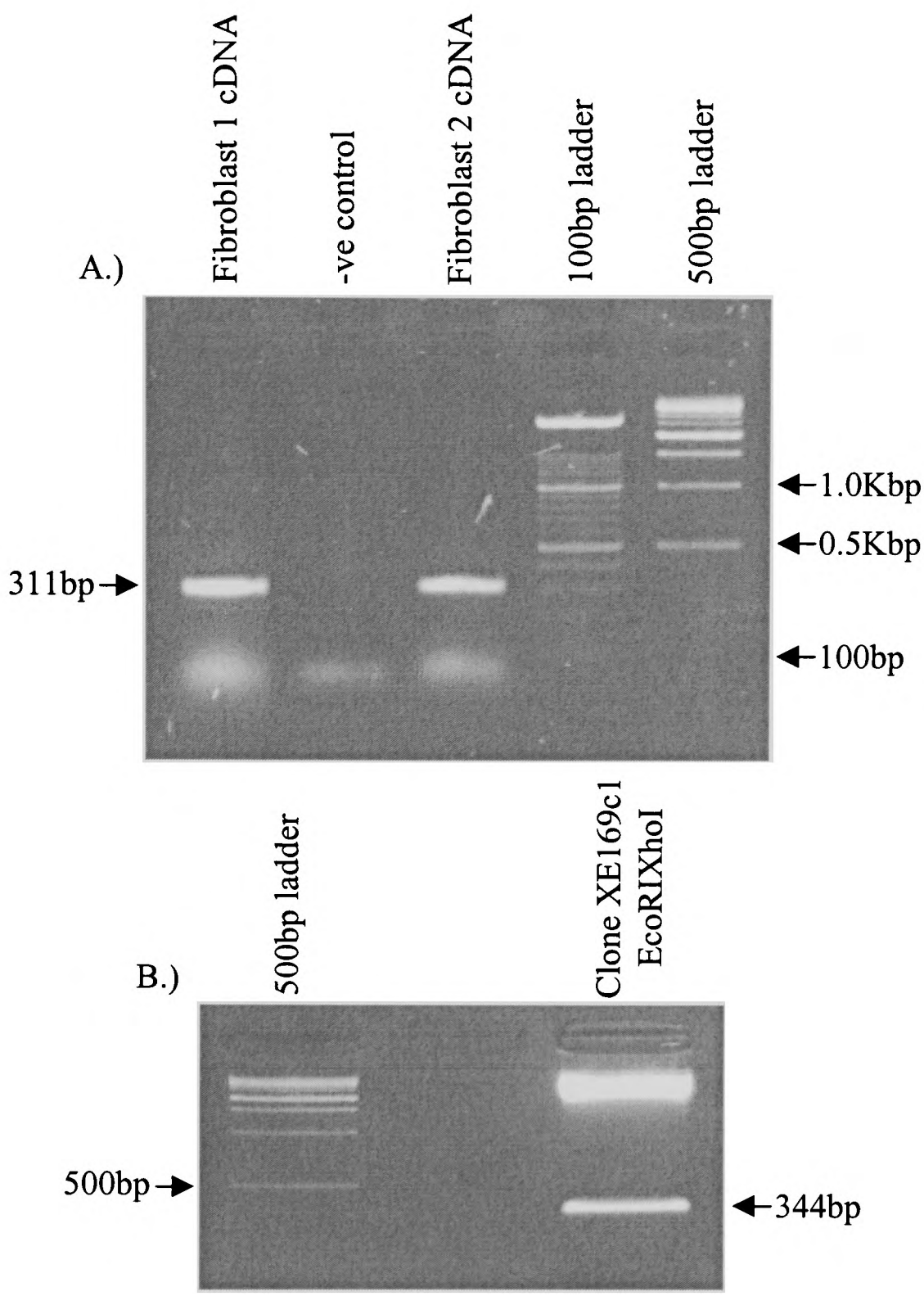


Figure 5.4: Generation of the human SMCX cDNA probe.
A probe containing the 5' end of the Human SMCX cDNA was made by RT-PCR. Total RNA was prepared from tissue culture fibroblasts and used as a substrate for first strand synthesis of cDNA. The primers XE169F1 (5'-CCGGGGTGGTACTTTTGAG-3') and XE169R1 (5'-TACTGCTACCGCCTCTTCG-3') were used to amplify a 311bp product (A) which was then blunt-end cloned into the EcoRV site of the vector pZErO-2. The clone XE169c1 was isolated and sequenced to confirm that it contained the correct insert. Probe for use in screening libraries was prepared from XE169c1 miniprep DNA by gel extraction of a 344bp EcoRI/XhoI fragment (B). The probe was used to screen both the Human LLNL X-only cosmid and Mouse PAC RPCI21 libraries to isolate clones containing the promoters of the SMCX/Smcx genes.

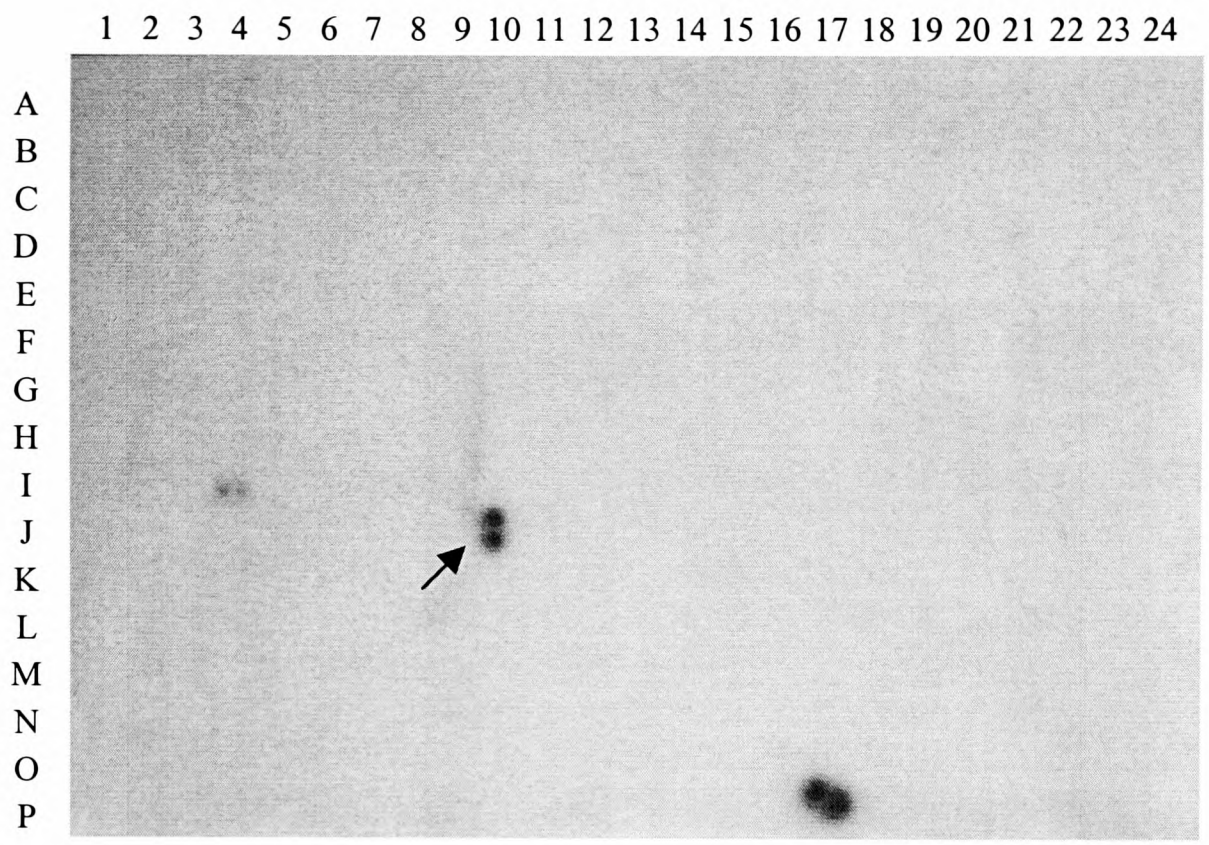


Figure 5.5: Screening of the Human LLNL X-only cosmid library with the probe Xe169c1
The figure shows two strong positives on an auto-radiograph of an Xe168c1 probed library filter. Clones are gridded in oriented double spots on the library filters to aid identification. The cosmid clone U34J10 is indicated with an arrow and was one of three clones ordered from the HGMP-RC for further analysis.

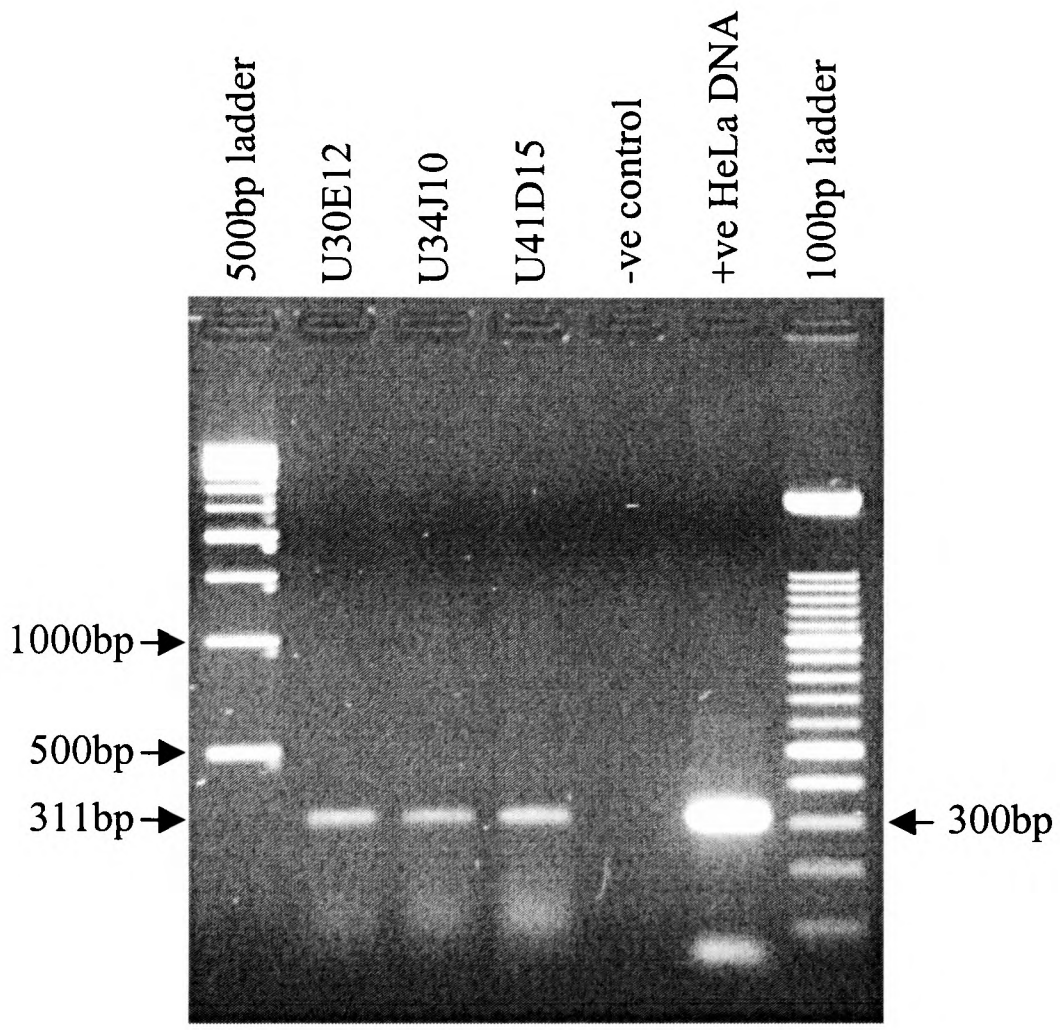


Figure 5.6: PCR screening of cosmid clone DNA for SMCX content. The cosmid clones U30E12, U34J10 and U41D15 were used in colony PCR reactions with the primers XE169F1 and XE169R1 which were used to make the Xe169c1 5'-cDNA probe to ascertain whether any of them contained those parts of the SMCX gene. The figure shows that all three clones and a sample of human genomic DNA all produce the same 311bp PCR product. This demonstrates not only that all three cosmids were correctly isolated as being Xe169c1 positive, but also that the probe is positioned within an exon.

observation that the band produced from the genomic DNA clones is the same length as the Xe169c1 probe (approx. 300bp) demonstrates that the probe is entirely contained within one exon.

Cosmid DNA from clones U30E12, U34J10 and U41D15 was digested with the restriction enzymes *Bam HI, Eco RI* or *Hind III* and separated by electrophoresis at 1.5V/cm on a 0.8% (w/v) agarose gel to achieve good resolution of fragments between 2-5kb (Figure 5.7). The digest fragments were transferred to nylon filters by alkali transfer Southern blotting (Ref.) as described (see Methods 2.29.2) and the filters probed with the Xe169c1 probe. The filters were washed down to a background count of approximately 10cps and then exposed to X-OMAT film at -70°C overnight. Post-exposure, the films were developed manually to suitable intensity and contrast. Figure 5.8 shows the auto-radiograph of the hybridisation in which all three clones produce identically sized positive bands. The most conveniently sized fragments for sub-cloning are the 1.3kb *Bam HI* and 2.6kb *Eco RI* fragments (approximate sizing). There is only one positive band in the *Eco RI* lane of each clone suggesting that the entire Xe169c1 probe is contained within the 2.6kb fragment.

The 2.6kb *Eco RI* fragment was excised from a preparative gel of *Eco RI* digested DNA prepared from the cosmid clone U34J10 and purified by gel extraction. Approximately 10ng of this purified fragment was sub-cloned into the *Eco RI* site of the vector pZErO2 then transformed into TOP10 cells by electroporation. After culture on selective plates, bacterial clones were picked and screened by colony PCR using the primers XE169F1 and XE169R1 to see whether any contained the 311bp fragment derived from the 5'end of the *SMCX* gene. Figure 5.9A shows the results of the PCRs demonstrating that clones XE01 and XE03 produced a band of the correct size. Plasmid DNA from these clones was digested with *Eco*

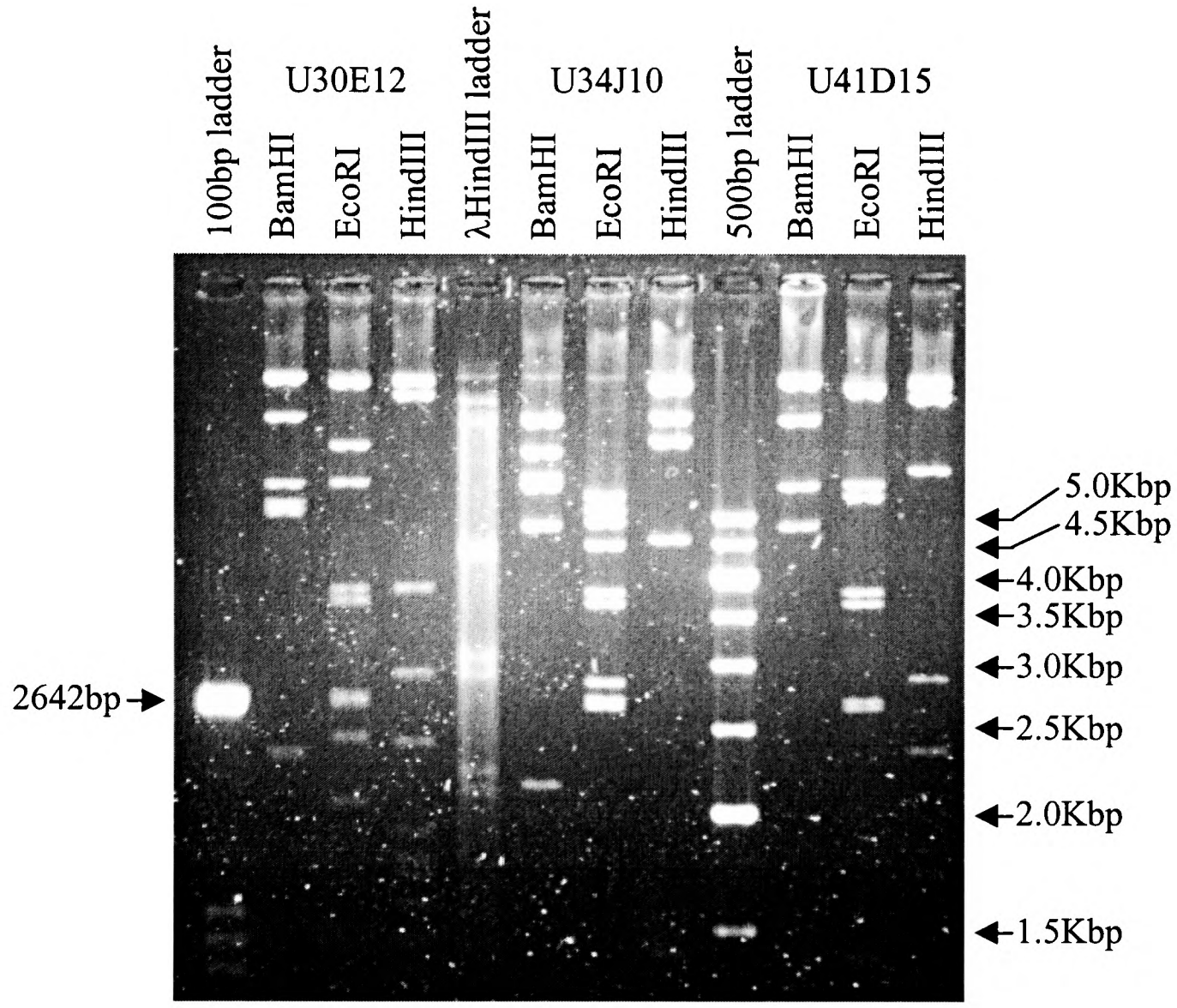


Figure 5.7: Restriction digests of Xe169c1 positive cosmid clones
DNA was prepared from the three Xe169c1 positive cosmid clones by a standard alkaline-lysis protocol and then digested with the restriction enzymes BamHI, EcoRI and HindIII. The fragments were separated by electrophoresis in a 0.8% (w/v) agarose gel at 1.5V/cm for a length of time suitable to achieve good resolution of fragments in the 2-5Kbp range (approx. 2.5 hours). Notably all three cosmids produce different restriction fragment patterns suggesting that they are non-identical overlapping clones. This gel was subsequently blotted onto nylon filters and probed with the Xe169c1 probe (see Figure 5.8).

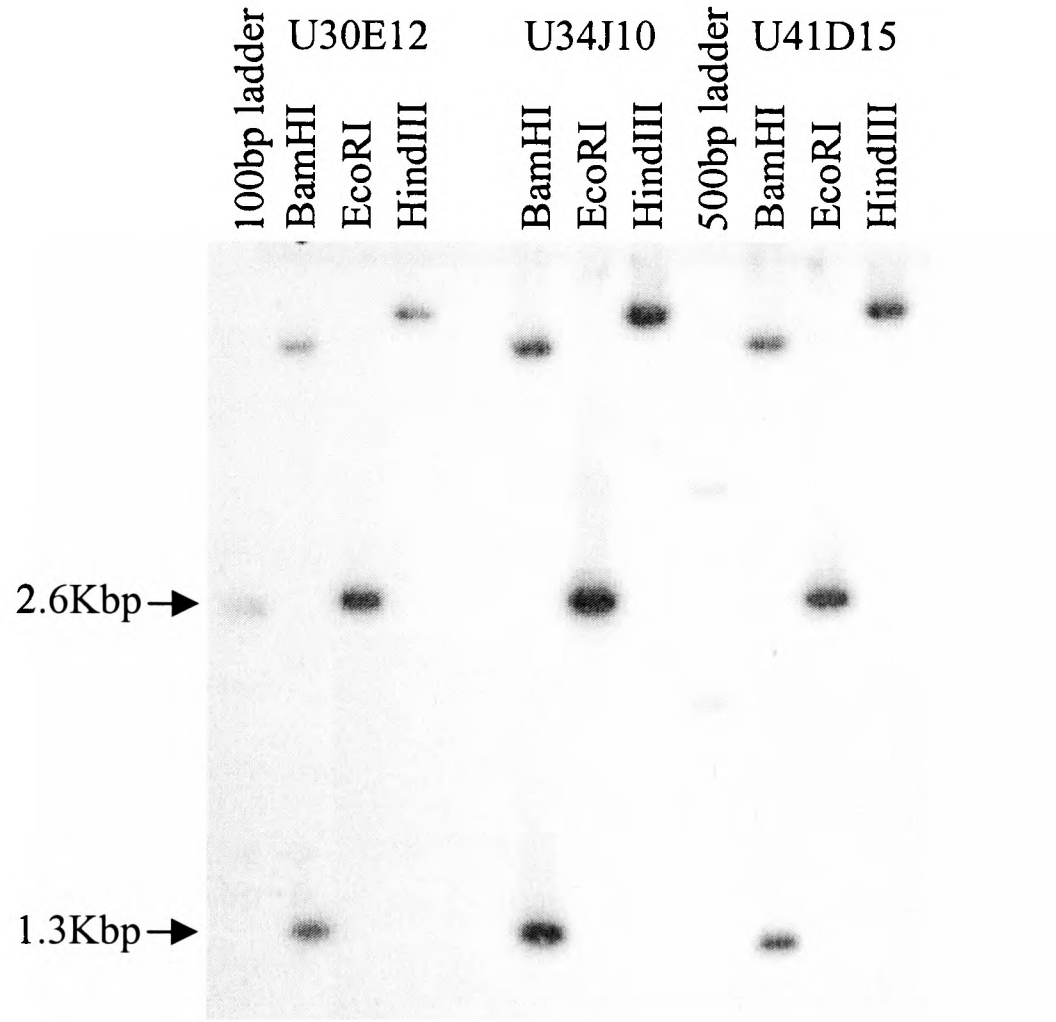


Figure 5.8: Auto-radiograph identifying the Xe169c1 positive fragments of the cosmid clones U30E12, U34J10 and U41D15.
Restriction digests of DNA isolated from the three cosmid clones was separated by agarose electrophoresis (see Figure5.7) and then Southern blotted onto nylon filters by alkaline transfer (see Methods 2.29.2). The nylon filter was probed with radio-labeled Xe169c1 probe to identify fragments of the cosmid digests that contained the 5'-end of the SMCX gene. The filter was washed to a background count of approx. 10cps and then exposed to X-OMAT AR film overnight at -70°C. The film was developed manually to suitable contrast and intensity. Each clone is seen to produce identically sized Xe169c1 fragments. Two bands are present in the BamHI lane, one high molecular band and one at 1.3Kbp.The EcoRI lane has one 2.6Kbp band and the HindIII lane has one high molecular weight band. As only one Xe169c1 positive band was found in the EcoRI lane and it was of suitable size to clone it was selected as the target fragment for cloning the 5'end of the SMCX gene.

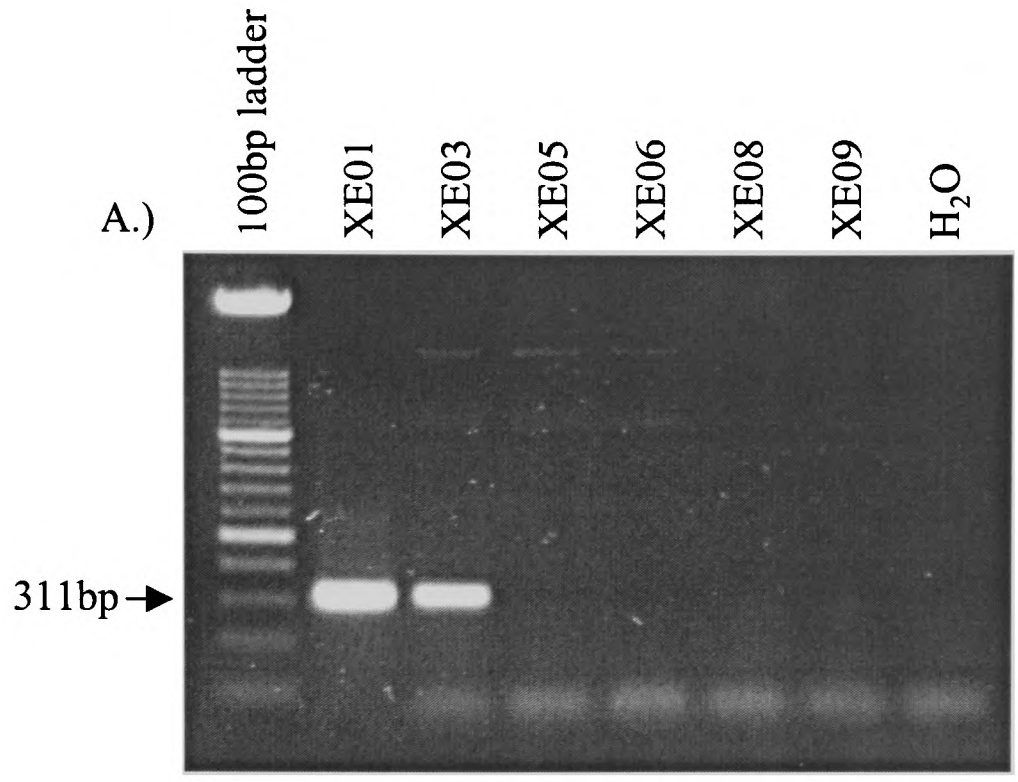


Figure 5.9: Sub-cloning Xe169c1 positive fragments of the cosmid clone U34J10
The 2.6Kbp Xe169c1 positive EcoRI fragment identified by hybridisation was selected to be the fragment sub-cloned to isolate a manageable clone for sequencing of the human SMCX proximal upstream region. A preparative digest of U34J10 cosmid DNA was made with EcoRI and the fragments separated by agarose gel electrophoresis. The 2.6Kbp fragment was excised from a gel, purified and used in a ligation reaction with EcoRI pre-cut pZeRO2 zero-background cloning vector. The ligation products were transformed into competent TOP10 strain *E.coli* by electroporation and vector containing clones selected for by culture on Kanamycin containing plates overnight. Surviving clones were screened by colony PCR with the primers XE169F1 and XE169R1 to identify those containing the 311bp Xe168c1 genomic interval. Two clones, XE01 and XE03 were found to produce PCR products of the correct size. In order to determine whether any of the clones contained an insert of approx. 2.6Kbp miniprep DNA was prepared from overnight cultures of several clones and digested with EcoRI. Separation of the restriction fragments by electrophoresis demonstrated that only clone XE01 contained an insert of the correct size (B).

RI and clone XE01 was found to have an insert of 2.6kb (see Figure 5.9B) consistent with it containing the desired genomic fragment. Clone XE01 plasmid DNA was used as template in PCR reactions to determine the orientation of the coding section within the insert. This was achieved by using the primer sets XE169F1/T7 and SP6/XE169R1 which determine the orientation with respect to either the SP6 or T7 primer sites which are at opposite ends of the multiple cloning site (Figure 5.10C). The PCR reactions produced a product with the XE169F1/T7 of 1.6kb (Figure 5.10A) and no product was obtained with the SP6/XE169R1 set. The knowledge of the position of a *Bam HI* site within the Xe169c1 probe made it possible to predict that there would be approximately 1.0kb of upstream sequence in this clone (Figure 5.10C).

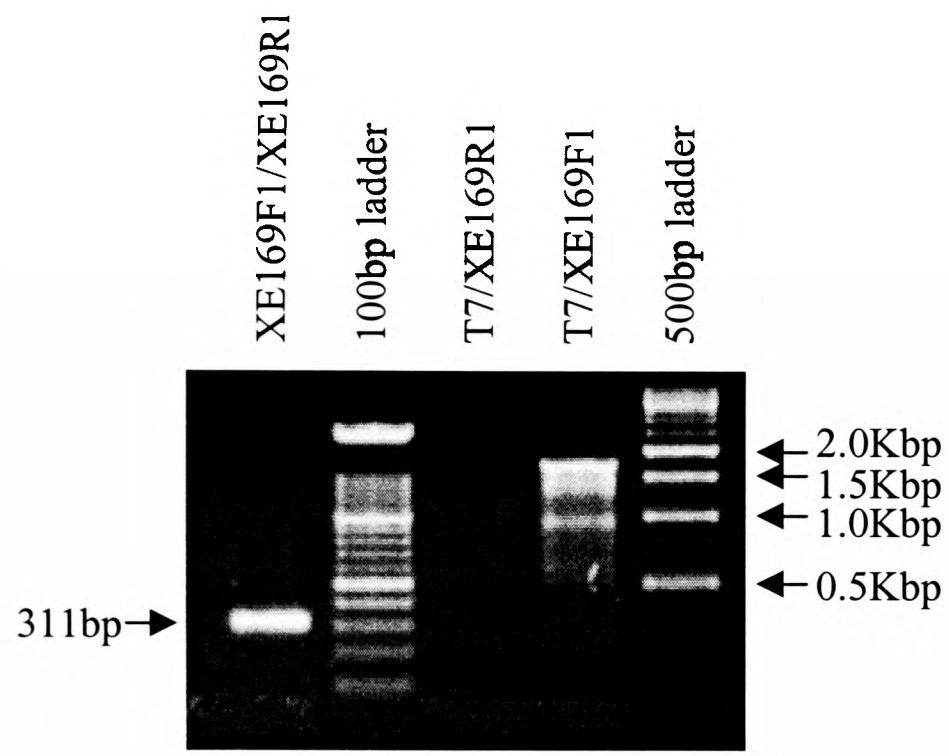
A 1.2kb *Bam HI/Eco RI* fragment containing *SMCX* upstream sequence was sub-cloned from the XE01 clone into the vector pZErO2 and two independent clones containing the correct sized insert (XEBE6 and XEBE9) were sequenced by fluorescent cycle sequencing using the primers SP6 and T7 (see Methods 2.42).

The entire nucleotide sequence of the human promoter XEBE clones is shown in Figure 5.11 with the first exon and the position of the Xe169c1 probe fragment indicated. A total of 1010bp of upstream sequence was obtained from these clones, further upstream sequence could be isolated in the future from the cosmid clone U34J10.

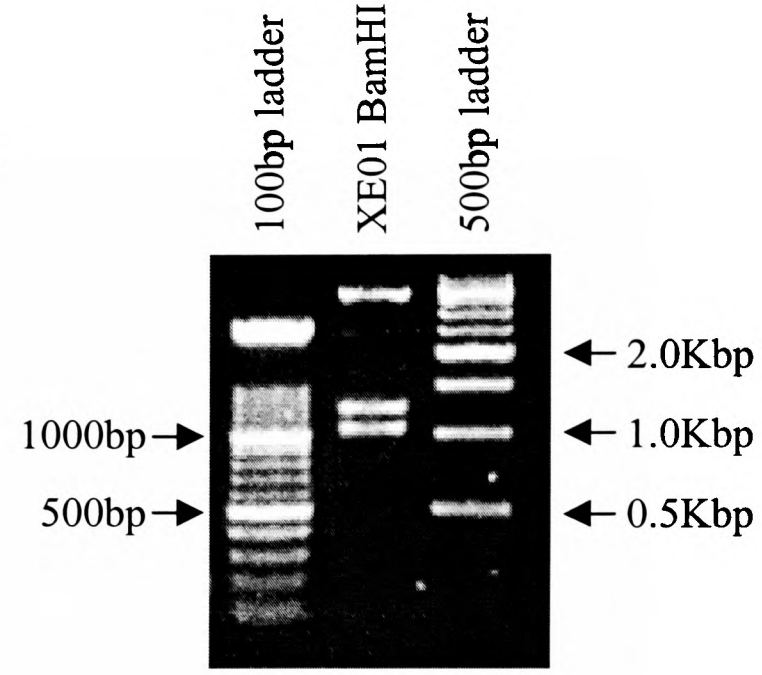
5.2.3 Analysis of the Human *SMCX* upstream region.

The human upstream region was analysed for potential sites of methylation and CpG-islands as previously described for the *PDHA1* (Chapter 3), *NDP* and *Ndp* (Chapter 4) gene promoters. The human *SMCX* upstream region was found to contain no CpG-islands, but

A.)



B.)



C.)

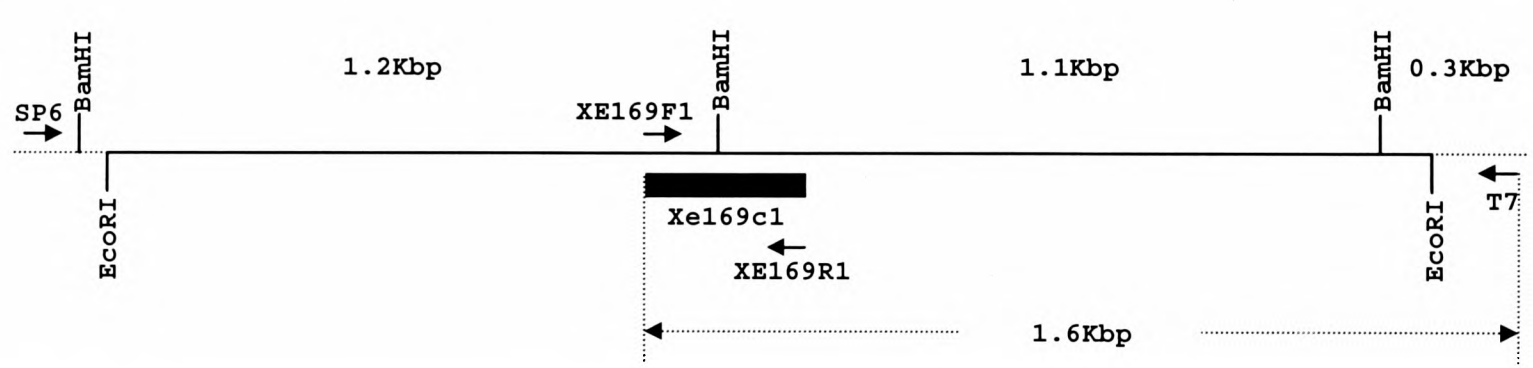


Figure 5.10: Sub-cloning the upstream fragment of the human SMCX gene from cloneXE01.
In order to determine the orientation of the SMCX coding sequence in the XE01 clone a set of PCR reactions were performed on diluted miniprep DNA using the primer pairs XE169F1/XE169R1, T7/XE169R1 and T7/XE169F1. The first PCR confirmed again that the XE01 clone contained the 311bp fragment identical to that of probe Xe169c1. Only the T7/XE169F1 produced a product of the remaining two PCRs indicating that the coding sequence was oriented as shown in (C). A digest of the miniprep DNA with BamHI showed that the clone was bisected by a BamHI site producing fragments of 1.1 and 1.2Kbp (B). The fragments could be ordered because the position of the BamHI site in the SMCX cDNA is in the published sequence and the T7/XE169F1 PCR produced a product of 1.6Kbp. All of these experiments led to the conclusion that the clone XE01 should contain approximately 1Kbp of sequence upstream of the published start of the SMCX cDNA.

EcoRI

GAATTCATAGTCTGATGGGGGAAAACAAAATTACTTAATCAGCATGTACGCAAGTACTCA
AATATTTGCACAGTAGGTGTTCAAACTATATGCAGTAGGTCCACAGCTGATGAATGACAC
AGATGATCCTCCAATCAGTAGATGCCTGATGTATATGTACAAAGTACTGGTTTGATGATA
CTCATTCAAAGTACTGTAGTTGCAATGAACCTATTCAGAAACTGCTCAAGGAACATGCAC
ACGATAGGACTAAATGTACATACAGTTGGTACCAAATGCACAGGGGCTATACACAATAAA
TACAATCAAAGTACACGGATGTTTGGTAAGTGCACTGTAGGCTCACAAATGGGCAAACAG
GTGTTTTACACAGTTGACACTGGATAAATACGCCGTTGCTAAACATCCAATAGTTAAAGT
ATACATACACACAGGCTTATTTTGTAGGAGTATATGCCACAAATGTGTACATGGTAAAAG
TGAATACAGAACAGTTGAAAGCGTTCATTAAAATGCTGACTCCCACGTGACTTTTGGGGG
GTAAGGAATCGAAACCCCATCTTCTCCTACCCGTAACCTTGGATGAAATAAAGATTTCCT
TCCTTTACTTGGGGACTCAGAAATCCCTAAAAAATCCATAACCCCCGGGGGGAGGTGAGC
AGCTTCCCTGCCTTCCTTTATTTCAAAGCCCCGAACGTTGTCAGGTGTCCTAAAGGCTGG
AGCCTAGTACATCTCCCCCCATCCCGCGGGCCAGCATAAGGGGTCCTTTGCAATCAGGCT
CCCAGAGCCGCCTTCCGCGGCCCATACCGTTAAGACCAGGGAGAAGGGACAGAGAACACT
GGTGCGGGGCTAGCGACCTGGTTGAGGTCTGCGGGGGAGTGCCGGTTTGACTCCACCACC
CACGGACACCAGCTCCTCTTCCGGTCTCCATAACGCGCATGCGCATCACTGGCCTTTAAA
CGTGCACCGCCGGTCCATCCGGAAAGACGATCCGGCAAACTAATTACAAT**CTTGTTCCTC**
CGCCGTTGCAATGAACTATTTTCTCTCAGTCCGGGGTGGTACTTTTGAGTAACCCCTTCC
AAATAGAAACCCATACCAGCCTATTTAGCTCGGTCTCCACTATATGAAGGTTTCCTAGGG
CTTGGTGTGACGCAACGTATACGAGGCTCGGAAGGACACCCCGCGGAAGGATCC

BamHI

Figure 5.11: Nucleotide sequence of the human XE169 *Bam HI/Eco RI* promoter clones XEBE6 and 9.
The most efficient way to obtain upstream sequence from the clone XE01 once it had been characterised (Figure 5.10) was to sub-clone the 1.2Kbp EcoRI/BamHI fragment from it and then to sequence the resulting clone from both ends. The 1.2Kbp fragment was easily sub-cloned from cloneXE01 as it was the only fragment near the size that had the correct cohesive ends. Purification of the band from a preparative restriction digest and subsequent ligation into prepared pZeRO2 vector produced many clones containing the correct sized insert (data not shown). Two of these clones XEBE6 and XEBE9 were sequenced entirely from both ends by the vector multiple cloning site flanking primers SP6 and T7. The resulting sequence contained 1010bp of upstream sequence and as expected the entire of the Xe169c1 probe sequence (**underlined**). SMCX cDNA sequence is shown in (**bold**), restriction enzyme sites are boxed.

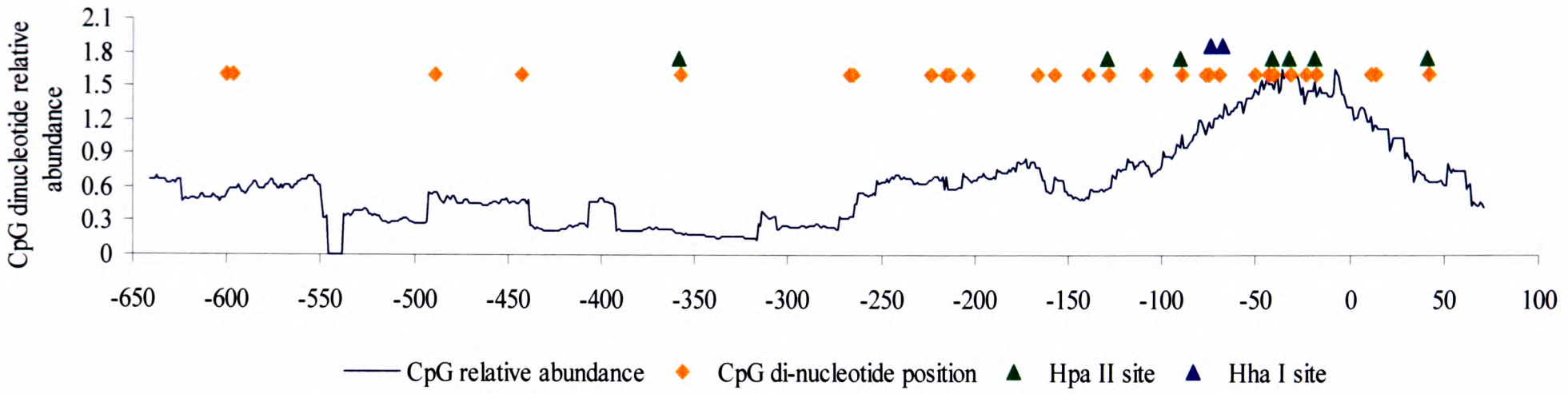
contains two continuous sections which satisfy the "island criteria" (see Chapter 4 section 4.2.1) from –262 to –145bp (117bp) and –140 to –19bp (121bp) (Figure 5.12). The section from –262 to –19bp (243bp) narrowly misses classification as a CpG-island, an additional cytosine or guanine residue anywhere within 50bp either side of position –142bp would allow this section to qualify as a CpG-island. This illustrates the limitations of the definition of a CpG-island to one which serves as a tool for gross classification applied across large sections of DNA. It is not well suited to the detailed analysis of individual genomic loci.

Analysis of this promoter sequence reveals 43 CpG di-nucleotides, 2 *Hpa II* and 7 *Hha I* sites with 70% of CpG sites concentrated within 250bp of the transcription start site. Initial analysis of the upstream sequence revealed no putative TATA or CAAT boxes, but indicated that several Sp1 sites were present suggesting that this gene may be a housekeeping or widely expressed gene. The positions of putative transcription factor binding sites in the human *SMCX* promoter are detailed in comparison with the murine *Smcx* promoter in section 5.2.6.

5.2.4 Screening the Murine RPCI21 genomic PAC library.

The comparative study of gene promoters between man and mouse can reveal highly conserved sections of their promoters and help to determine the positions of functional cis-acting elements and methylatable sites.

The murine RPCI21 genomic PAC library (see Materials 2.11) was screened with the human Xe169c1 cDNA probe to isolate genomic PAC clones containing the 5'-end of the murine *Smcx* gene. The human cDNA probe was used because only the 3'-end of the murine cDNA has been cloned and published (Genbank accession number L29563) (Wu et al., 1994). Cross-hybridisation between species can be an extremely effective way of isolating



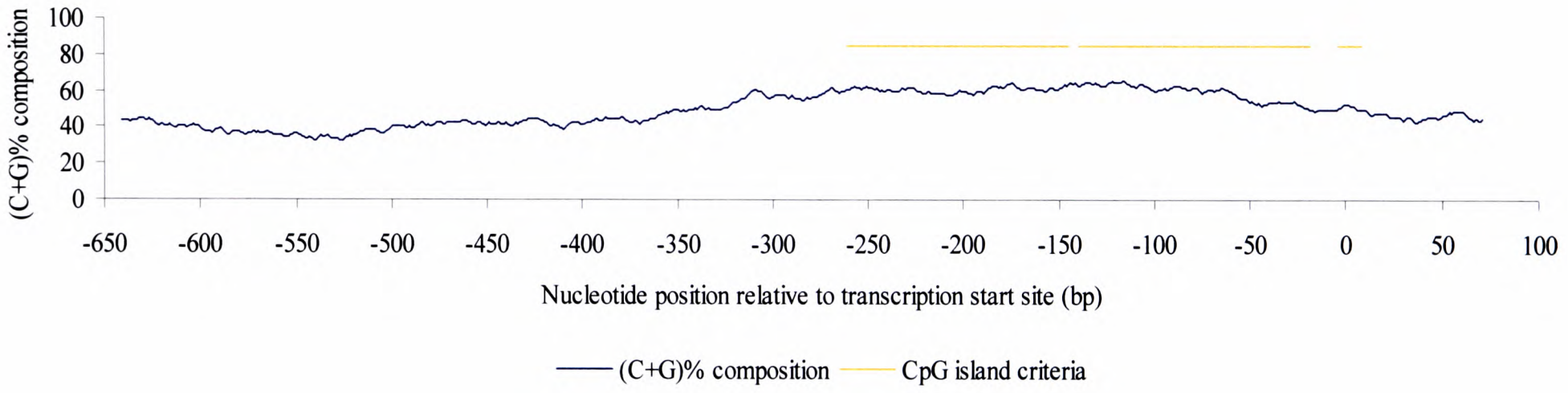


Figure 5.12: CpG-Plot of the Human SMCX promoter region.
The GCG-Wisconsin program CpG-Plot was used to generate co-ordinates for graphs of CpG di-nucleotide relative abundance (RA) and (C+G)% content against nucleotide position for the SMCX upstream region. The upper graph shows the CpG RA and is annotated with the positions of individual CpG sites, the positions of HpaII and HhaI sites. The lower graph shows the (C+G%) content and an assessment of whether the sequence contains a CpG-island. The points which satisfy the island criteria are shown in orange. If these criteria are satisfied by a continuous stretch of 200bp then the section qualifies as a CpG-island. The section displayed here contains 43 CpG di-nucleotides and narrowly misses classification as a CpG-island. The proximal region is however, clearly CG-rich with 70% of CpG sites clustered within the first 250bp of the transcription start site.

homologous genes based on their high similarity at the nucleotide level. The RPCI21 library was particularly useful for the isolation of the *Smcx* gene because it was constructed from female 129/SvevTACfBR mouse genomic DNA and therefore contained no Y-chromosome sequences precluding the possibility of isolating clones of the highly homologous Y-linked *Smcy* gene (Agulnik et al., 1994).

The Xe169c1 probe was radio-labelled prior to hybridisation then denatured and hybridised to the seven library filters overnight at 65°C (see Methods 2.32). Post-hybridisation, the filters were washed to a low stringency and then exposed to X-OMAT film at -70°C overnight (see Methods 2.33). Photographic films were developed manually to a suitable intensity and contrast. A total of 12 positives were obtained and the strongest 4 of these (489g2, 567c6, 600j10 and 601f7) were ordered from the HGMP-RC for further analysis. Figure 5.13 shows an example auto-radiograph of a panel from the screening with the positive clone 600j10 identified by an arrow.

The PAC containing bacterial clones were cultured and DNA prepared from them by a modified alkali-lysis protocol (see Materials 2.20). This DNA was digested with the restriction enzymes *Bam HI*, *Eco RI* and *Hind III* and the digest fragments separated by electrophoresis in a 0.8% (w/v) agarose gel at 0.5V/cm to allow good resolution in the 2-5kb range (Figure 5.14). The digest fragments were transferred to nylon filters by alkali transfer Southern blotting (see Methods 2.29.2) and the filters probed with radio-labelled Xe169c1 probe. The filters were washed down to a background count of approximately 10cps and then exposed to X-OMAT film at -70°C overnight. Post-exposure, the films were developed manually to suitable intensity and contrast.

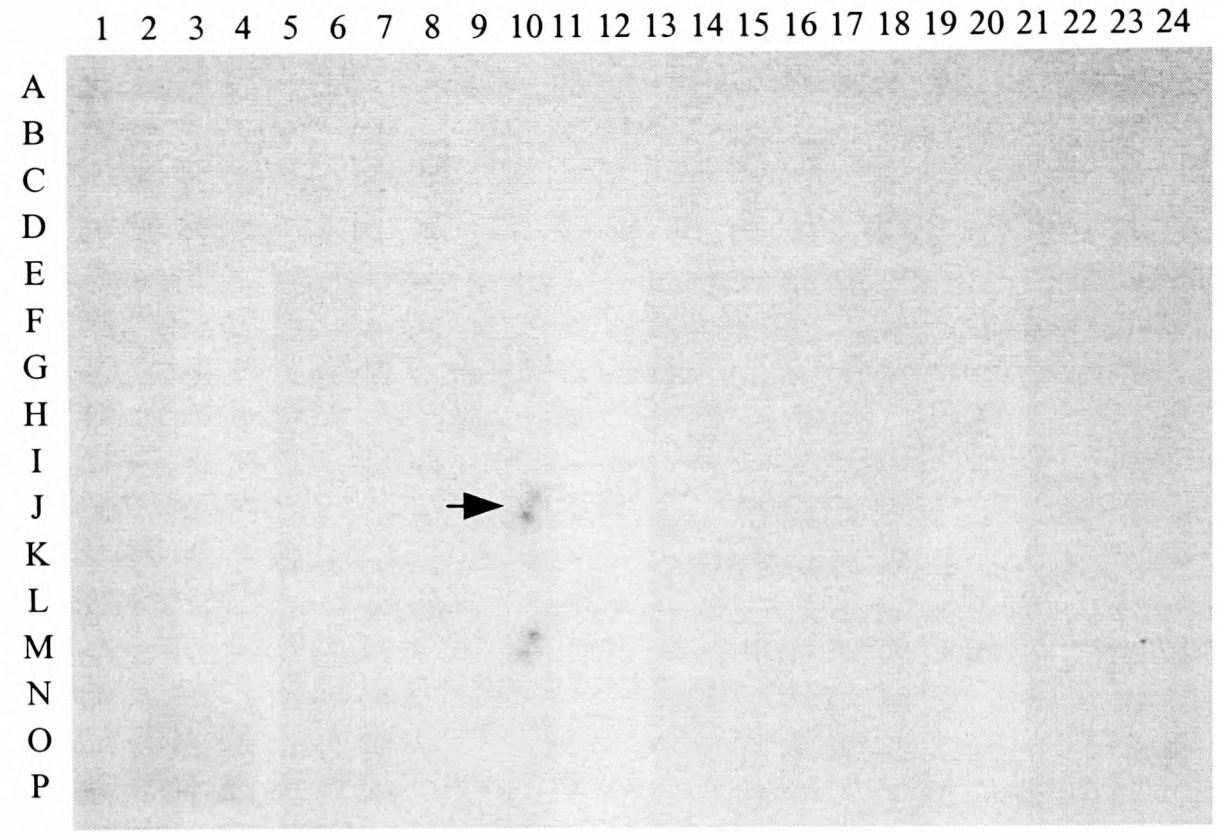


Figure 5.13: Screening of the murine RPCI21 PAC library with the human cDNA probe XE169c1.
The RPCI21 mouse PAC library was probed with the human 5' cDNA clone XE169c1 to identify clones containing the 5' end of the murine Smcx gene. The figure shows a panel from an auto-radiograph containing 2 of the 12 strong positives identified during the screening. The clone marked by an arrow is the clone 600j10, identification is simplified by the gridding pattern on the filter (background) and the orientation of the spots. A total of four strong positives were ordered from the HGMP-RC, 489g2, 567c6, 600j10 and 601f7 for further analysis.

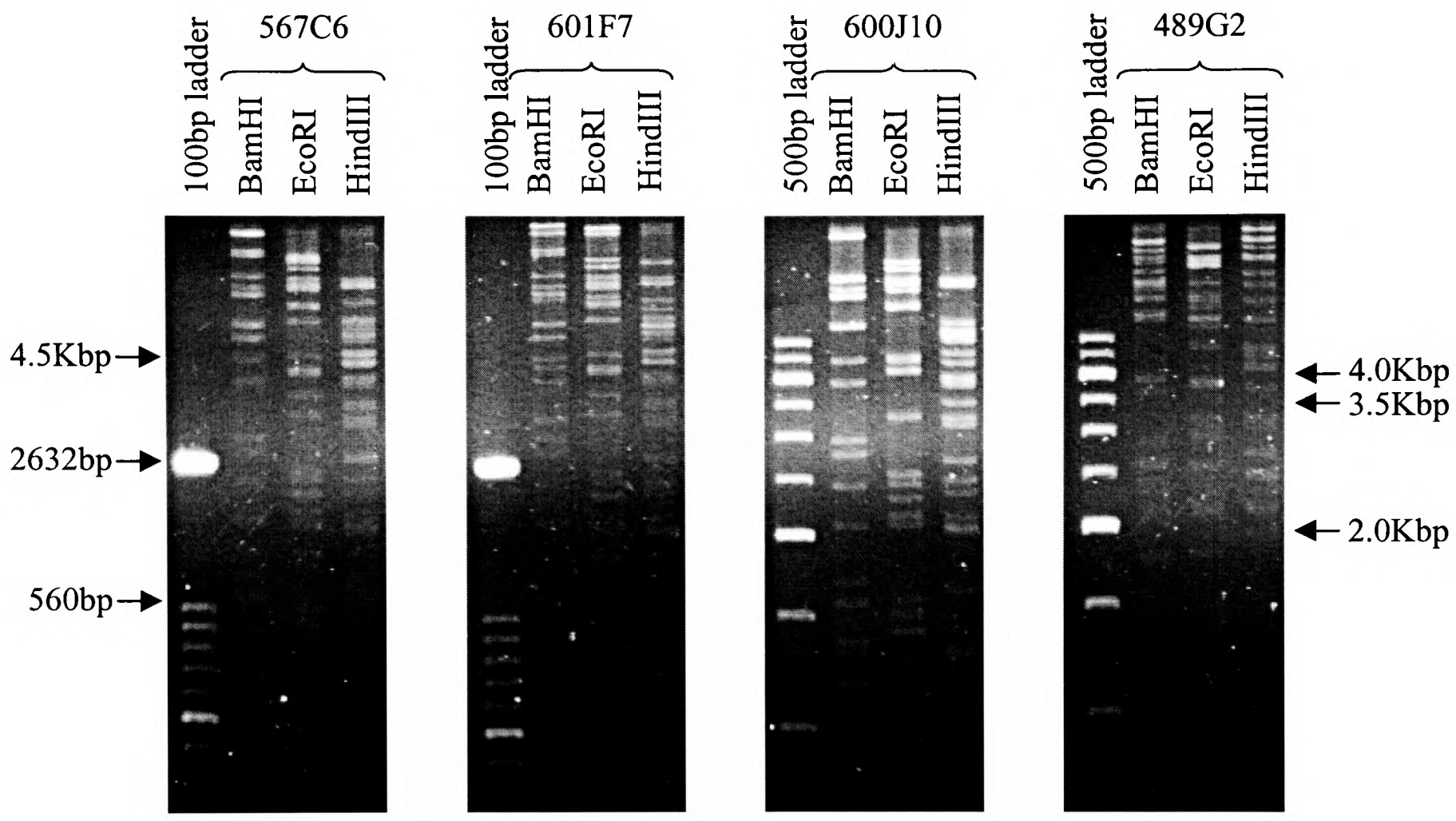


Figure 5.14: Restriction digests of PAC DNA from the four Xe169c1 positive clones.
DNA was extracted from four Xe169c1 positive PAC clones by a modification of the standard alkaline-lysis protocol (see Methods 2.20) and digested with either BamHI, EcoRI or HindIII. Fragments were separated in a 0.8%(w/v) agarose gel at 0.5V/cm to allow good resolution of fragments in the 2-5Kbp range. The digest patterns are complex, but all four clones appear to be different suggesting that they are non-identical. This gel was subsequently blotted onto nylon filters for use in hybridisation experiments to identify Xe169c1 positive fragments suitable for sub-cloning.

Figure 5.15 shows the results of the PAC digest hybridisation with the probe Xe169c1 in which clones 567c6, 601f7 and 600j10 are shown to contain single common positive fragments in each digest lane. The clone 489g2 contains different high molecular weight positive fragments and is not used further in this thesis. The identically sized positive fragments in the clones 567c6, 601f7 and 600j10 are a 750bp *BamHI* fragment, a 4.5kb *Eco RI* fragment and a 3.0Kbp *Hind III* fragment.

To clone the 4.5kb *Eco RI* Xe169c1 positive fragment, two 5µg aliquots of PAC DNA from the clone 600j10 were digested with *Eco RI* and the fragments separated by electrophoresis in a 0.8% (w/v) agarose gel at 1V/cm until the fragments around 4.5kb were well resolved (Figure 5.16). The doublet at 4.5kb was excised from both lanes and pooled prior to purification by gel extraction. The purified restriction fragments were cloned into the *Eco RI* site of the vector pZErO2 and the resulting ligation products transformed into TOP10 cells by electroporation. The transformants were cultured on selective plates overnight then picked and re-streaked onto fresh plates. Plasmid DNA was prepared from 20 random clones and analysed by restriction digestion with *Eco RI* to see if they contained the correct sized inserts (data not shown). Five clones appeared to have inserts in the right size range so plasmid DNA from each of these was digested with *Bam HI* and *Eco RI* separately and the resulting fragments separated by electrophoresis (Figure 5.17A). The DNA fragments were transferred to a nylon filter by alkali transfer Southern blotting and then probed with the Xe169c1 probe as described above. Figure 5.17B shows the auto-radiograph from the hybridisation confirming that clone 19 contains the 750bp positive *Bam HI* fragment and a positive *Eco RI* fragment of approximately 4.5kb.

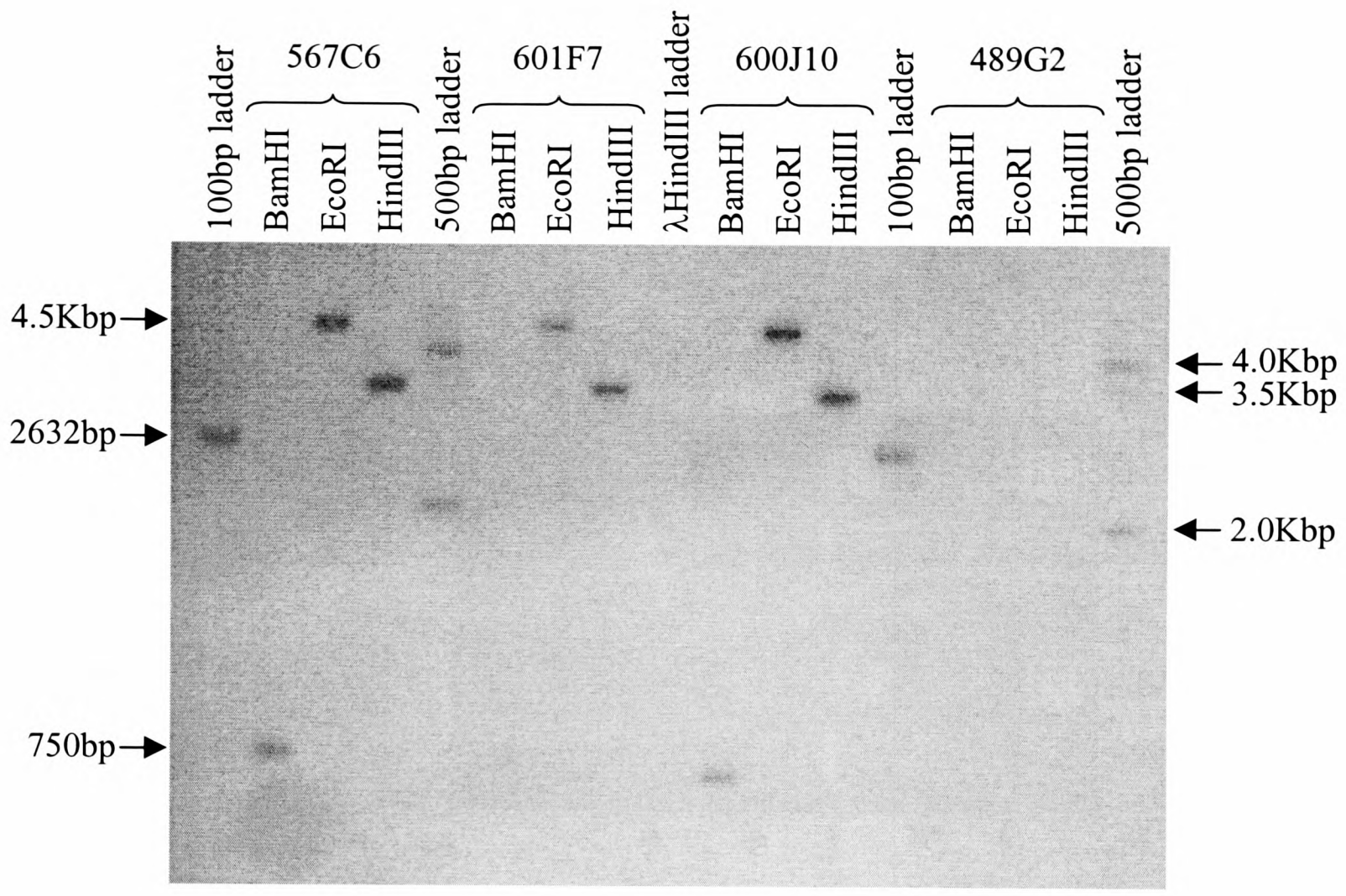


Figure 5.15: Hybridisation of the probe Xe169c1 to restriction fragments of the positive PAC clones.
Restriction digests of PAC clones 567c6, 601f7, 600j10 and 489g2 were separated on an agarose gel (Figure 5.14) and then blotted onto a nylon filter by alkaline transfer. The filter was used in a hybridisation experiment with the Xe169c1 probe to identify restriction fragments containing the 5'-end of the murine Smcx gene. Clones 567c6, 601f7 and 600j10 each share the same Xe169c1 positive fragments, one 750bp BamHI, one 4.5Kbp EcoRI and one 3.0Kbp HindIII fragment. Clone 489g2 had only very high molecular weight positive fragments and was not used further. The 4.5Kbp EcoRI fragment was chosen as the target to be sub-cloned to isolate the Smcx upstream sequence as it was the largest of the positive fragments that could be easily cloned.

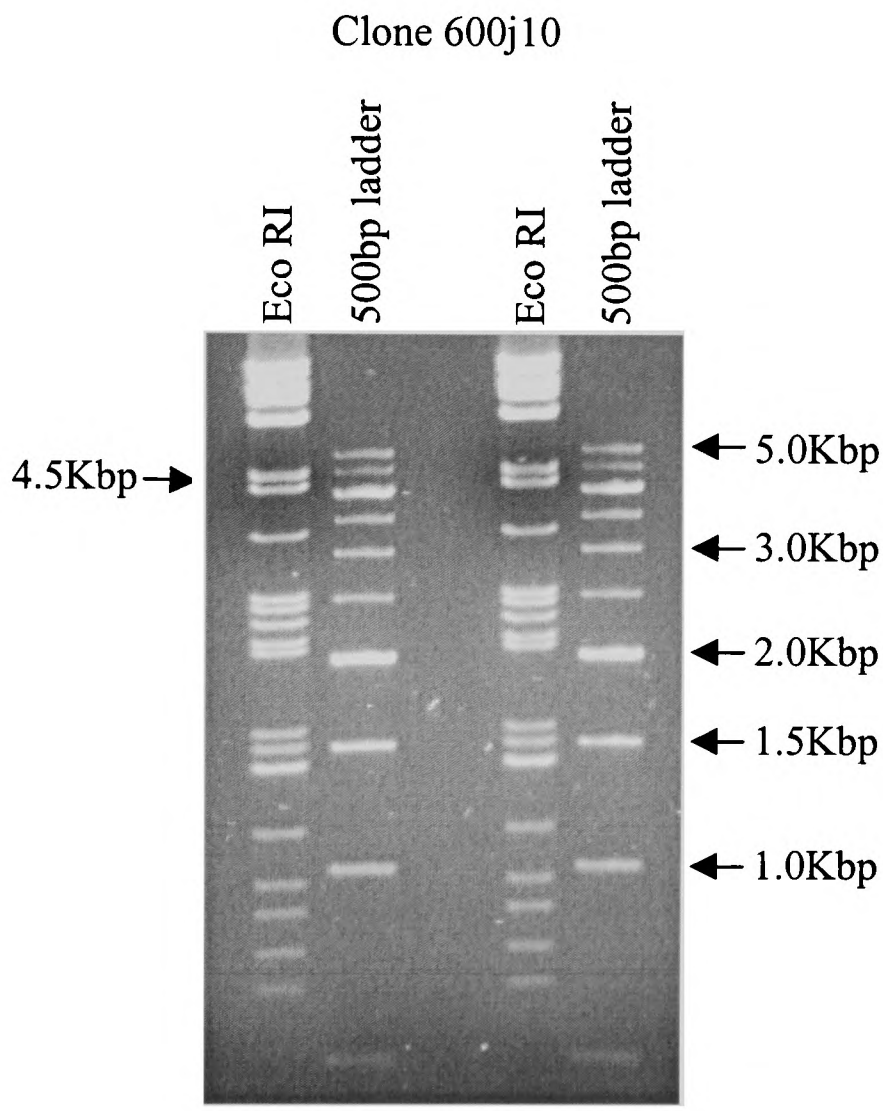


Figure 5.16: Preparative Eco RI digests of PAC DNA from the clone 600j10. To clone the Xe169c1 positive 4.5Kbp EcoRI fragment 5µg duplicates of PAC DNA were digested with EcoRI. Fragments were separated in a 0.8%(w/v) agarose gel at 0.5V/cm until the doublet at 4.5Kbp was well resolved from the other digest fragments.The doublet was excised, purified and used as insert in a ligation reaction with pre-cut pZErO2 vector DNA. The ligation products were transformed into competent *E.coli* by electroporation and plasmid containing clones selected for by culture on Kanamycin plates.

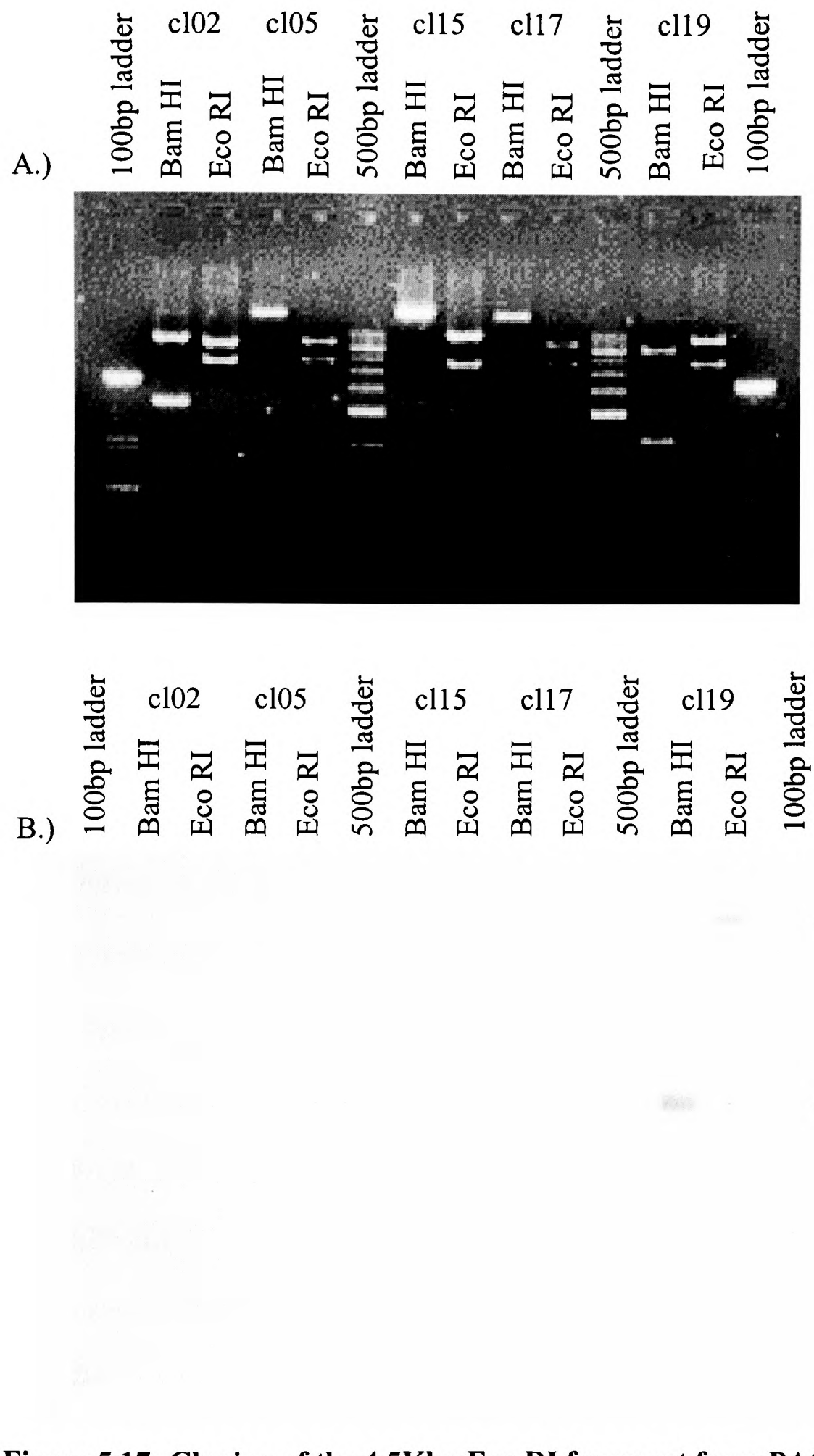


Figure 5.17: Cloning of the 4.5Kbp Eco RI fragment from PAC clone 600j10
DNA was prepared from 20 randomly selected bacterial clones generated from the previously described 4.5Kbp EcoRI fragment cloning experiment (Figure 5.16 and text). Digests with EcoRI identified five clones cl02, 05, 15, 17 and 19 which appeared to have inserts of the right size. These were then digested with BamHI to determine whether the predicted 750bp BamHI fragment was present. Clone cl19 had a BamHI fragment of the correct size and an EcoRI insert of 4.5Kbp (A). The digest gel was blotted onto a nylon filter by alkaline transfer and probed with the Xe169c1 probe (B). The only positive bands in the resulting auto-radiograph were for the cl19 digest producing the expected 750bp BamHI and 4.5Kbp EcoRI Xe169c1 positive fragments. This clone was subsequently used to isolate Smcx upstream sequence.

As no genomic sequence was available for the murine *Smcx* gene, the most efficient way to find the upstream sequence in the 4.5kb Eco RI clone was to sub-clone the 750bp Bam HI clone and sequence it entirely. This sequence could then be used to design a primer at the 5' end of the cDNA pointing upstream to allow sequencing of the proximal upstream region.

Plasmid DNA from clone 19 was digested with *Eco RI*, *Bam HI* and with both enzymes together to confirm the size of the insert and to separate out the 750bp *Bam HI* fragment for purification (Figure 5.18A). The estimated size of the *Eco RI* fragment from this digestion pattern is approximately 4.3kb and the digest shows that in the *Bam HI* lane the 750bp band is well separated from any other fragments. The 750bp fragment was excised from this gel and purified by gel extraction. An aliquot of 10ng of purified fragment was cloned into the *Bam HI* site of the vector pZErO2 and the resulting ligation products transformed into TOP10 cells by electroporation. The transformants were cultured overnight on selective plates and individual colonies re-streaked on fresh plates. Plasmid DNA was prepared from 6 randomly selected clones and digested with *Bam HI* to identify clones with the correct sized insert. Figure 5.18B shows that all 6 of the clones contained a 750bp insert. Plasmid DNA from two of the clones (XeB2 and XeB3) was sequenced by fluorescent cycle sequencing using the primers SP6 and T7 which flank the multiple cloning site. The entire 762bp sequence of their *BamHI* inserts is shown in figure 5.19.

The entire XeB2 insert sequence was gap aligned to the human *SMCX* cDNA sequence to determine whether the correct clone had been isolated and to define which part of the cloned sequence represented the 5'-end of the murine gene (Figure 5.20). The alignment shows that the murine insert shares 78.5% sequence identity with the human *SMCX* cDNA. The homology ends at position 24 of the XeB2 insert sequence which corresponds to the

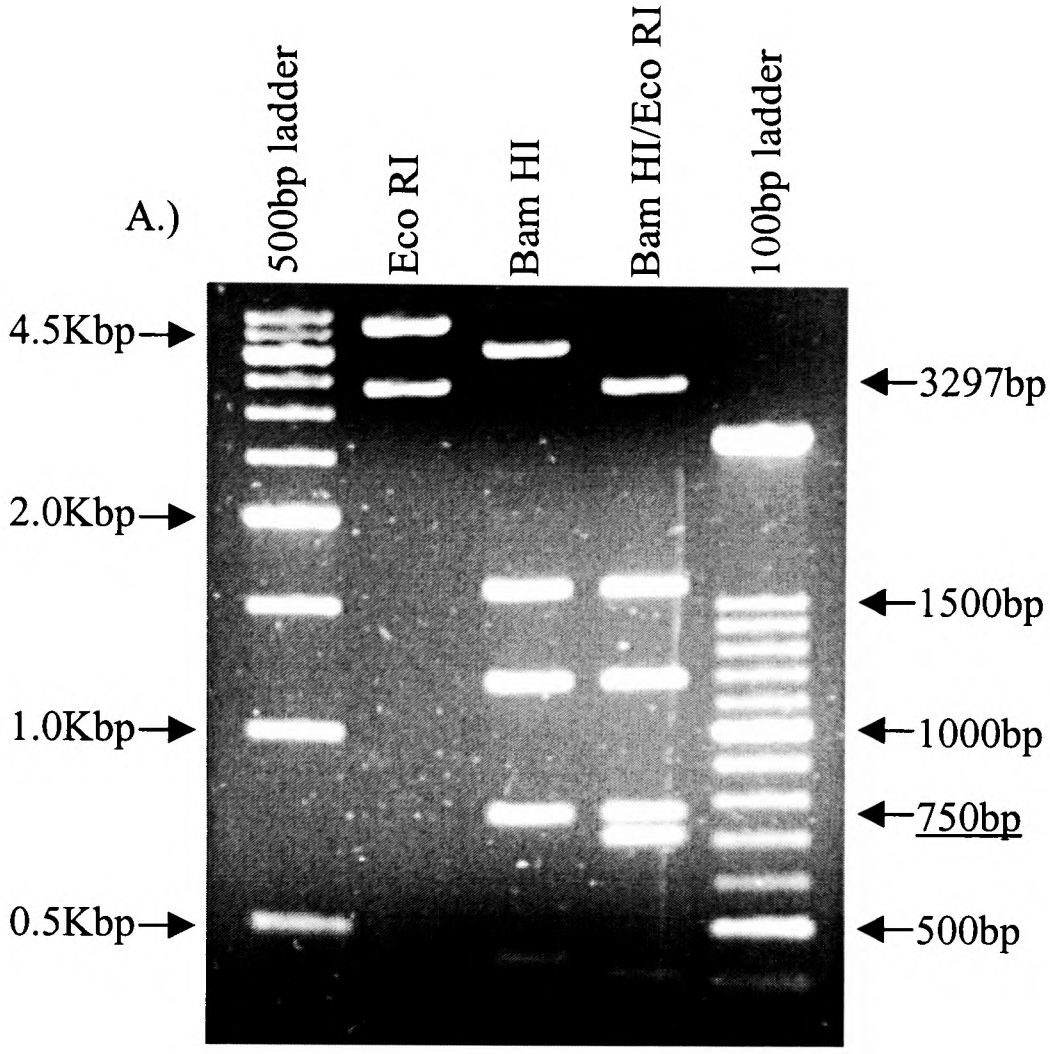


Figure 5.18: Digest of clone cl19 and sub-cloning of the 750bp BamHI fragment. DNA was isolated from the Xe169c1 positive murine sub-clone cl19 and digested with EcoRI, BamHI and EcoRI/BamHI to confirm the size of the insert and map the restriction sites (A). The digest showed that the 750bp band in the BamHI digest was easily separable from the other fragments and that the total insert size was likely to be closer to 4.3Kbp. The 750bp band was excised from the gel and cloned into the BamHI site of pZErO2 and the ligation products transformed into competent TOIP10 *E.coli* by electroporation. Plasmid containing clones were selected for by culture on Kanamycin plates and DNA prepared from six randomly chosen colonies by rapid miniprep. BamHI digests of DNA from all six clones confirmed that they contained the desired 750bp BamHI fragment (B)

BamHI

GGATCCGGCCATCGATTTACAATTTTGTTCTTCCGCCAATGAAATGAACTATTTTCTCTT
AGCCTGAGAAATGTTTTAACTAACCACCTCCAAATAAGGGACATATACAACCCGATTTCT
TTCGGTCATCACATGATGAAGGCTTCCTAGGTTTTTTTCTGACTCAAAGTAAACGGAACT
CGGAGAGACACCTTGCGGAGGGATCTGGTTTGTTGTGGTGAGGTACAGAGGTGGGGGGGA
GGAGACGCTGACAAACCAAGATGGCGGTGGCAGCGGTGAAACCTGCCGCGACTGGGACTT
AACTGTAGTAGTGAAGGCTTCAGTAGTTGGGAGGCAGCGGTAAGGTGTGGAAGAAACGGA
ACAGGACTAAGAACCTGTAAAAGGAACCTCAGCCTGAGACTTTTAGAGCACGGCACCCGC
CCCCCTAAGCCCTTGACGGGCGGTGGCCGTCCTGCTTAGGGCCTAGACTCCGAGAAGTAC
CTCGGTTTCAGAGAGCGGCGGCGCCATGCGTCCTTAAGGGCGACCTGAGCCCTCCCGATC
TCTGGCCCAGCCCTCGGGCCCACCATGGAGCTGGGGTCCGACGATTTCCTACCGCCGCCG
GAGTGTCCGGTGTTCGAGCCTAGCTGGGCCGAGTTCCGAGACCCTCTTGGCTACATCGCG
AAAATCAGACCCATCGCCGAGAAGTCCGGCATTTGCAAGATCCGCCCACCCGCGGTAAGG
CTGGGAGCTGATGGTCGCCAGCGGCGAGTTTGTGGTGGATCC

BamHI

Figure 5.19: Nucleotide sequence of the Xe169c1 positive clones XeB2 and XeB3. High-quality miniprep DNA was prepared from two of the six clones previously identified as containing the 750bp Xe169c1 positive BamHI fragment (XeB2 and XeB3). Their inserts were sequenced entirely from either end using the vector MCS flanking primers SP6 and T7 by fluorescent cycle sequencing. The 762bp sequence is shown above with the BamHI restriction sites boxed.

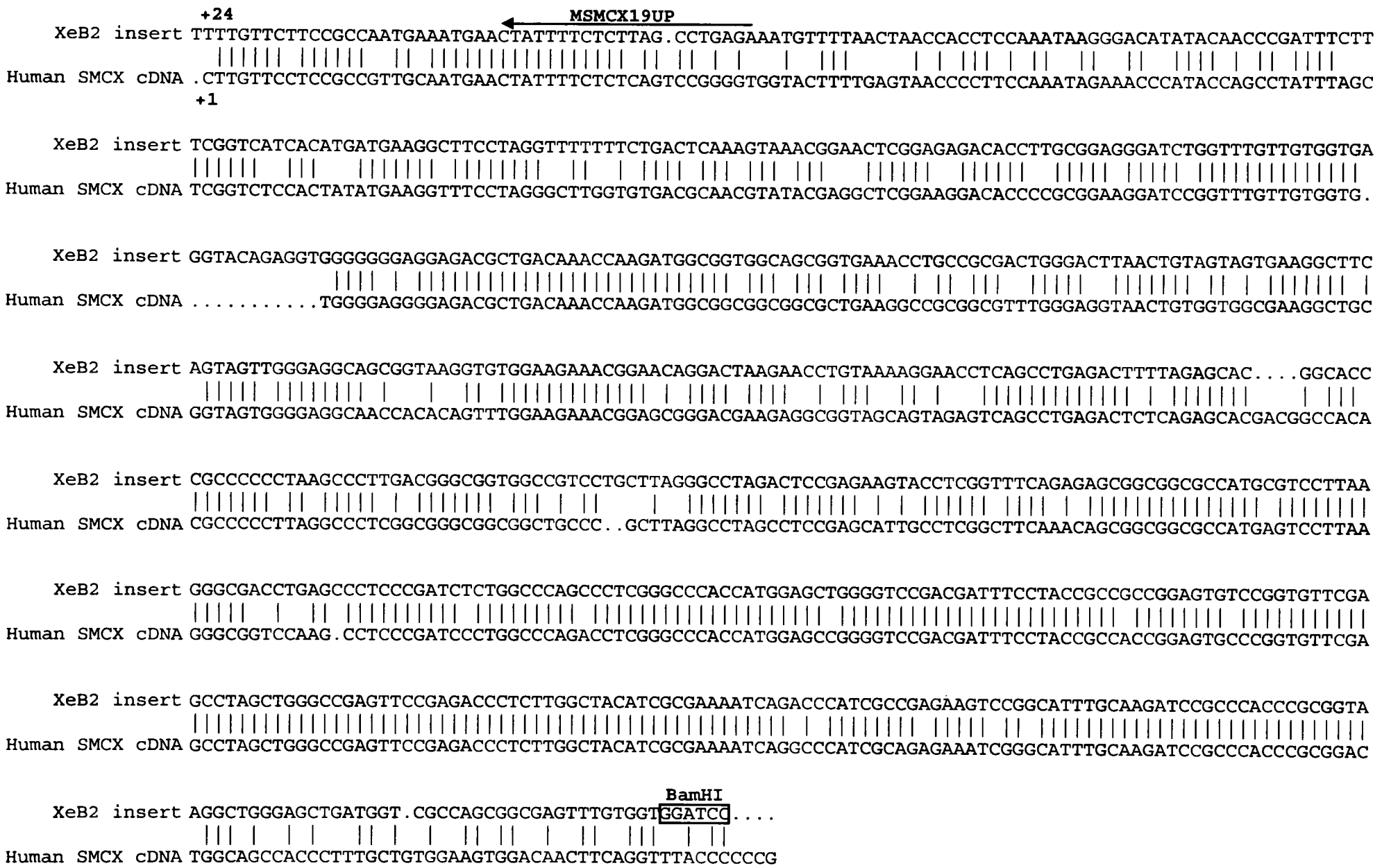


Figure 5.20: Gapped alignment of the 762bp XeB2 clone sequence against the human Xe169cDNA sequence.
The sequence from the Xe169c1 murine BamHI clone XeB2 was gap aligned against the full human SMCX cDNA sequence to determine how much upstream sequence had been obtained. The alignment shows that the XeB2 clone only contains 23bp of upstream sequence so a new sequencing primer MSmcx19Up was designed to isolate more sequence using the original 4.5Kbp EcoRI cl19 clone as template. Interestingly the two sequences show 78.5% identity in this alignment and at least one intron appears to be present although this would need to be confirmed by sequencing the 5'-end of the murine Smcx cDNA and aligning it to the murine genomic sequence.

transcription start site of the human cDNA sequence. This means that the sequence orientation shown for the XeB2 insert of figure 5.19 is the same as that of the coding strand of the murine *Smcx* gene and that in total the clone contains 23bp of upstream sequence. To isolate more upstream sequence, the primer MSMCX19UP (5'-CTCAGGCTAAGAGAAAATAG-3') was designed to anneal to the opposite strand of the 5' end of the XeB2 insert (see Figure 5.20). Plasmid DNA from the original clone 19 was used as template for fluorescent cycle sequencing using the MSMCX19UP primer and an additional 528bp of upstream sequence was generated (Figure 5.21).

5.2.5 Analysis of the Murine *Smcx* gene upstream region.

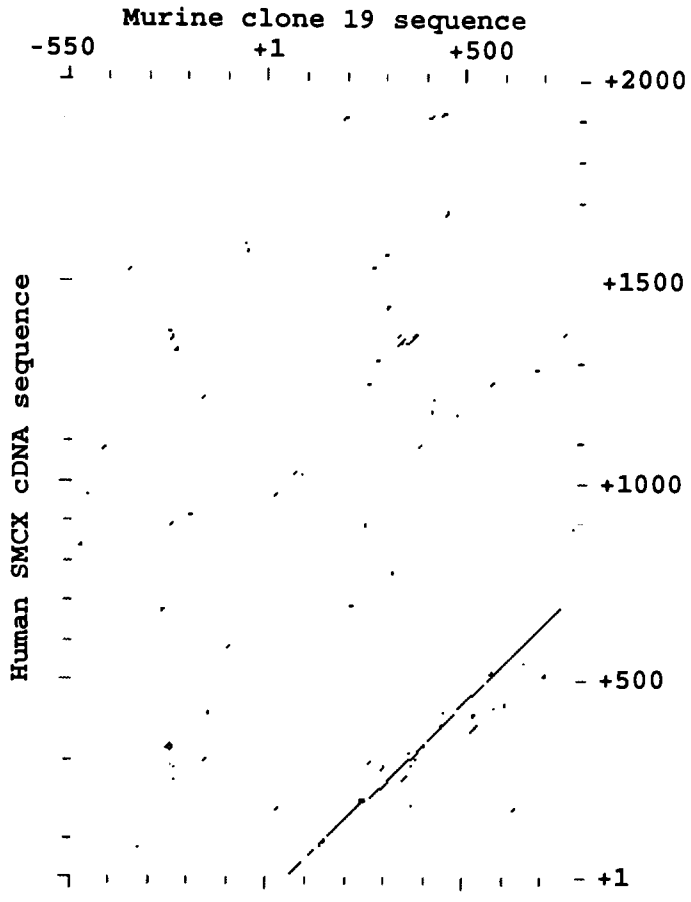
The entire sequence generated from clone 19 was compared to the human *SMCX* cDNA by using the programs Compare and Dotplot (see Methods 2.56.4). Figure 5.22 shows the dotplots which illustrate the high degree of similarity between the two sequences. Some of the gaps in the diagonal line are likely to be introns, but this would have to be confirmed by comparing the sequence to the murine *Smcx* cDNA sequence which has yet to be isolated.

The murine upstream sequence was analysed for the presence of methylatable sites and CpG-islands as described above for the human upstream sequence and for other sequences in this thesis. Figure 5.23 show a CpG-plot of the murine *Smcx* upstream region which was found to contain a CpG-island from –322 to –42bp (280bp), 5 *Hpa II* and 7 *Hha I* sites and 76 CpG di-nucleotides. Analysis of the upstream sequence reveals no TATA or CAAT box homologies, but several consensus binding sites for the transcriptional activator Sp1 suggesting that *Smcx* may be a widely expressed or housekeeping gene. A detailed analysis of the nucleotide sequences of the human and murine *SMCX*/*Smcx* upstream sequences is presented in the next section.

CATATTATACAAAGTGTTTATATAGTTACAATGTATACCAAACAGTCGAAAGCATTCTTT
AAAATGCTGATGGCCAATGATACTGGGGAGAATAATGCATTTTTACTCTACGTGGGACTG
AAGCTTCTAAAACTCCACATTCTCTAGGGAGTGAGGAGTTTCCTTCTTTCATTTAAAAAC
AAAAACCTCAGAGGTGTAAGTCTGGTAAATCTTCAATTCACTCAGCAGTAGACACGCGGA
ATGGGAGGGGGAGGGGGAGGAAGGGGGCGGAGGTTCCTCGAAATCACCGCCTTCCGTGGT
TCTCACTGAAAGGACCTGAGAGAAGCAGCAAAGGACATTGGGGAGGAGACAACCGGCTTG
GGATCTAGGGTGTGTGCGTGCGCGTGCGCGAGCGTGTGCGAGCGCGCATGGGCGCGTACT
AGTTTAGGCTCACTACTTATGGACACTAGCTCCTCTTCCGGTTTGTCTAGAGACTGCGCA
TGCGCATCACTGGCCTTTAAACGTGCATCGCCGGTCCATCCGGTAAAAGGATCCGGCCAT
CGATTTACAATT**T**TGTTCTTCCGCCAATGAAATGAACTATTTTCTCTTAGCCTGAGAAAT
GTTTTAACTAACCACCTCCAAATAAGGGACATATACAACCCGATTTCTTTCGGTCATCAC
ATGATGAAGGCTTCCTAGGTTTTTTTCTGACTCAAAGTAAACGGAACTCGGAGAGACACC
TTGCGGAGGGATCTGGTTTGTTGTGGTGAGGTACAGAGGTGGGGGGGAGGAGACGCTGAC
AAACCAAGATGGCGGTGGCAGCGGTGAAACCTGCCGCGACTGGGACTTAACTGTAGTAGT
GAAGGCTTCAGTAGTTGGGAGGCAGCGGTAAGGTGTGGAAGAAACGGAACAGGACTAAGA
ACCTGTAAAAGGAACCTCAGCCTGAGACTTTTAGAGCACGGCACCCGCCCCCCTAAGCCC
TTGACGGGCGGTGGCCGTCCTGCTTAGGGCCTAGACTCCGAGAAGTACCTCGGTTTCAGA
GAGCGGCGGCGCCATGCGTCCTTAAGGGCGACCTGAGCCCTCCCGATCTCTGGCCCAGCC
CTCGGGCCCACCATGGAGCTGGGGTCCGACGATTTCCTACCGCCGCCGGAGTGTCCGGTG
TTCGAGCCTAGCTGGGCCGAGTTCCGAGACCCTCTTGGCTACATCGCGAAAATCAGACCC
ATCGCCGAGAAGTCCGGCATTTGCAAGATCCGCCCACCCGCGGTAAGGCTGGGAGCTGAT
GGTCGCCAGCGGCGAGTTTGTGGTGGATCC

Figure 5.21: Nucleotide sequence of the XeB2 clone and the additional upstream sequence generated by sequencing clone 19 with the primer MSMCX19UP.
The sequencing primer MSmcx19Up (underlined) was designed at the 5'end of the sequence derived from the 762bp BamHI clones XeB2 and XeB3 to obtain further upstream sequence of the Smcx gene. Miniprep DNA from the 4.5Kbp EcoRI clone cl19 was used as sequencing template and produced a further 528bp of upstream sequence. In total 551bp of sequence upstream of the putative transcription start site (**bold** and forward arrow) was isolated. The BamHI sites are boxed to show the sequence contained within the XeB2/3 clone inserts.

A.)



B.)

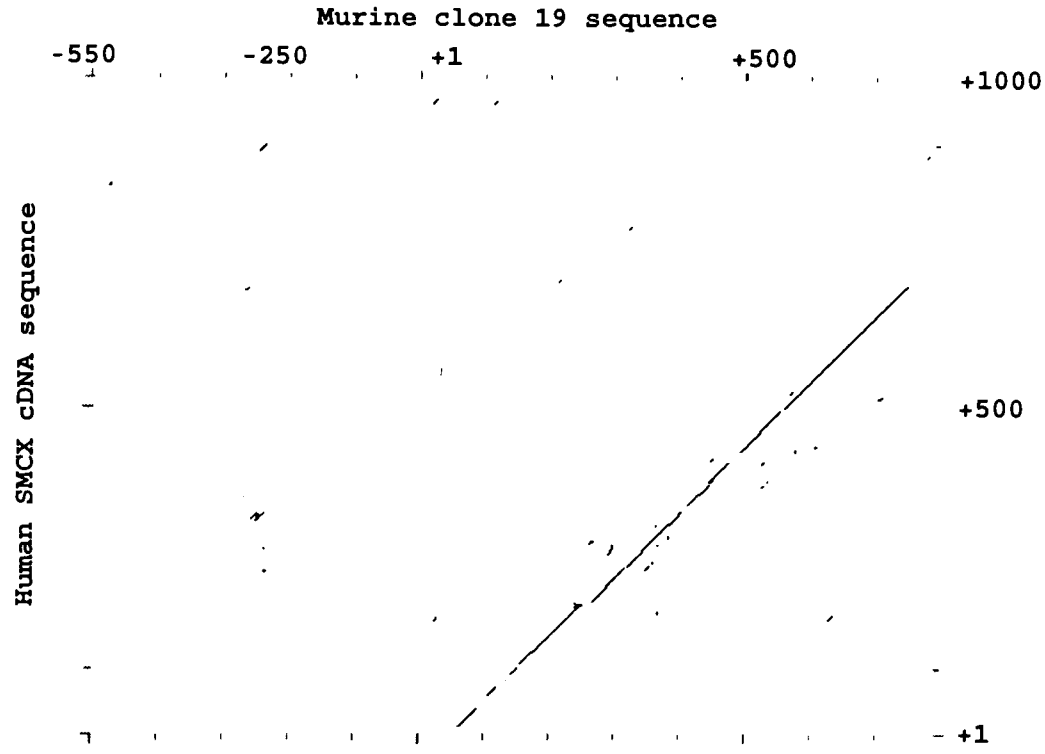


Figure 5.22: Dotplots between the Human SMCX cDNA and murine clone 19 sequences.

The GCG-Wisconsin programs Compare and Dotplot were used to produce graphical representations of the similarity between the murine Smcx genomic sequence isolated from cl19 and the full length human SMCX cDNA. The first low-resolution plot (A) shows the very high level of similarity between the two sequences as a near continuous diagonal line. This line stops as the murine cl19 sequence begins to goes upstream of the transcription start site. In the narrower field graph (B) it is apparent that there are breaks in the similarity diagonal which in at least some cases will be introns in the murine sequence. Notably similarity seems to fall away right at the 5'-end of the cDNA which could either mean that there are several small exons there or perhaps that the first exon is not translated and poorly conserved.

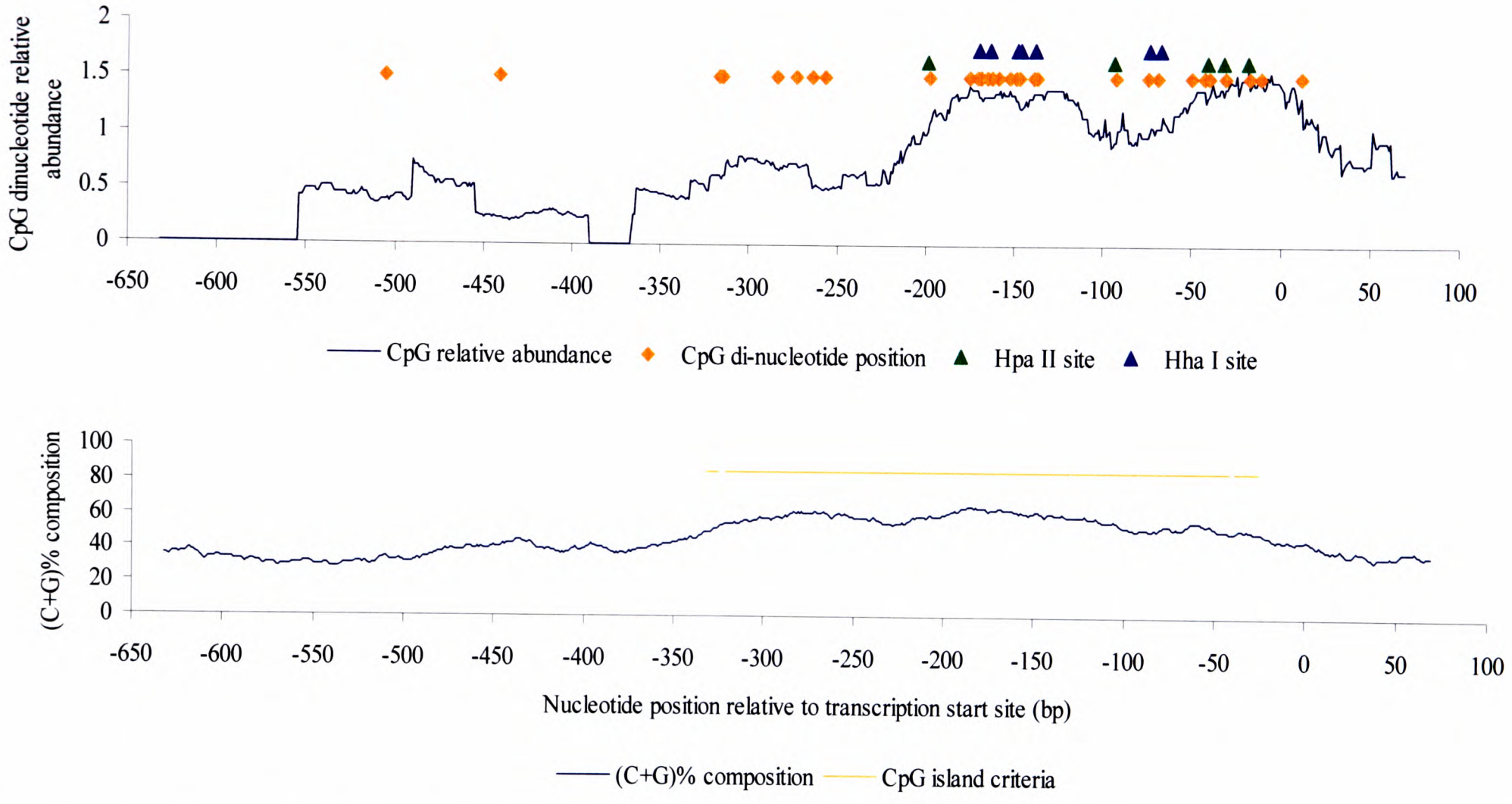


Figure 5.23 CpG-Plot of the Murine Smcx upstream region.
The GCG-Wisconsin CpG-Plot program was used to generate co-ordinates for plotting the CpG relative di-nucleotide abundance and (C+G)% content against nucleotide position for the Smcx upstream sequence isolated in this chapter. The position of the 76 CpG sites is marked on the graph as are the positions of HpaII and HhaI restriction sites. The lower graph shows the (C+G%) content and an assessment of whether the sequence contains a CpG-island. The points which satisfy the island criteria are shown in orange. If these criteria are satisfied by a continuous stretch of 200bp then the section qualifies as a CpG-island. This section satisfies the criteria over at least 250bp and therefore qualifies as a CpG-island.

5.2.6 Comparison of the human and murine *SMCX/Smcx* upstream sequences.

The human and murine upstream sequences were compared to each other in order to determine whether there were any conserved areas between them. The sequences were aligned using the programs Gap (see Methods 2.56.2) and Compare/Dotplot (see Methods 2.56.4). The Gap alignment shows that the two sequences are highly similar in the section proximal to the transcription start site with the first 132bp of upstream sequence being 82% identical (Figure 5.24). The similarity decreases substantially after position –132bp with the remaining 209bp being only 36% identical. The similarity between the two sequences is visualised easily when presented as a dotplot in which similar regions are seen as diagonal lines (Figure 5.25). The figure illustrates the high level of identity from -130 to +200bp but distal to this there is no significant similarity apart from a small section centred around –500bp (see Figure 5.25 and 5.26A). This conserved section is 62bp long and 80% identical between man and mouse suggesting that it may have some structural or functional significance.

Figure 5.27 shows a gapped alignment of the proximal section of the human and mouse upstream sequences with the positions of conserved transcription factor consensus binding sites indicated. All of these are contained within the proximal 72bp of upstream sequence and include sites for the transcription factors NRF, Sp1, LBP-1 and H4TF2. It is possible that this very small, but highly conserved (93% identical) section of upstream sequence represents the core promoter of the human and murine *SMCX/Smcx* genes. It should be noted, however, that there are other conserved sections between the two promoters which do not contain consensus biding sites for known transcription factors. It is possible that other as yet unidentified transcription factors could bind to one or more of these areas. The highly

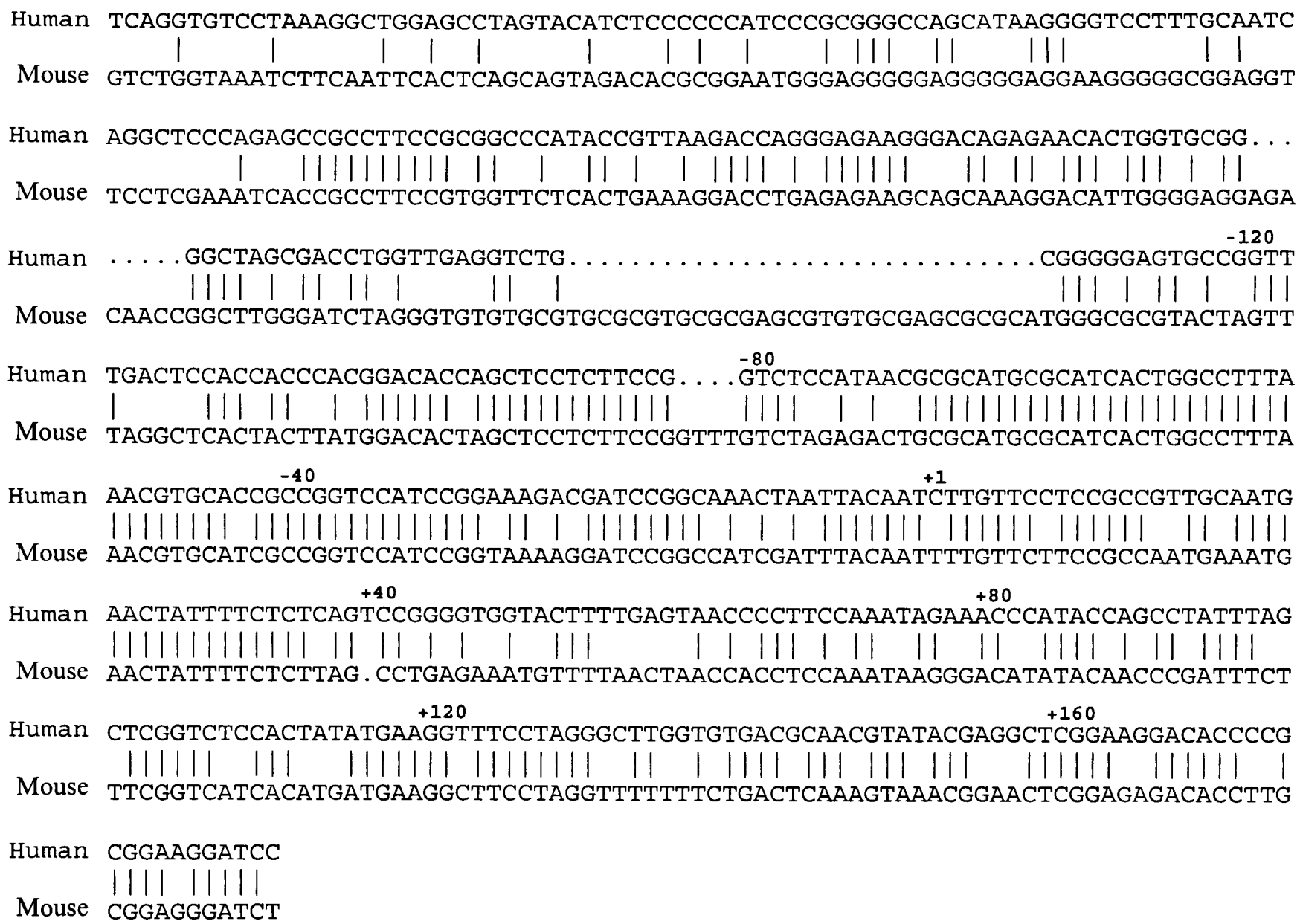
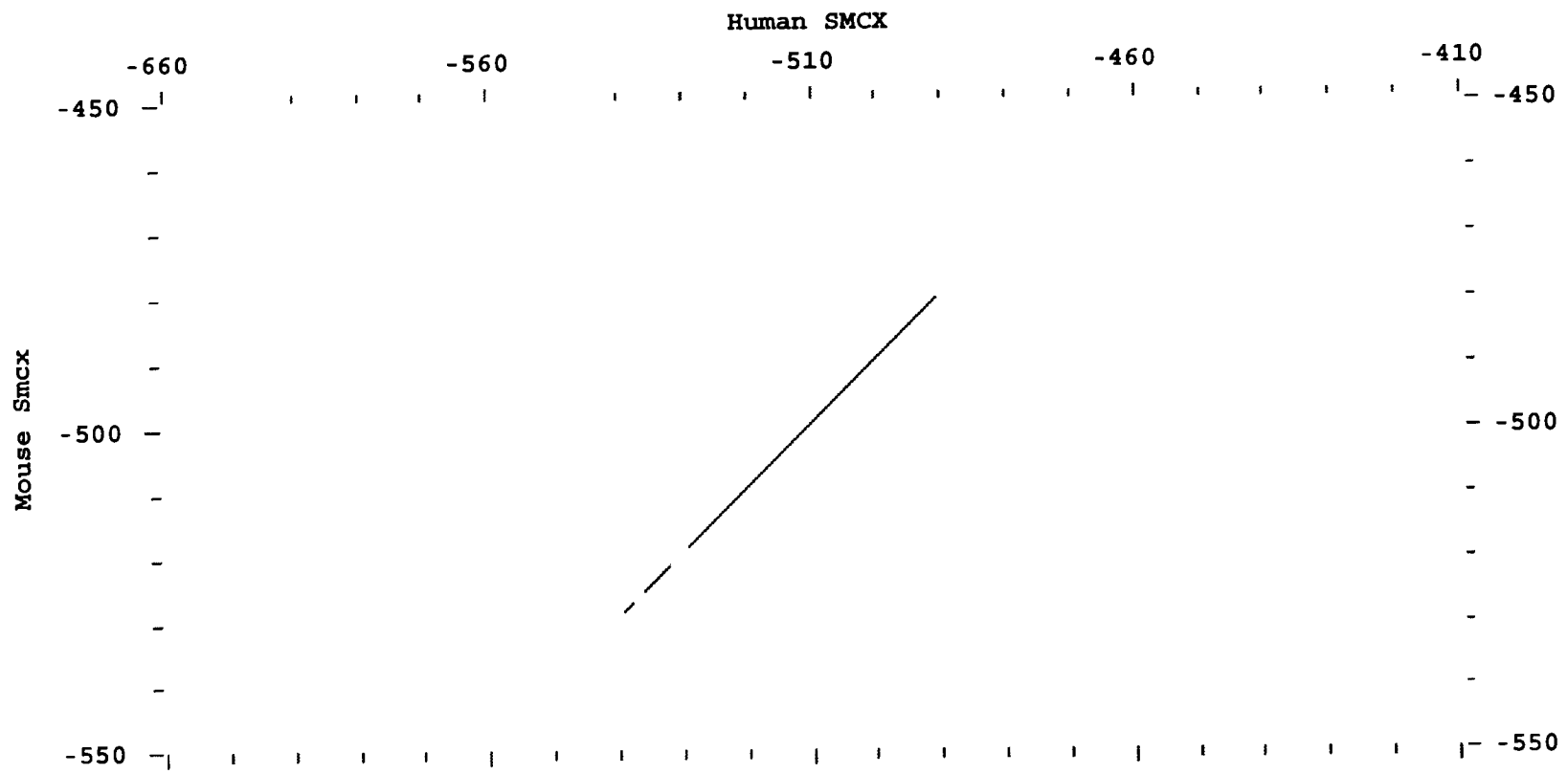


Figure 5.24: Gapped alignment of the human and mouse SMCX/Smcx sequences.
The full length human and murine SMCX/Smcx sequences isolated in this chapter were gap-aligned to each other to assess the similarity between them and to identify any conserved sections of sequence. The first 132bp of upstream sequence is highly similar (82% identical) but the similarity then falls away sharply (36% identity). This pattern is commonly seen at the promoters of widely expressed genes where a strong core-promoter is conserved and little or no other upstream similarity is apparent. Notably there does not appear to be anything similar to a TATA or CAAT box in either sequence.

A.)



B.)

```
Human SMCX  CAAA.TGTGTACATGGTAAAAGTGAATACAGAACAGTTGAAAGCGTTCATTAAAATGCTGAC
            ||||.|||.||.||.||.|.|.||.||||..||||||.||||||.|||.|||||||.|||||
Mouse Smcx  CAAAGTGTTTATATAGTTACAATGTATACCAAACAGTCGAAAGCATTCTTTAAAATGCTGAT
```

F2F
(Pit-1)

Figure 5.26: Dotplot and alignment of a conserved section between the human and murine SMCX/Smcx upstream regions.
A close-up of the highly conserved (80% identical) 62bp upstream section of the human and murine SMCX/Smcx genes (A) and the sequence alignment showing a conserved transcription factor binding site. The isolation of the section is interesting as it is reminiscent of other housekeeping gene promoters which have tissue-specific regulatory components as upstream elements outside the core-promoter. In this case there is a conserved Pit-1 transcription factor site present although as yet undefined transcription factors or regulatory proteins may bind to the other conserved sites. The Pit1 protein is a transcription factor which acts to up-regulate pituitary and other brain specific genes. Whether Pit-1 affects the level of SMCX/Smcx expression in the brain is not known at present.

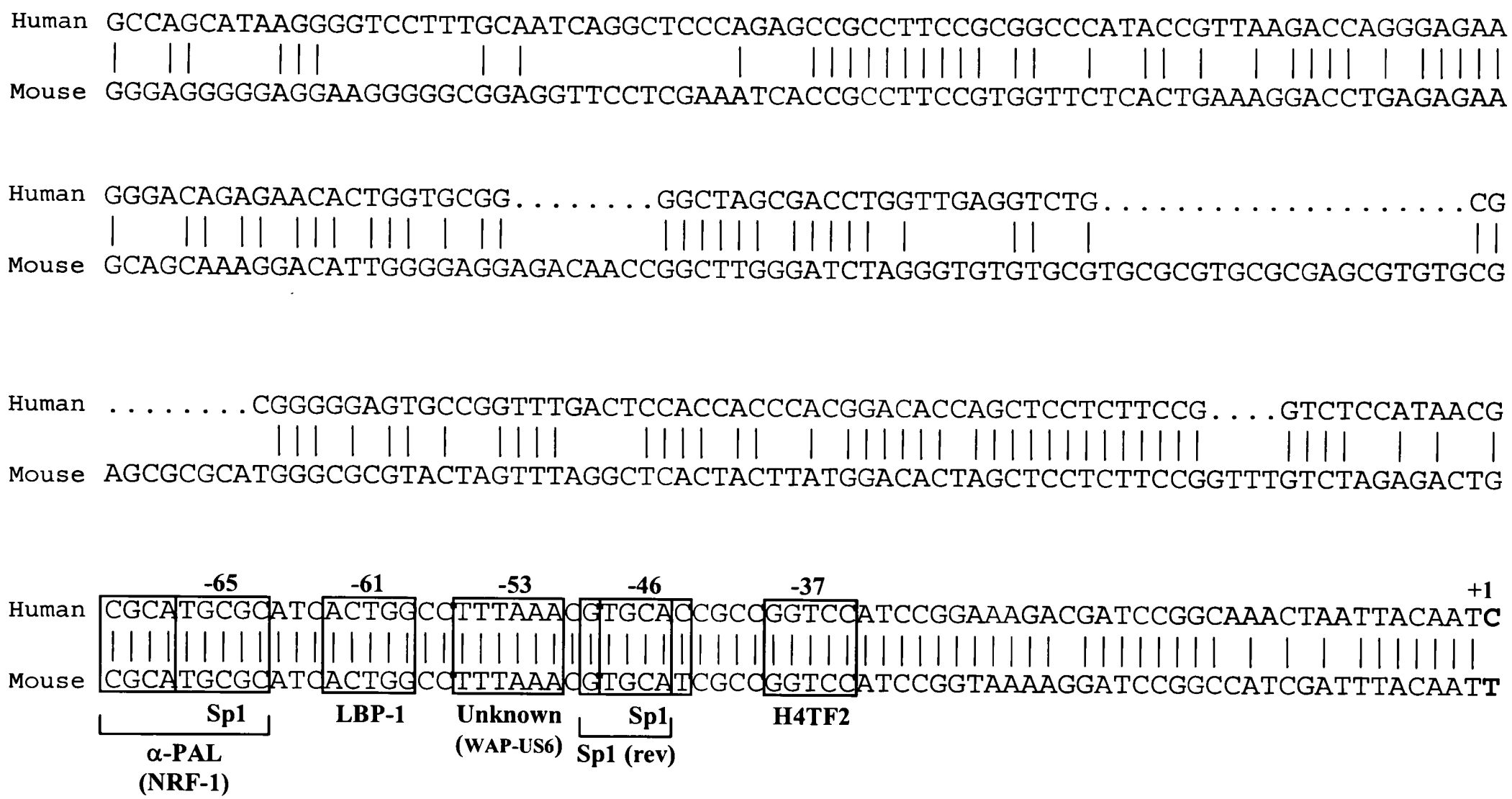


Figure 5.27: Gapped alignment and conserved consensus transcription factor binding sites between the human and murine SMCX/Smcx upstream sequences.

Positions of conserved transcription factor binding sites are superimposed onto the aligned human and murine SMCX/Smcx upstream sequences. Sites for the proteins αPAL/NRF-1, LBP-1, H4TF2, Sp1 and WAP-US6 are present. The transcription factors αPAL/NRF-1 and LBP-1 are involved in the up-regulation of genes involved in the control of growth and metabolic activity whereas H4TF2 is responsible for ensuring that genes are switched on particularly early in development when many other transcription factors are not expressed. The Sp1 protein is a ubiquitously expressed transcription factor used by most housekeeping genes as a powerful transcriptional activator. The Sp1 protein can interact positively directly through the transcription initiation complex through TBP. WAP-US6 is a protein of unknown function.

conserved upstream section centred around –500bp contains a binding site for the transcription factor Pit-1 (see Figure 5.26B). It is possible that other conserved parts of this region could contain binding sites for as yet undefined transcription factors.

As described in chapter 4, the conservation of methylatable sites could help to identify those that are potentially involved in the control of a gene's expression. Figure 5.28 shows the alignment of the human and murine proximal upstream sections with the positions of CpG di-nucleotides highlighted. The alignment shows that there are 11 conserved CpG sites in the available upstream sequences with 8 of these lying within the proximal 132bp. The positions of 6 of these 8 CpG sites saturate the positions of the conserved transcription factor binding sites, demonstrating that CpG sites are conserved in those parts of the sequence which contain potential regulatory elements. This parallels the findings for the *PDHA1*, *NDP* and *Ndp* gene promoters described in chapters 3 and 4.

5.2.7 Expression analysis of the *SMCX/Smcx* gene.

i) RT-PCR experiments.

Previous reports have suggested that *SMCX* and *Smcx* are widely expressed genes (Carrel et al., 1996; Sheardown et al., 1996; Wu et al., 1994). In order to confirm that the genes are similarly expressed in the mice used in this laboratory, RT-PCR experiments were performed on total RNA isolated from multiple murine tissues. This was essential for ongoing work characterising expression patterns and promoter methylation profiles for these genes in different tissues and for the selection of a suitable cell line for transfection studies (section 5.2.7(iii), below).

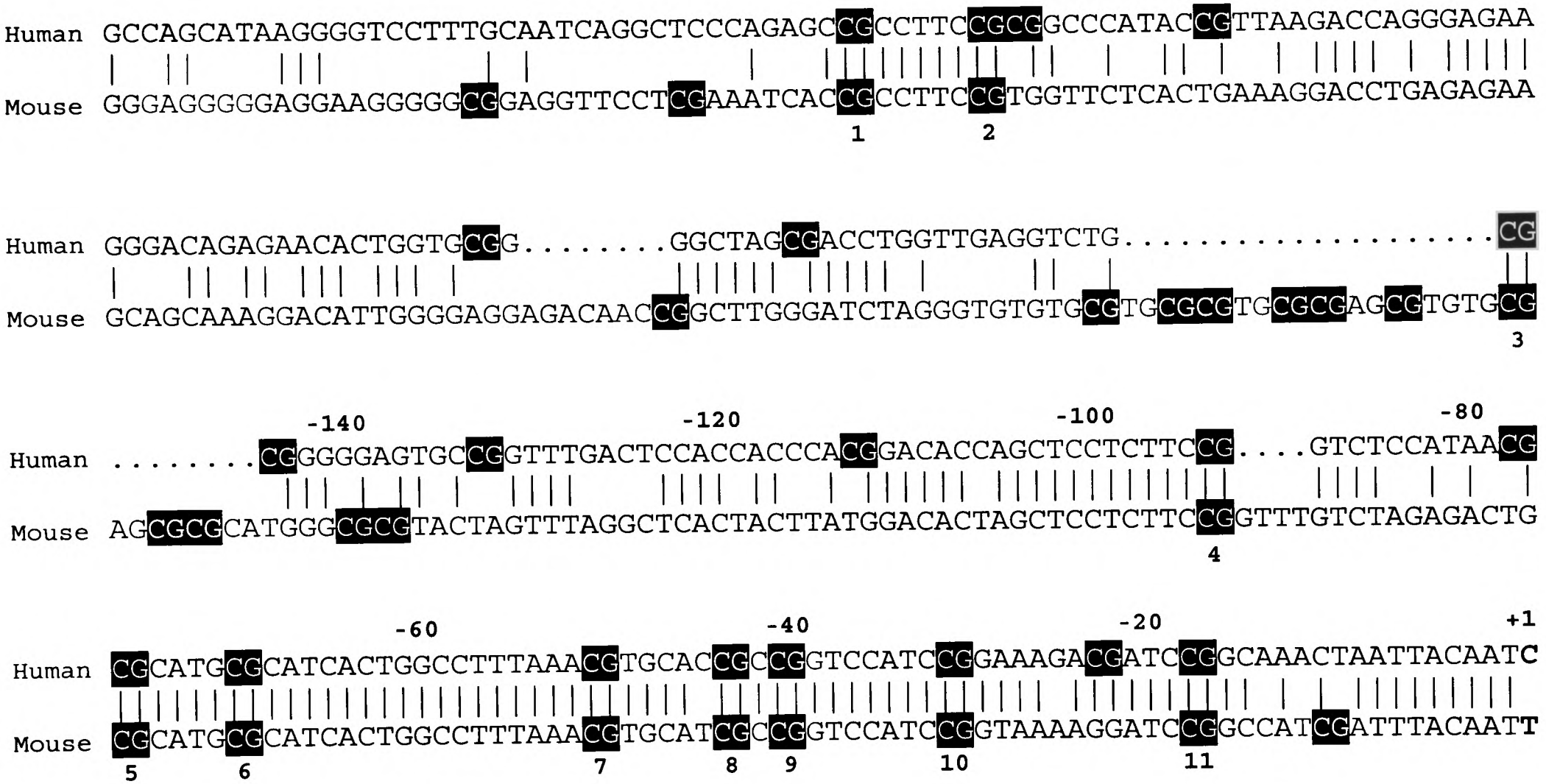


Figure 5.28: Alignment of the human and murine SMCX/Smcx upstream sequences showing the positions of CpG di-nucleotides. The positions of CpG di-nucleotides are highlighted in black and conserved sites are numbered. In total there are eleven conserved CpG-sites 8 of which (4-11) saturate the region in which conserved transcription factor binding sites are found. This is consistent with the concept established in chapters 3 and 4 that CpG sites are frequently found to be conserved directly at or near to key transcription factor binding sites at the promoters of genes. The human and murine SMCX/Smcx proximal upstream sequences are densely populated with CpG sites. If these are methylated *in-vivo* it seems likely that this would affect gene expression. Analysis of methylation at the promoters of the SMCX/Smcx genes *in-vivo* could help to explain some of the mechanisms behind escape from X-inactivation. The sequences shown in this figure represent ideal targets for analysis as they are small and highly conserved.

Total RNA was extracted by an adaptation of the method of (Ref.) from murine samples of brain, heart, kidney, liver, muscle, spleen and whole eye and from human retinal tissues. The RNAs were used to make first-strand cDNA (see Methods 2.45) for use in PCR reactions. The murine RT-PCR reactions used the primers MSMCX3PF1 (5'-GCTGGTAGCCCTACAGAGACTG-3') and MSMCX3PR1 (5'-TCTGTGGCTACCTTTTCAGGAC-3') and the human RT-PCR reactions used the primers HSMCX3PF1 (5'-GACCTGTGTCAGGACTGGTTC-3') and HSMCX3PR1 (5'-CTGGACCTTAGGCATATCCTTG-3') with conditions as shown in Table 5.1 (below).

Murine *Smcx* RT-PCR (297bp product) MSMCX3PF1:MSMCX3PR1 Qiagen buffer	96°C for 2'00'' 96°C for 0'30'' 60°C for 0'30'' 72°C for 0'15'' } 35 cycles
Human *SMCX* RT-PCR (438bp product) HSMCX3PF1:HSMCX3PR1 Qiagen buffer	96°C for 2'00'' 96°C for 0'30'' 60°C for 0'30'' 72°C for 1'00'' } 35 cycles

Table 5.1: Primers and conditions for the human and murine *SMCX/Smcx* RT-PCR reactions.

Figure 5.29 shows the results of the murine RT-PCR experiments in which it can be seen that the *Smcx* gene is expressed in the brain, heart, kidney, liver, muscle and spleen, but not in whole eye tissue. The quantity of RT-PCR product in each of the expressing samples appears to be different, consistent with the variable levels of expression by tissue that have previously been reported (Carrel et al., 1996; Sheardown et al., 1996). The RT-PCR results are not quantitative, but do indicate that the *Smcx* gene is widely expressed.

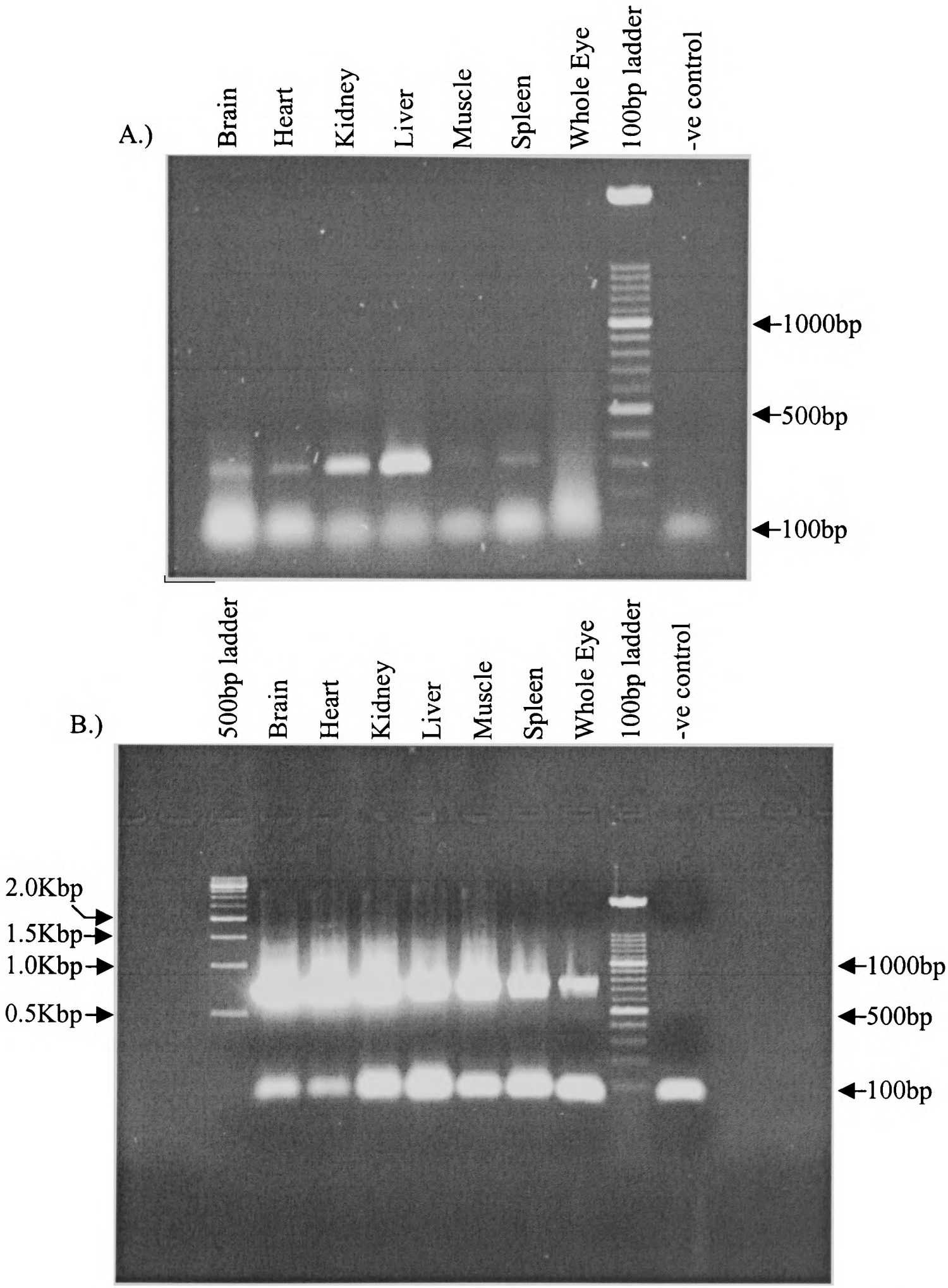
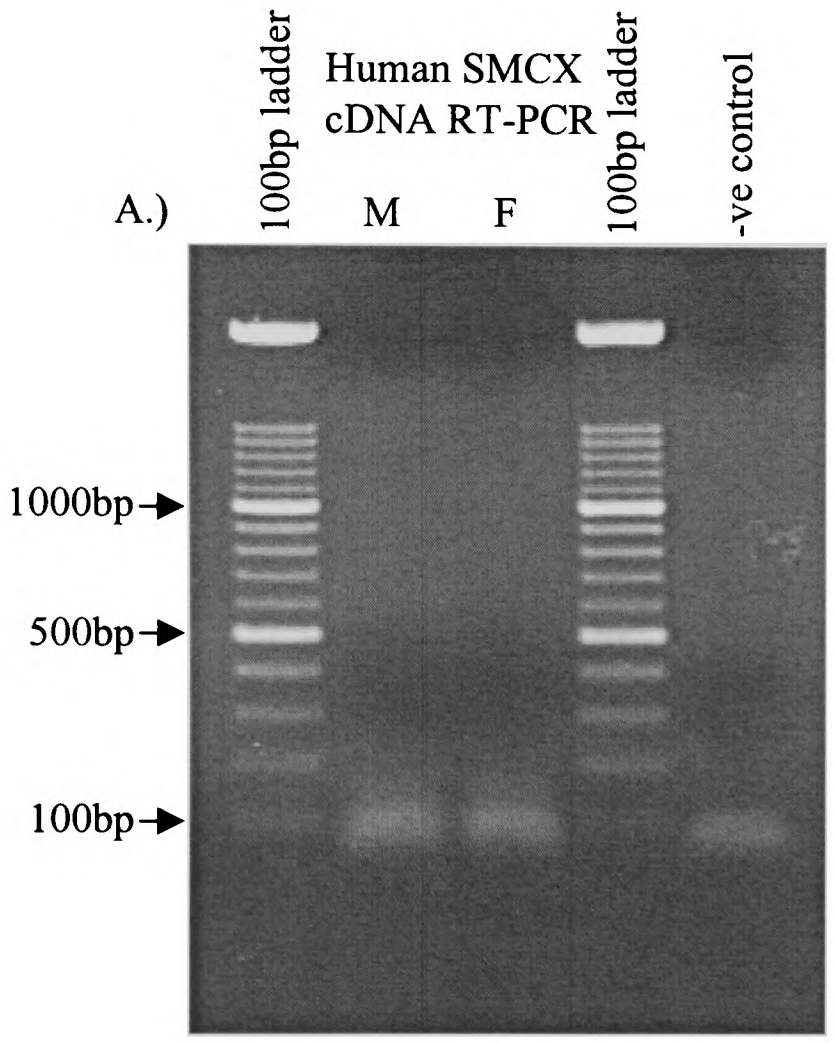


Figure 5.29: Analysis of Smcx expression in murine tissues by RT-PCR.
The expression profile of the murine Smcx gene was assessed by performing RT-PCR on RNA isolated from multiple tissues. The PCR primers MSmcx3PF1 and MSmcx3PR1 were used to amplify a 297bp product from the 3'-end of the cDNA (A). The Smcx gene appears to be expressed in many murine tissues including the brain, heart, kidney, liver, muscle and spleen. The PCR is not quantitative, but the product yield varies considerably between samples. As a control for the amount of RNA used in each reaction, RT-PCR was performed on the same multiple tissueRNA samples, but using cDNA primers to the ubiquitously expressed Pdha1 gene (B). The yield of PCR product in the Pdha1 RT-PCR reactions is much greater then for the Smcx gene, but appears far more uniform. This suggests that the variable yields seen for Smcx were reflecting the level of Smcx expression rather than the variable yield of the RNA preparations.

The apparent lack of expression of the *Smcx* gene in the whole eye was of interest because the predicted protein sequence of the gene is most similar to the retinoblastoma binding protein-2 (see section 5.1, above). In order to further investigate whether the apparent lack of expression of *SMCX/Smcx* in the eye was genuine RT-PCR experiments were performed on human retinal tissue and male and female murine whole eye tissue. Figure 5.30 shows the results of the RT-PCR experiments on human retinal RNA which confirm that the *SMCX* gene is not expressed in either male or female adult human retina. Figure 5.31 shows the same experiment on male and female murine whole eye RNA which shows no expression in the male eye, but some expression in the female eye. This is an intriguing result and suggests that *SMCX/Smcx* is expressed in some parts of the eye, but not the retina and that this expression might be limited to females, perhaps as a result of escape from X-inactivation in the female. This possibility is discussed further in section 5.3 (below).

ii) Generation of expression constructs.

Expression constructs were made of the human *SMCX* sequence from the clone XEBE6 which contains the genomic region –1010 to +184bp. This insert was cloned into the expression vector pEGFP-1 in both orientations. The cloning strategy is shown schematically in figure 5.32A and is made possible by the different order of restriction sites in the multiple cloning sites of the vectors pZErO2 and pEGFP-1. In pZErO2 the order of the relevant cloning sites is *Hind III - Bam HI - Eco RI* and in pEGFP-1 the order is *Hind III – Eco RI – Bam HI*. The excised inserts from the pZErO2 clone XEBE6 were gel purified and 10ng of each purified fragment was used in ligation reactions with prepared pEGFP-1 vector (see Methods 2.39.3). The ligation products were transformed into TOP10 cells by electroporation and cultured overnight on selective plates (see Methods 2.37). Bacterial clones containing the correct inserts were identified by restriction digestion of rapid miniprep plasmid DNA (see



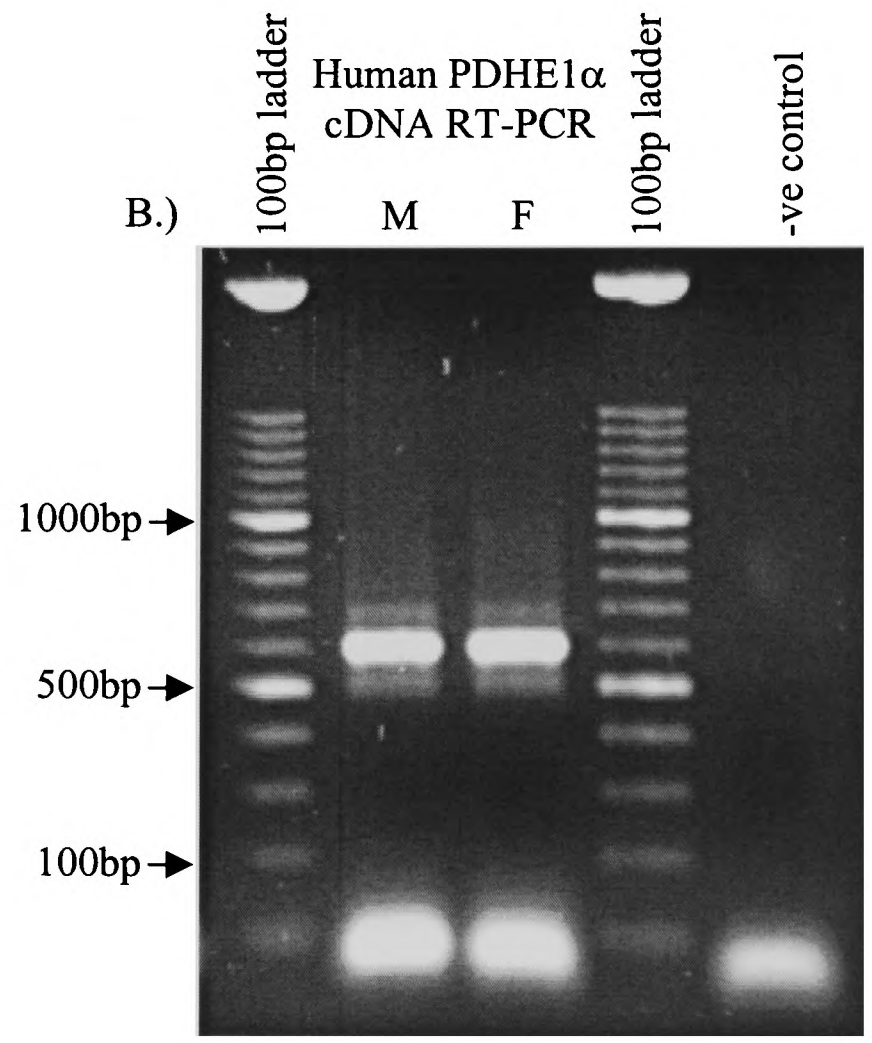


Figure 5.30: Analysis of SMCX expression in male and female retinal tissue by RT-PCR
RNA isolated from primary human retinal tissue was used as template in RT-PCR reactions to determine the expression status of the SMCX gene in the human retina. The primers HSMCX3PF1 and HSMCX3PR1 were used to amplify a 431bp PCR product from male and female retinal RNA (A). Control RT-PCR reactions for the human PDHA1 gene produced identical yields of product from both male and female retina (B). These results suggest that the SMCX gene is not expressed in the human retina. This implies that even though the protein is most similar to retinoblastoma binding protein-2 (RBBP-2) it does not share the same expression profile.

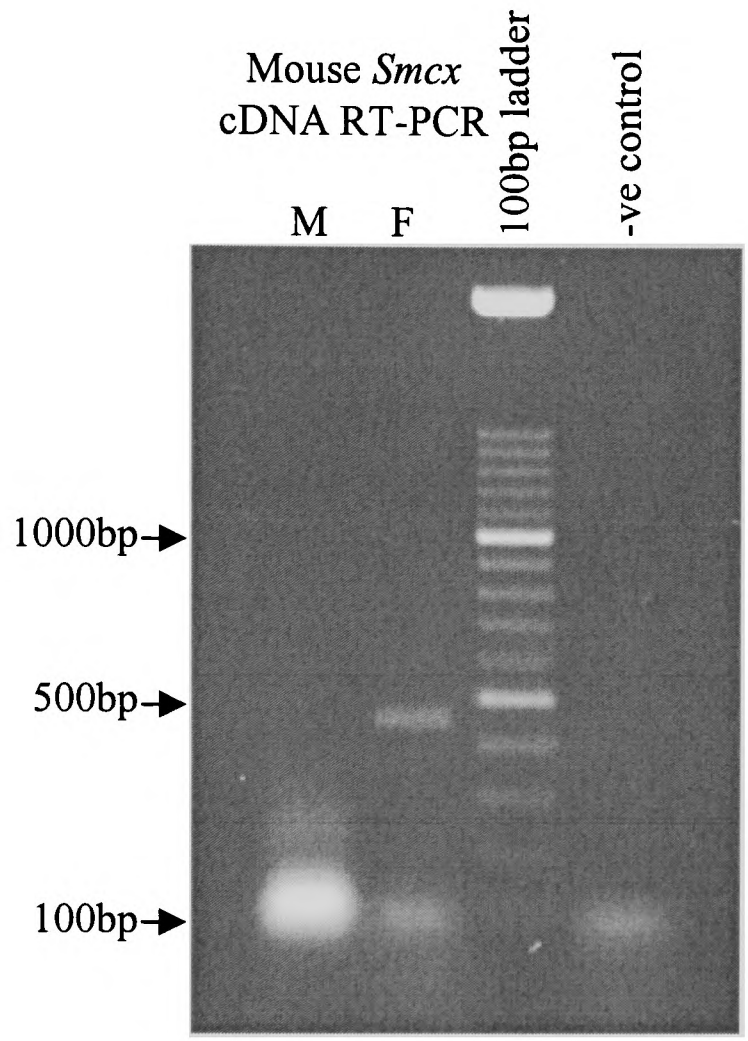
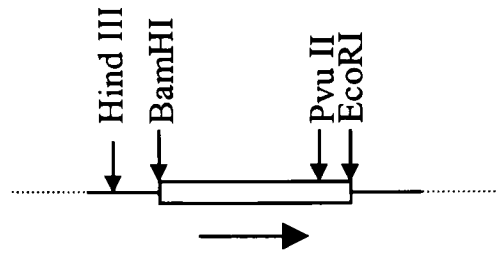


Figure 5.31: Analysis of Smcx expression in murine eye tissue by RT-PCR
RNA isolated from murine whole eyes was used as a template in RT-PCR reactions with the primers MSmcx3PF1 and MSmcx3PR1 to determine whether the Smcx gene was expressed anywhere in the eye. Unexpectedly a PCR product was reproducibly generated in female, but not male samples. This is unusual because both male and female cells should have one copy of the gene on an active X-chromosome which should be identical between the sexes. A possible explanation for the result is that the expressed allele is actually on the inactive X-chromosome of the female. This possibility is considered further in section 5.3.

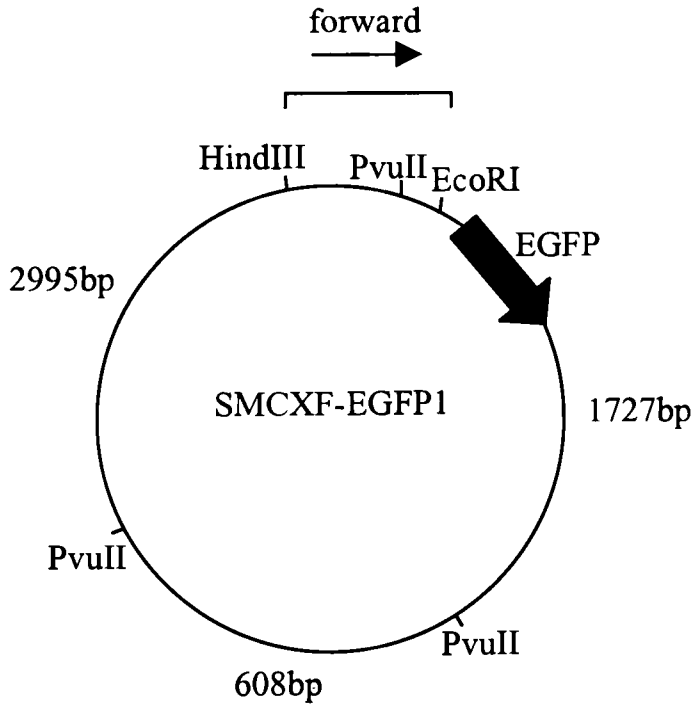
A.)



SMCX promoter-Forward-pZErO2
(XEBE6-pZErO2)

Sub-clone the 1.2Kbp SMCX upstream section into pEGFP-1 in the forward orientation by cloning into the HindIII/EcoRI sites

Sub-clone the 1.2Kbp SMCX upstream section into pEGFP-1 in the reverse orientation by cloning into the EcoRI/BamHI sites



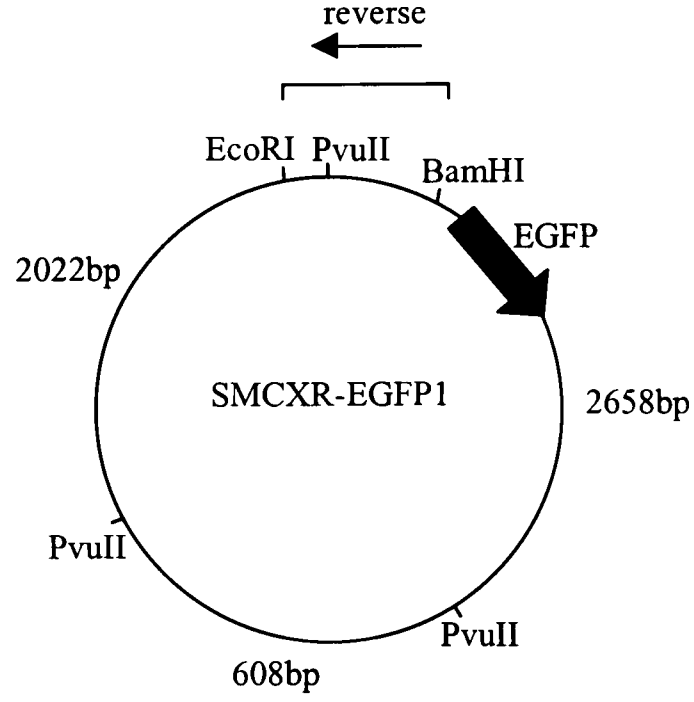


Figure 5.32: Generation of the human SMCX pEGFP-1 expression constructs.
The schematic shows the cloning strategy to make forward and reverse oriented constructs. The 1.2Kbp BamHI/EcoRI fragment containing the region -1010 to +184bp was excised from the vector XEBE6 with either the flanking enzymes HindIII/EcoRI or EcoRI/BamHI and these were cloned into prepared pEGFP-1 vector (A). The order of MCS enzyme sites in the pZeRO2 and pEGFP-1 vectors is different thereby allowing both orientations to be easily constructed (see section 5.2.7(ii) for explanation).

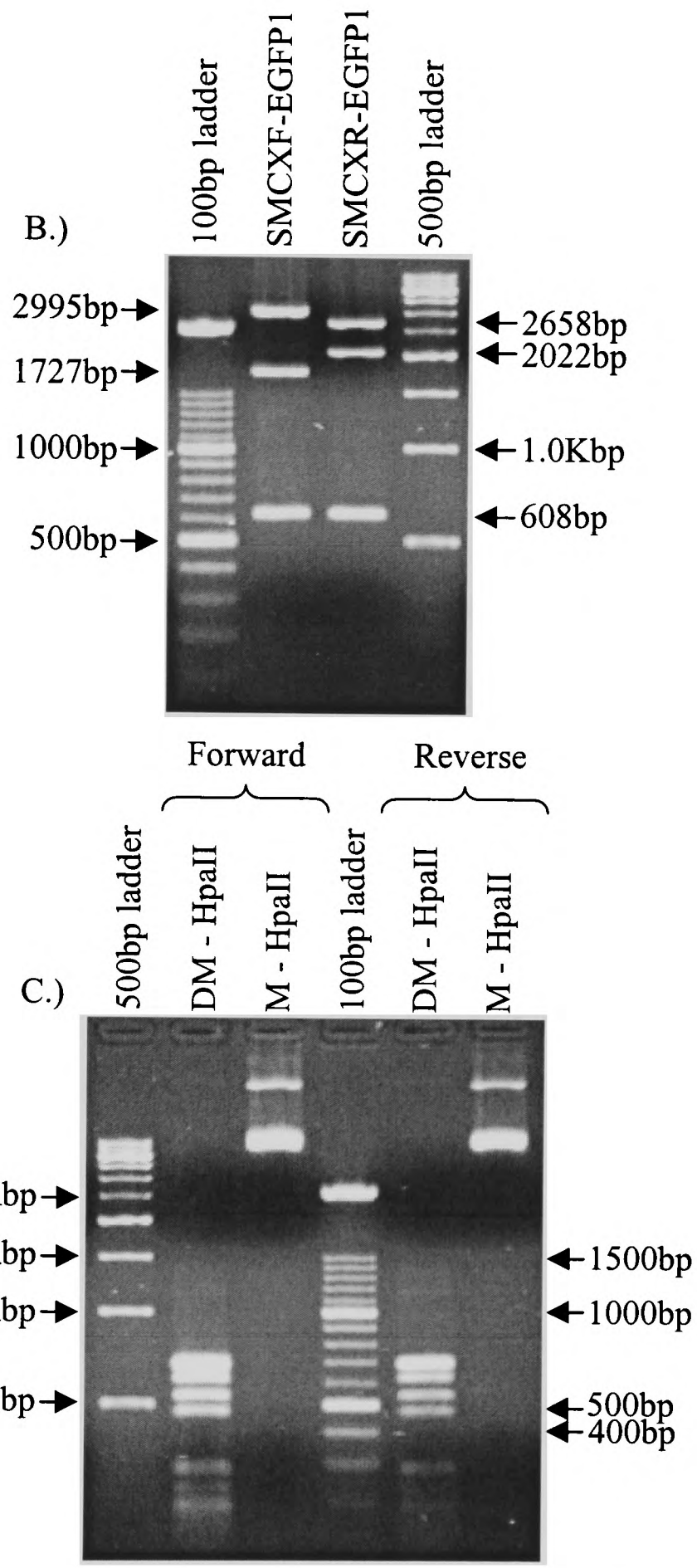


Figure 5.32: Generation of the human SMCX pEGFP-1 expression constructs (contd.) Forward and reverse orientation clones were generated as described in 5.32A. The orientation of inserts in the clones was achieved by exploiting the asymmetric position of a PvuII site in the insert (A). Briefly, rapid miniprep DNA was isolated from several clones from each cloning experiment and digested with PvuII. The reulting restriction patterns are notably different for the two possible orientations (B). A forward and a reverse clone were identified by this approach and partially sequenced to confirm that they contained the correct insert (data not shown). The selected clones, SMCXF-EGFP1 and SMCXR-EGFP1 were subsequently used in transfection experiments to assess the potential of the cloned fragment to act as a functional promoter. In order to assess the effect of methylation of the SMCX upstream sequence on reporter gene expression the vectors were artificially methylated *ex-vivo* with SSsI methylase. The methylation status of the vectors prior to transfection was determined by digesting aliquots with the methylation sensitive enzyme HpaII (C). (DM)-Dummy methylated, (M)-SSsI methylated (see section 5.2.7(iii)).

Methods 2.23). Figure 5.32B shows a restriction digest of two such clones with the enzyme *Pvu II* which generates fragments unique to forward and reverse orientation clones as the *SMCX* insert contains an asymmetrically positioned *PvuII* site as illustrated in figure 5.32A. The two clones were also partially sequenced using the XE169F1 primer to confirm that they contained the correct inserts (data not shown).

iii) Transfection of SMCX-EGFP1 expression constructs into human kidney 293 cells.

It was important to determine whether the novel upstream sequence of the human *SMCX* gene was capable of acting as a functional promoter. This was tested by transfecting the forward and reverse SMCX-EGFP1 expression constructs into human tissue culture cells. The effect of methylation on the ability of the constructs to drive reporter gene expression was also assessed by artificially methylating the plasmid DNA with the enzyme *Sss I* methylase prior to transfection (Figure 5.32C and Methods 2.48). "Dummy" methylated construct DNA (Figure 5.32C) was also transfected into cells to show that methylation itself rather than the steps involved in the artificial methylation protocol was responsible for observed differences in reporter gene expression between unmethylated and methylated constructs.

RT-PCR analysis had previously shown that the *Smcx* gene was widely expressed so it was likely that all but eye derived cell lines were suitable for the experiment. The human kidney 293 cells used in chapter 4 were selected for the transfection experiments as they were readily available, un-transformed and RT-PCR experiments had shown that the murine *Smcx* gene was expressed in the kidney (Figure 5.29).

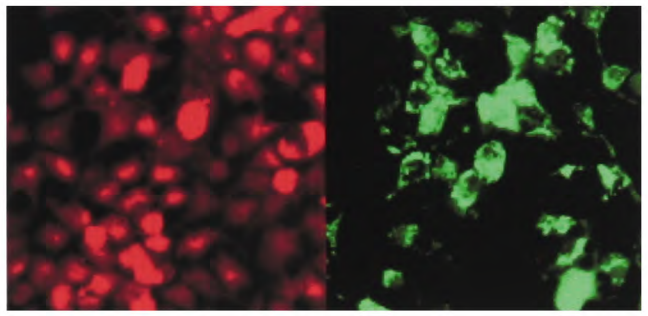
Transfection experiments were carried out as described previously (see Chapter 4 section 4.2.4(ii)) using the Fugene™6 transfection reagent (see Methods 2.51). Briefly, aliquots of

$1x10^6$ 293 cells were plated out the evening before transfection in 8-well plates containing sterile cover-slips and 3ml of culture medium (see Methods 2.50). For each transfection 2μg of plasmid DNA was mixed with prepared Fugene reagent and applied directly into the wells containing the growing 293 cells. The cells were then cultured for 48 hours prior to fixation and mounting (see Methods 2.52-54). Slides were visualised with a Zeiss (Axioskop) photo microscope under UV-illumination using filters designed for the viewing of either FITC (green) or Texas Red (red) fluorescent signals. Images were captured by a mounted CCD camera connected to a MacRed computer running the IPlab™ and IPlab extensions™ software packages. Images were taken at the same exposure and magnification to allow direct comparison of results and were stored as TIF compressed image files.

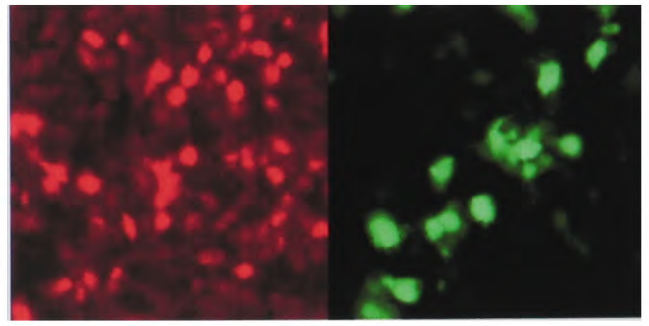
Figure 5.33 shows the results of transfecting the mtEGFP and forward SMCX-EGFP1 constructs into the human kidney 293 cells. The mtEGFP construct was used as a control for transfection efficiency in this cell line and to allow the comparison of replicate experiments (Figure 5.33A). The unmethylated and "dummy" methylated forward constructs clearly demonstrate that the section of the *SMCX* gene cloned in front of the EGFP cDNA is capable of acting as a functional promoter in these cells (Figure 5.33B and C). When the 293 cells are transfected with SMCXF-EGFP1 construct DNA methylated *ex-vivo* with the enzyme *Sss I* methylase no EGFP protein is observed (Figure 5.33D). Identical transfection experiments were also performed with the SMCXR-EGFP1 construct, but no reporter gene expression was detected in either unmethylated, "dummy" methylated or methylated constructs indicating that the *SMCX* gene promoter operates in an orientation dependent manner (data not shown).

The conclusion from these experiments is that the cloned upstream section of the human *SMCX* gene is capable of driving reporter gene activity in human kidney 293 tissue culture

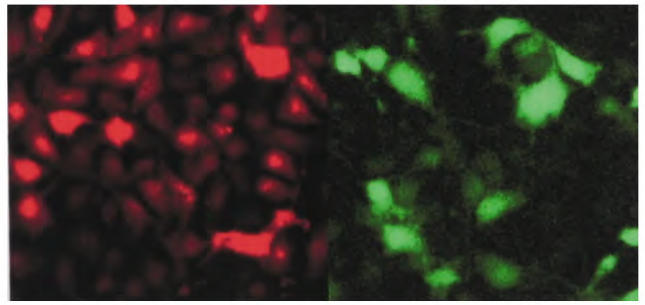
A.)



B.)



C.)



D.)

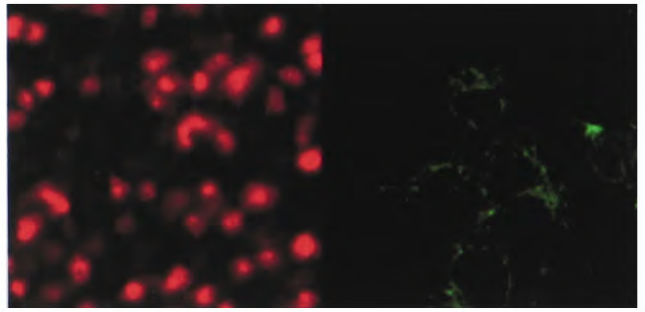


Figure 5.33: Human SMCXF-EGFP1 expression construct analysis in human kidney 293 cells.
A variety of SMCX-EGFP preparations were used in transfection experiments with the Fugene™6 reagent. The transfection efficiency in human kidney 293 cells was qualitatively assessed by transfecting cells with the mitochondrially targeted mtGFP expression construct (A) as described in section 5.2.7(iii). To assess the potential of the SMCX upstream sequence to act as a functional promoter an aliquot of purified vector was tranfected into the human kidney 293 cells using the same conditions as for the mtGFP construct (B). The image shows a high level of EGFP expression with a typical cytoplasmic signal distribution similar to that observed previously for the PDHA1-EGFP1 constructs (Chapter3). To assess the potential of methylation to affect the expression of the EGFP reporter gene from the SMCX promoter the construct was artificially methylated with SSs I methylase prior to transfection (see Figure 5.32C). As a control, another aliquot of the vector was subjected to a dummy methylation reaction then used in idnetical transfection experiments. The dummy methylated produced identical expression results to the untreated construct demonstrating that the methylation procedure itself does not affect reporter gene expression (C). The fully methylated construct did not express any EGFP signal above background, indicating that methylation of the SMCX promoter sequence switches off reporter gene expression *in-vitro*.

cells in an orientation dependent manner. This expression is ablated by the methylation of the construct *ex-vivo* suggesting that if methylation occurs at this promoter *in-vivo* it may be capable of down-regulating the genes expression. Previously published studies of *SMCX/Smcx* expression indicate that the gene escapes X-chromosome inactivation to different extents by tissue. It is tempting to speculate that tissue specific methylation patterns at the promoter of this gene might correlate with the degree of escape. This could be tested directly by bisulphite sequencing of the murine *Smcx* promoter in different tissues exactly as performed for the *Ndp* gene in chapter 4.

5.3 Discussion.

The availability of high quality gridded genomic libraries has greatly facilitated the isolation of genomic fragments. In this chapter two different kinds of gridded libraries were used to isolate genomic clones containing the upstream sections of the human and murine *SMCX/Smcx* genes. This has provided a resource for future studies of the two promoters to define *cis*-acting elements and sites of methylation *in-vivo*. The determination of the structure and regulation of the promoters of genes which escape X-inactivation will contribute greatly to our understanding of the process of escape and of the X-inactivation process as a whole. The *SMCX/Smcx* protein is one of only three non-PAR encoded protein known to escape X-inactivation in both man and mouse and as such provides a target for study in a comparative manner.

i) Isolating the human *SMCX* gene promoter.

A number of potential problems had to be considered during the design of the *SMCX* cDNA probe, Xe169c1. As with all genomic hybridisation experiments, it was not known whether

the probe would cross-hybridise to other sequences. This was a valid concern particularly with *SMCX* as it was suspected to be a transcription factor and as such was likely to contain common domains with other transcription factors. A blast search of the Genbank database with the entire human cDNA revealed regions of similarity to the Zinc-finger and Chromo-domain moieties (see section 5.1.4, above).

To minimise the possibility of cross-hybridisation to clones of other transcription factor genes, the probe was designed to a 5' section of the *SMCX* cDNA which contained no significant identity to any other sequence in the Genbank database. The possibility of isolating spurious clones was further reduced by screening an X-chromosome flow-sorted genomic library. This screening produced 8 strong positives, 3 of which were ordered from the HGMP-RC. All three clones produced identically sized positive bands when digests were probed with the isolating Xe169c1 probe (Figure 5.8). In addition the digests had different band patterns implying that they were non-identical overlapping clones (Figure 5.7). Subsequent sub-cloning from one of these cosmid clones (U34J10) generated clones XE01, XEB6 and XEB9 which after sequencing were found to contain 1010bp of upstream *SMCX* sequence (Figure 5.11).

ii) Isolating the murine *Smcx* gene promoter.

Ideally the murine *Smcx* gene promoter would have been isolation from an X-chromosome specific genomic library as for the human promoter, but no such library was available. However, a high quality gridded murine PAC library was available which had unusually been constructed using genomic DNA from a female mouse. This precluded the possibility of the Xe169c1 probe isolating clones containing the highly similar murine *Smcy* gene by cross-hybridisation to the *SMCX* derived probe. The RPCI21 murine PAC library was made by the

BACPAC resource centre and distributed by the HGMP-RC (see Materials 2.11). The average insert size in the library is approximately 145kb so any clones isolated with Xe169c1 probe were likely to contain sizeable amounts of novel upstream sequence.

A potential disadvantage of using PAC clones to isolate smaller fragments is that due to their large insert size PACs produce complex restriction patterns with the enzymes commonly used in sub-cloning experiments (e.g. Figure 5.14). As a result it can be difficult to specifically sub-clone a fragment of choice especially when other fragments in the digest are of similar size. Somewhat counter-intuitively this is why it is frequently desirable to sub-clone fragments in the 3-5kb or greater range as they are often more easily resolved and are less numerous than the smaller fragments. This is illustrated well in figure 5.16 which shows the preparative digest of clone 600j10 with *Eco RI*. In this digest there are at least 13 fragments in the size range 0.5-2.5kb and only 3 in the size range 3-5kb. Clearly this will vary by genomic interval, but as a general rule was observed in the sub-cloning experiments carried out during the course of this thesis.

Digests of the various PAC clones were hybridised to the Xe169c1 probe to identify which fragments contained the 5' end of the gene. A 4.6kb *Eco RI* and a 750bp *Bam HI* were shown to be Xe169 positive. The 4.6kb Eco RI fragment was sub-cloned from PAC 600j10 into pZErO2 producing clone cl19. This clone was digested and hybridised again with the Xe169c1 probe showing that it contained within it the 750bp *Bam HI* fragment. This fragment was sub-cloned from cl19 and sequenced entirely as the fastest means of generating novel upstream sequence (Figure 5.19). Gap alignment of this against the previously isolated human sequence demonstrated that only 23bp of upstream sequence was present in the 750bp *Bam HI* clone. As a result, cl19 DNA was used as template in sequencing reactions with a

primer designed to point upstream from the most 5' known murine sequence. This resulted in the generation of an additional 528bp of upstream sequence, but more could be readily obtained by further sequencing of the cl19 clone or of other upstream sub-clones from the PAC 600j10.

iii) Comparison of the human and murine proximal upstream sequences.

The power of comparative analysis between promoter pairs is in the assumption that conserved sections of sequence are of functional or structural significance to the expression of the gene. The validity of this assumption can only be tested by suitable experimentation. Many comparative promoter studies have been published including recent studies of the Norrie disease protein (Kenyon and Craig, 1999), adenosine deaminase (Brickner et al., 1999), lipoprotein lipase (Bey et al., 1998), transferrin (Chaudhary and Skinner, 1998), tuberous sclerosis-2 (Kobayashi et al., 1997), L-plastin (Lin et al., 1997), fibroblast growth factor receptor-3 (PerezCastro et al., 1997) and oxytocin (Young et al., 1996) genes. All of these studies confirm that the comparative approach can be extremely useful in identifying potential structural and regulatory elements of gene promoters.

Comparative analysis of promoter sequences although frequently productive can also be misleading. It is possible that sequences could be conserved between homologous promoters when they have no structural or functional role. Conversely sequences could appear to share no identity and yet harbour elements performing the same function (an example of convergent sequence evolution). Finally, elements may have evolved differentially by species that are critical in the control of the genes, but which could never be defined by comparative sequence analysis.

Promoters are composed of many elements some of which are involved directly in the initiation of transcription others are required for regulatory roles involving the spatial and temporal control of gene expression. In the course of divergence between species, spontaneous mutation will alter DNA sequences including promoter sequences. If the amount of mutation was high enough over this time period then the expectation would be that essential promoter elements would remain whereas other non-essential sequences would be mutated. If this were the case then comparison of promoter sequences between the two species would be extremely informative with the conserved sections likely to be of importance to the control of gene expression. If, however, the amount of mutation is lower then it is likely that sections of the two promoters would remain similar irrespective of their role in the control of gene expression. It is likely that most gene promoters are juxtaposed somewhere between these two extremes. These ideas are significant because they imply that promoters or sections of promoters that are susceptible to higher rates of mutation are better suited to this kind of analysis than those that are not.

Methylation of genomic DNA causes an increase in the rate of spontaneous mutation by as much as nine-fold and has been used to explain the over-representation of disease causing point mutations at CpG sites (Bird, 1980). Methylation at the promoters of genes would therefore increase the rate of mutation at certain sites within that promoter. The active and inactive X-chromosomes share comparable levels of methylation (Bernardino et al., 1996), but the inactive X-chromosome has methylation concentrated at the promoters of many genes unlike autosomal genes (Antequera and Bird, 1993; Norris et al., 1991). This means that promoters of genes on the X-chromosome which are subject to methylation would experience increased relative rates of mutation and therefore many would be well suited to comparative promoter sequence analysis.

The human and murine *SMCX/Smcx* gene promoters only show high levels of similarity very close to the start of transcription (Figures 5.25 and 5.27). The first 132bp of upstream sequence are 82% identical consistent with the idea that this region includes the core promoters of the genes. The lack of similarity upstream of –132bp could mean that these promoters are subject to high rates of mutation. The observation of a small conserved section of sequence around –500bp (Figure 5.26) could therefore be significant as it is conserved amongst a region of little similarity. There are no sequences similar to the TATA or CAAT boxes commonly found at the promoters of tissue-specific genes consistent with the observation that this gene is widely expressed (Figure 5.29). In total eight conserved transcription factor binding sites were found including three putative Sp1 sites, Pit-1, NRF-1, LBP-1, WAS-US6 and H4TF2 sites. The transcription factors NRF1 and LBP1 are ubiquitously expressed and found in the promoters of genes involved in growth and metabolic control (Novotny et al., 1998). The transcription factor Sp1 is ubiquitously expressed and a potent transcriptional activator found in the promoters of most housekeeping genes. The H4TF2 transcription factor is found commonly in the promoters of several histone protein genes, its role at the promoters of other genes is unclear, but it has been implicated in up-regulating gene expression prior to differentiation (Labella and Heintz, 1991). The Pit-1 gene is expressed in a tissue specific manner directing expression of the prolactin (PRL) and growth hormone (GH) genes in the anterior pituitary gland (Vila et al., 1993). The WAS-US6 has not been described in the literature, but was conserved in the two promoters and identified in the TRANSFAC database. Remarkably three of the conserved transcription factor binding sites are of transcription factors which activate genes involved in the control of growth; four of the remaining five conserved sites are potent transcriptional activators which are ubiquitously expressed. The indication from this comparative analysis is that the

SMCX/Smcx genes which have been proposed to be transcription factors themselves are involved in growth processes. The observation that the *Smcy* gene is expressed as early as the 2-cell embryo strongly suggests that the *SMCX/SMCY* and *Smcx/Smcy* genes play a critical role during the earliest stages of growth and development.

The level of amino-acid identity between the X-linked and Y-linked copies of these genes is so high (95% for human, 91% for mouse) that they have been presumed to be functionally identical although this has yet to be formally demonstrated. If they are functionally equivalent then males would express two copies and females only one should the gene be subject to X-inactivation, resulting unusually in a case of haplo-insufficiency on behalf of the female. The *Smcx* gene partially escapes X-inactivation with increased escape through ontogeny so that early in development the gene is subject to inactivation. Dosage equivalence is maintained between the sexes at later stages during development with reports that the degree of escape can be as high as 70-80% in some tissues (Carrel et al., 1996; Sheardown et al., 1996).

The differential control of *Smcx* gene expression manifest as variable degree of escape from inactivation must have a molecular basis. The tight inactivation of the gene early in development results in haplo-insufficiency in females which would be most prominent midway through embryonic development. The subsequent variable release of this control will result in clear sex specific differences in gene expression at key points during ontogeny. An excellent molecular candidate for this complex and re-programmable control is DNA methylation at the *SMCX/Smcx* gene promoters. The comparative analysis of the promoters in this chapter not only detailed conserved transcription factor sites, but also conserved sites of potential methylation. There are eleven conserved CpG sites in the proximal promoter seven of which saturate the conserved transcription factor sites. Reporter gene constructs of the

SMCX promoter demonstrated that *ex-vivo* methylation of these CpG sites was capable of switching off the promoter in tissue culture cells. The isolation of the human and murine *SMCX/Smcx* promoters reported in this chapter provides the necessary resources to test the relationship between escape from inactivation and promoter methylation for these genes *in-vivo*.

RT-PCR experiments were used to demonstrate the wide expression profile of the *Smcx* gene in primary murine tissues. This produced an unexpected result in murine male and female whole eye tissue because the female, but not the male sample produced a product band (Figure 5.31). Control bands looked identical between the two samples so differences in cDNA quantity are unlikely to explain the result. Further investigations are needed to determine whether these RT-PCR results are correct, although they have been consistently reproduced. These results, if validated, could prove important to understanding the control of gene expression in the context of the X-inactivation process.

Constructs containing the novel human sequence upstream of a reporter gene when transfected into human kidney 293 cells showed high levels of reporter gene expression indicating that the cloned sequence was a functional promoter. Methylation of the construct *ex-vivo* with *Sss I* methylase ablated this function suggesting that methylation of the promoter *in-vivo* may be capable of regulating the expression of the gene. This system of assessing the effect of methylation on promoter sequences is now well established and was discussed in chapter 4. The murine promoter was not used in identical reporter gene construct experiments because the only murine cell line available at the time was the 661W retinal cell line used in chapter 4. The unusual RT-PCR results described for the murine *Smcx* gene above in eye tissue raised the possibility that eye derived cell lines were not suitable for transfection

experiments and that cells from any of the other *Smcx* expressing tissue types would be preferable. Confirmation that the novel murine sequence is also a functional promoter is required.

Future work will involve the detailed analysis of promoter methylation of the human and murine gene promoters and correlation to degree of escape in a model system such as the Searle's translocation mouse. *In-vitro* dissection of the *cis*-acting elements of the two promoters will be carried out in suitable tissue-culture cell lines with a deletion arrays of reporter gene constructs. *DNase I* hypersensitivity mapping and *in-vivo* footprinting of the promoters will determine changes in hypersensitive sites and bound proteins in different tissues. In conclusion the *SMCX/Smcx* genes are targets for the definition of properties involved in escape from the X-inactivation process. Critically, because the murine homologue of an escaping human gene also escapes inactivation, studies in the mouse are possible.

Chapter 6 - General Discussion

The analysis of the human Pyruvate dehydrogenase E1α (*PDHA1*) and the human/murine Norrie disease protein (*NDP/Ndp*) gene promoters in this thesis generated detailed methylation profiles from a variety of genomic DNA sources. These data were interpreted in the context of existing literature and were found to be consistent with a number of possible roles for methylation in the control of gene expression on the X-chromosome. Genomic PAC clones spanning the *Pdha1* gene were isolated to allow the future cloning of the murine promoter so that sequence and methylation profiles can be compared to those of the human gene. The promoters of the human and murine *SMCX/Smcx* genes were cloned and sequenced to allow similar methylation studies to be performed on a gene that escapes the X-inactivation process. The *Smcx* promoter sequences were compared to identify sections of identity, methylatable sites and conserved consensus binding sites for known transcription factors. Promoter-reporter gene constructs for all three sets of gene sequences were transfected into tissue culture cells to demonstrate that they could act as functional promoters. These constructs were also artificially methylated to assess the potential of DNA methylation to repress their promoter activities.

6.1 General features of promoter comparison and sequence analysis.

The promoter sequence analyses of the *PDHA1*, *NDP/Ndp* and *SMCX/Smcx* genes in this thesis highlighted some important features of gene promoters in general as well as some features specific to promoters on the X-chromosome. The promoters of the *PDHA1* and *SMCX/Smcx* genes were found to contain consensus transcription factor (TF) binding sites for regulatory elements commonly associated with widely expressed genes such as AP-2 and

SP1. Conversely the promoters of the *NDP/Ndp* genes did not contain elements associated with widely expressed genes, but contained a putative TATA box. In addition the promoters were found to have their most highly conserved sections (*NDP/Ndp* 90%, *SMCX/Smcx* 82%) within the first one to two hundred base pairs of their transcription start sites. The promoters of the *PDHA1* and *Smcx* genes had a CpG-island, unlike the *SMCX* promoter which had substantial sections of GC-rich sequence but missed the CpG-island criteria. The *NDP/Ndp* gene promoters did not have a CpG-island, but again had several sections of high GC-content. The majority of housekeeping genes and less than half of tissue-specific genes have CpG-islands at their 5′-ends, so the promoters of these X-linked genes reflect the CpG-island associations in the genome as a whole.

The CpG-islands that are predominantly associated with promoters of housekeeping genes are hypo-methylated on autosomes and hyper-methylated on the inactive X-chromosome (Norris et al., 1991). The murine genome has 20% less CpG-islands than the human (Antequera and Bird, 1993). This difference has been attributed to the shorter generation time of the mouse and the increased rate of mutation caused by cytosine methylation. It seems likely that, in the presence of promoter methylation on the inactive X-chromosome, the promoters of X-linked genes are subject to higher rates of mutation at CpG-sites than their autosomal counterparts. As was discussed in Chapters 4 and 5 when the murine *Ndp* and *Smcx* promoter sequences were compared to the human they were found to have fewer CpG-sites consistent with this idea (see Figures 4.34 and 5.28). It also appeared that the sequence differences within the highly conserved sections of the proximal promoter regions were predominantly at sites in the mouse where CpG di-nucleotides lay in the corresponding human promoter. This could have profound implications to the way in which X-linked gene promoters evolve and the selection of transcription factors used in the regulation of X-linked

genes compared to autosomal genes. A more systematic analysis of large numbers of X-linked and autosomal gene promoters would reveal whether these observations are significant, but at present the number of sequences available in the genome databases is too limited. This will certainly change in the next few years as the human genome project is completed.

6.2 Interpreting methylation profiles.

The first decision that has to be made when characterising genomic methylation is which method to use. In practice it is simpler and faster to use methylation sensitive restriction enzyme based approaches, but these are extremely limited in the number of sites they can assess and are unreliable. As has been discussed one of the principle concerns when using these approaches is that complete digestion of unmethylated restriction enzyme target sites is difficult to achieve. This is compounded by the difficulty of designing positive controls for complete digestion of genomic DNA within each experiment. The numerous methods available for studying methylation were described in Chapter 1 (see section 1.16) each having their own advantages and disadvantages. Bisulphite sequencing was the most widely used and refined method available to assess methylation at nucleotide resolution and was chosen for use in this project.

Many studies have established that methylation profiles from tissue culture cells are different to those of primary tissues (Saluz and Jost, 1993) (and references therein). This is supported by the methylation studies at the *PDHA1*, *NDP* and *Ndp* gene promoters in this thesis. The methylation profiles at the human *NDP* promoter between WeriRb1 cells and primary retinal tissue (see Figure 4.11) revealed striking differences. The tissue culture cells were found to have a high level of methylation (≈100%) on the presumed inactive X-chromosome in

contrast to retinal tissue which was only lightly methylated (≈40%). In addition the presumed active X-chromosome in WeriRb1 cells was lightly methylated (≈30%) in some sequences whereas the retinal tissue was found to be unmethylated in all. The increased levels of methylation in tissue culture cells could be the result of high passage cell culture which mimics the increase in methylation at lightly methylated genomic sites that is generally observed during development (Bestor and Tycko, 1996). Alternatively the increased methylation could be part of the mechanism of the cell transformation process in which dynamic changes in methylation profiles have been widely reported in cancer cell lines when compared to untransformed cell lines of the same individual (Bird, 1996). As was discussed in Chapter 4, the WeriRb1 cell line had a polymorphism within the region of study between the X-chromosomes which segregated in an ordered manner between classes of methylation profile. The highly methylated sequences always associated with one allele whereas the unmethylated and partially methylated sequences always associated with the other. This was consistent with the cancer line being clonally derived and the methylation patterns faithfully inherited by maintenance methylation. This could not explain the third partially methylated set of sequences which were associated with the allele on the presumed active X-chromosome. It was proposed that this third class of methylation sequence was the consequence of de-differentiation of the transformed cells in which retinal specific gene expression would be down-regulated.

Methylation sequences from the *PDHA1* promoter in HeLa cells and the *NDP* promoter in WeriRb1 cells were homogeneous whereas those at the *Ndp* promoter in murine primary tissues were extremely heterogeneous. These differences in methylation between tissue culture cells and primary tissues may be explained by heterogeneity within the samples themselves. It is likely that genomic DNA isolated from individual organs or tissues will be

derived from many different cell types having different, but related methylation patterns. Tissue culture cells on the other hand are often derived from clonal sources (this is likely to be the case for both HeLa and WeriRb1 cells) and would be expected to have more stable homogeneous methylation profiles. The murine brain, heart, kidney, liver, muscle, spleen and whole eye tissue samples would certainly have been composed of many different cell types with markedly different gene expression profiles. The human retinal tissue on the other hand was almost exclusively comprised of retinal epithelium although a small amount of the neuro-epithelial layers would have survived the dissection (personal communication Dr. Sally Boulton). It is reasonable to argue that the purer the population of cells used in bisulphite methylation analysis (be they tissue culture or primary) the more homogeneous the methylation profiles are likely to be. However, none of the work reported here or in any other published studies using methylation sequencing have used purified cell-types from primary tissues in their analyses. Without studying methylation in such samples it would be unwise to argue that the heterogeneous methylation observed in primary tissues was the result of differential methylation within cells of the same type. It is important to note that the heterogeneity observed in profiles derived from primary tissues has been used to refute the significance of site specific methylation in gene repression. The human retinal tissue in this thesis was nearly all retinal epithelium (one cell type) and produced homogeneous methylation profiles supporting the possibility that methylation is not as heterogeneous in primary tissues as is currently thought. Further studies should aim to characterise methylation in pure populations of primary cells to resolve this issue as it is critical to the understanding of the role of methylation *in-vivo*.

Other specific results from the methylation analyses at the *PDHA1* and *NDP/Ndp* gene promoters are suggestive of a site-specific role for methylation in gene repression. The CREB

site in the *PDHA1* promoter lies within a putative c-fos/c-jun site and has been shown to bind nuclear proteins from cells expressing the *PDHA1* gene and to bind recombinant c-fos/c-jun heterodimers in DNase I footprinting experiments (Chang et al., 1993). This site is saturated with 6 CpG di-nucleotides all of which are methylated on the presumed inactive X-chromosome of HeLa cells. Methylation of CRE sites has been shown to repress protein binding and to repress CRE regulated gene expression (Iguchi-Ariga and Schaffner, 1989). These results taken together support the idea that specific methylation across the CRE binding site of the *PDHA1* promoter in HeLa cells would be capable of preventing transcription factor binding and down-regulating *PDHA1* expression *in-vivo*. The promoter of the *NDP/Ndp* genes contains a conserved binding site for the NRL transcription factor which is very closely related to c-fos and c-jun and is capable of making heterodimers with them (Kerppola and Curran, 1994). The NRL site is saturated with 4 CpG di-nucleotides (see Figure 4.34) which are methylated on the presumed inactive X-chromosome of WeriRb1 cells, human retinal cells and in all non-expressing murine tissues studied (see Figures 4.11 and 4.27). The third class of methylation sequence derived from WeriRb1 cells is only methylated across the NRL site (see Figure 4.6). If this site is critical in the control of *NDP* gene expression then it is possible that this light methylation is concentrated at the precise sites required to achieve down-regulation of *NDP* on the active X-chromosome, as part of the de-differentiation process. In summary, the distribution of methylation across these transcription factor binding coincides with sites that are implicated in the control of the specific genes expression. It is plausible that this methylation is coincidental or consequential to gene repression, but the evidence presented here and elsewhere are extremely supportive of a functional role for methylation in gene repression at both specific sites and through action distributed across local regions.

Analysing methylation data is extremely difficult. There are a large number of variables involved when working with both tissue culture cells and primary tissues. The first problems in interpreting methylation data in this thesis were when methylation was assessed at the *PDHA1* promoter in female lymphocytes. Indications from msPCR experiments had suggested that at least 60% of the samples were methylated across one region of the promoter, but no bisulphite sequences were produced that had methylated sites. This discordance was rigorously investigated, but no explanation could be provided other than incomplete digestion of genomic DNA in some of the msPCR experiments. Certainly contaminants from blood DNA extractions can interfere with restriction digestion of DNA, PCR and cloning experiments, but no convincing evidence of this existed in these experiments. There was also no evidence that the bisulphite PCR itself was biased between methylated and unmethylated copies as both classes of sequence derived from HeLa cells were equally represented. The very limited number of methylation sequencing studies that have been published using DNA from lymphocytes make it impossible to comment on the validity of these results. The only paper that has mentioned lymphocyte profiles commented that they were heterogeneous and hypo-methylated when compared to other samples (Pfeifer et al., 1990). The data in this thesis do not prove that there is no methylation at the female *PDHA1* promoter in lymphocytes. Importantly a standard hybridisation experiment using methylation sensitive restriction enzymes would not necessarily resolve this issue as complete digestion cannot be guaranteed. The best way to resolve this issue would be to analyse methylation at the *PDHA1* and *NDP* promoters in a variety of human tissues using the same genomic DNA samples. The human *NDP* bisulphite protocol has already been used successfully to determine methylation profiles in primary tissue and would therefore act as a control for contamination effects on the *PDHA1* analysis.

The profiles generated from murine tissues clearly illustrate the difficulty of dissecting out profiles and assigning these to the active and inactive X-chromosomes in expressing and non-expressing tissues (see Figure 4.27). The variation in the murine *Ndp* bisulphite sequences in primary tissues and the generally high level of methylation across the region made this task particularly difficult. The expectation was that the *NDP/Ndp* gene promoters would be heavily methylated on the inactive X-chromosome and on the active X-chromosome in non-expressing tissues and would be unmethylated on the active X-chromosome in expressing tissues. This was clearly not the case. Instead the promoter was always methylated, but was hypo-methylated at the proximal promoter in some sequences from expressing tissues. These results were supported by the recently published analyses of the tissue transglutaminase (Lu and Davies, 1997) and galectin-1 (Salvatore et al., 1998) promoters. In these studies the whole nature of the methylation profiles was different to those of previous studies in which analysis has primarily been with DNA from tissue culture cells. Methylation was found to be distributed across the whole promoter region in all tissues but the proximal promoter was hypo-methylated in sequences of expressing tissues exactly as for the murine *Ndp* analysis in this thesis. Although these three studies provide results for tissue specific genes which are strikingly similar many more methylation studies need to be performed at the promoters of tissue specific genes before a strong relationship can be established between methylation and tissue specific repression.

The remaining problem in interpreting methylation profiles in this project was the assignation of the profiles to either the active or inactive X-chromosome. In every paper published to date the assignation of a methylation sequence to either chromosome has been made on the basis that the methylation sequence obtained from male cells and tissues is from the active X-chromosome. The sequences from female material are then compared to the male and those

that are most similar are assigned to the active and the remaining sequences assigned to the inactive X-chromosome. This approach is entirely appropriate for male samples, but is only arguably so for female ones. There is no published method to perform methylation sequencing so that the active and inactive X-chromosome sequences can be distinguished, but I propose one here. The Searle's translocation mouse has been described in detail elsewhere in this thesis (see section 5.1.2 and figure 5.3). Mice inheriting Searle's translocation have the same X-chromosome inactive in every cell (Lyon et al., 1964; Takagi, 1980). If a polymorphism is present or introduced (for example by inter sub-species breeding) at the X-linked gene to be subject to methylation analysis then it would be possible to assign the methylation profiles to the active or inactive X-chromosome. A concern would have to remain about the effects of the translocation on the two separated active X-chromosome fragments, but this approach could in principle provide a wealth of information from primary tissue and greatly simplify methylation analysis on the X-chromosome.

6.3 Assessing the effects of methylation on gene transcription.

The effect of methylation on gene transcription can be addressed *in-vitro* or *in-vivo*. A simple *in-vitro* approach has been used in this project in which promoter-reporter gene constructs were artificially methylated and transfected into tissue culture cells. For each gene promoter studied artificial methylation was found to down-regulate the expression of the reporter gene to a background level. The main concern about the validity of these results lay in the effect that methylation of flanking vector sequences might have on the expression of the reporter gene. As has been discussed in chapter 4, the mechanisms through which such an effect could be achieved are not thought to be in operation for transiently transfected non-replicating expression constructs. Transcription repression by DNA methylation is mediated through MeCP2 and other MBD proteins (see section 1.15). It seems likely that methylation of

promoters in reporter gene constructs would recruit proteins such as MeCP2 to occlude transcription factor binding sites and repress transcription (Nan et al., 1997). These experiments do not recapitulate gene control as it would occur at the endogenous locus, but they do demonstrate that methylation is capable of directing the repression of transcription at these gene promoters. Whether the same mechanism acts to achieve this *in-vivo* cannot be determined without perturbing the system through transgenesis.

6.4 Methylation and X-chromosome inactivation.

The methylation analysis at the promoter of the *PDHA1* gene has confirmed the prevailing concept that promoters of housekeeping genes are hyper-methylated on the inactive X-chromosome and hypo-methylated on the active X-chromosome. In total three such genes have had their methylation profiles characterised at nucleotide resolution, *PDHA1*, *HPRT* (Hornstra and Yang, 1994; Park and Chapman, 1994) and *PGK-1* (Pfeifer et al., 1990). The study has also confirmed that such promoters are not always totally methylated on the inactive X-chromosome as was originally suggested by some researchers (Hornstra and Yang, 1994; Pfeifer et al., 1990). The level of methylation on the presumed inactive X-chromosomes for the three genes are *PDHA1* (65%), *HPRT* (90% and 69%) and *PGK-1* (100%) so even though the promoters are structurally similar they have different methylation densities. This is likely to vary to some extent by tissue and cell-type, but nevertheless the conclusion remains the same and is in agreement with other independent methods of analysing methylation at the promoters of X-linked genes (Norris et al., 1991).

The methylation analyses at the promoters of the human and murine *NDP/Ndp* genes are important because they represent the first nucleotide resolution studies of promoter methylation at tissue specific genes on the X-chromosome. The first finding is that the human

NDP promoter is hyper-methylated on the inactive X-chromosome and hypo-methylated on the active X-chromosome in human retinal tissue and WeriRb1 cells. The murine *Ndp* promoter is hyper-methylated on both the active and inactive X-chromosomes, but hypo-methylation is broadly observed in expressing tissue types and is presumed to be on the active X-chromosome. Interestingly the active X-chromosome of male tissue samples are hyper-methylated at the *Ndp* promoter in the tissues studied including the heart and kidney in which some proximally hypo-methylated sequences were expected. The absence of such sequences is most likely due to the small number of male clones that were sequenced and the heterogeneity of cell type within the tissues as discussed in chapter 4. These results are significant because they imply that the correlation between methylation and tissue-specific repression could be maintained in the context of X-inactivation. In addition it adds weight to the importance of methylation in the X-inactivation process as a whole because it establishes the fact that heavy methylation can occur at the promoters of tissue-specific genes which do not have CpG-islands at their 5′ ends. As such this raises the possibility that methylation acts as a chromosome wide repression mechanism which can act wherever there are potentially methylatable sites.

Although methylation studies have not yet been performed at the novel promoters of the *SMCX/Smcx* genes their study can be expected to provide invaluable information about the mechanism of escape from inactivation. This is especially true of the *Smcx* gene because it is the only gene known to escape inactivation in the mouse (Agulnik et al., 1994; Wu et al., 1994) (with the exception of *Xist* which is a special case). It is therefore possible to study methylation of the *Smcx* gene in primary tissues throughout development. This is of particular importance as the *Smcx* gene has been proposed to escape inactivation differentially by tissue and through development (Carrel et al., 1996; Sheardown et al., 1996). If there is a

correlation between the degree and timing of escape from inactivation and methylation at the promoter of the *Smcx* gene it may be possible to use these data to isolate sections of the promoter required for escape to take place.

6.5 Future studies.

(i) General work

It would be interesting to establish whether Searle's translocation mice can be used in future methylation analyses to unequivocally assign a particular methylation profile to either the active or inactive X-chromosome. This method would greatly simplify the analysis of methylation profiles and should be applicable to practically all genes on the X-chromosome.

(ii) The *PDHA1/Pdha1* genes

The murine *Pdha1* promoter sequence could be cloned and sequenced from the genomic PAC clones isolated earlier in this project. The sequence could be used in comparative sequence analysis to identify conserved sites of potential methylation and transcription factor consensus binding sites. Bisulphite methylation analysis could then be performed on several primary tissues of the mouse to corroborate the studies from the human and determine if there is variation in promoter hyper-methylation on the inactive X-chromosome.

(iii) The *NDP/Ndp* genes

The methylation studies should be repeated in the Searle's translocation model system if it proves to be suitable (6.5(i), above). The possibility of further dissecting and dissociating tissues to isolate purer cell populations for study could be investigated. This would determine whether heterogeneity in methylation patterns is intrinsic to each cell or a feature of populations of cells. It should be possible to study the effect of methylation on the putative

NRL binding site by using recombinant NRL protein in DNase I sensitivity and gel-retardation experiments. The *NDP* promoter may prove to be a model system to study methylation mediated occlusion of transcription factor binding.

(iv) The *SMCX/Smcx* genes

It should now be possible to study the methylation profiles of the murine *Smcx* gene in multiple tissues throughout development either in wild-type mice or in the Searle's translocation model. If the promoter proves to be methylated and this correlates with degree of escape then promoter-deletion reporter gene constructs could be used to characterise elements of the promoter and relate these to the prevailing methylation profiles. These studies may lead to the isolation of sections of the promoter that are involved in the escaping mechanism of this gene.

6.6 Conclusion.

The methylation studies undertaken in this project have served to further define the precise methylation profiles associated with X-linked gene promoters and has proposed a number of ways in which these fit into our understanding of the X-inactivation process. Methylation has long been implicated in the mechanism of X-inactivation, but primarily at the level of maintenance through clonal inheritance of repressing methylation patterns at the promoters of X-linked genes. The work presented here and recent studies proving the link between methylation and the regulation of chromatin structure argues that the role of methylation is much greater. Methylation is likely to be part of a complex mechanism which can achieve X-inactivation only with the involvement of *Xist* RNA and many other components. I propose that it is very unlikely that X-inactivation can be correctly achieved *in-vivo* without genomic DNA methylation.

Bibliography

Adolph, S., and Hameister, H. (1990). In situ nick-translation of the hman metaphase chromosomes with the restriction enzymes MspI and HpaII reveals an R-band pattern. Cytogenetics and Cell Genetics *54*, 132-136.

Agulnik, A. I., Mitchell, M. J., Mattei, M. G., Borsani, G., Avner, P. A., Lerner, J. L., and Bishop, C. E. (1994). A Novel X-Gene With a Widely Transcribed Y-Linked Homolog Escapes X- Inactivation in Mouse and Human. Human Molecular Genetics *3*, 879-884.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool (BLAST). Journal of Molecular Biology *215*.

Alubaidi, M. R., Hollyfield, J. G., Overbeek, P. A., and Baehr, W. (1992). Photoreceptor Degeneration Induced By the Expression of Simian Virus- 40 Large Tumor-Antigen in the Retina of Transgenic Mice. Proceedings of the National Academy of Sciences of the United States of America *89*, 1194-1198.

Andreason, G. L., and Evans, G. A. (1988). Introduction and Expression of Dna-Molecules in Eukaryotic Cells By Electroporation. Biotechniques *6*, 650 et seq.

Antequera, F., and Bird, A. P. (1993). Number of CpG islands and genes in human and mouse. Proceeedings of the National Academy of Sciences (USA) *90*, 11995–11999.

Ariel, M., Robinson, E., McCarrey, J. R., and Cedar, H. (1995). Gamete-specific methylation correlates with imprinting of the murine *Xist* gene. Nature Genetics *9*, 312-315.

Avner, P., Heard, E., and Debrand, E. (1998). Cloning and Localization of the Murine Xpct Gene: Evidence for Complex Rearrangements during the Evolution of the Region around the Xist Gene. Genomics *48*, 296-303.

Avner, P., Prissette, M., Arnaud, D., Courtier, B., Cecchi, C., and Heard, E. (1998). Molecular correlates of the murine Xce locus. Genetical Research *72*, 217-24.

Baker, B. S., Gorman, M., and Marin, I. (1994). Dosage Compensation in Drosophila. Annual Review of Genetics *28*, 491-521.

Barr, M. L., and Bertram, E. G. (1949). A morphological distinction between neurones of the male and female and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature *163*, 676-677.

Beard, C., Li, E., and Jaenisch, R. (1995). Loss of methylation activates *Xist* in somatic, but not embryonic cells. Genes and Development *9*, 2325-2334.

Bensaada, M., Kiefer, H., Tachdjian, G., Lapierre, L. M., Cacheux, V., Niveleau, A., and Metezeau, P. (1998). Altered patterns of DNA methylation on chromosomes from leukemia cell lines: Identification of 5-methylcytosines by indirect immunodetection. Cancer Genetics and Cytogenetics *103*, 101-109.

Berger, W. (1998). Molecular dissection of Norrie disease. Acta Anat (Basel) *162*, 95-100.

Berger, W., Meindl, A., van de Pol, T. J., Cremers, F. P., Ropers, H. H., Doerner, C., Monaco, A., Bergen, A. A., Lebo, R., and Warburgh, M. (1992). Isolation of a candidate gene for Norrie disease by positional cloning. Nature Genetics *2*, 84.

Bernard, P., Gabant, P., Bahassi, E. M., and Couturier, M. (1994). Positive selection vectors using the F plasmid *ccd*B killer gene. Gene *148*, 71-74.

Bernardi, G. (1989). The isochore organisation of the human genome. Annual Review of Genetics *23*, 637-661.

Bernardino, J., Lamoliatte, E., Lombard, M., Niveleau, A., and Malfoy, B. (1996). DNA methylation of the X-chromosomes of the human female: an in situ semi-quantitative analysis. Chromosoma *104*, 528-535.

Bestor, T. H., and Tycko, B. (1996). Creation of genomic methylation patterns. Nature Genetics *12*, 363-7.

Bey, L., Etienne, J., Tse, C., Brault, D., Noe, L., Raisonnier, A., Arnault, F., Hamilton, M. T., and Galibert, F. (1998). Cloning, sequencing and structural analysis of 976 base pairs of the promoter sequence for the rat lipoprotein lipase gene. Comparison with the mouse and human sequences. Gene *209*, 31-38.

Bird, A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends in Genetics *3*, 342-347.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Research *8*, 1499-504.

Bird, A. P. (1996). The relationship of DNA methylation to cancer. Cancer Surveys *28*, 87-101.

Bone, J. R., Lavender, J., Richman, R., Palmer, M. J., Turner, B. M., and Kuroda, M. I. (1994). Acetylated histone H4 on the male X chromosome is associated with dosage compensation in Drosophila. Genes-Dev *8*, 96-104.

Bonthron, D. T., Handin, R. I., Kaufman, R. J., Wasley, L. C., Orr, E. C., Mitsock, L. M., Ewenstein, B., Loscalzo, J., Ginsburg, D., and Orkin, S. H. (1986). Structure of a pre-pro-von Willebrand factor and its expression in heterologous cells. Nature *324*, 270-273.

Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H. F., Avner, P., and Ballabio, A. (1991). Characterization of a Murine Gene Expressed From the Inactive X- Chromosome. Nature *351*, 325-329.

Brickner, A. G., Koop, B. F., Aronow, B. J., and Wiginton, D. A. (1999). Genomic sequence comparison of the human and mouse adenosine deaminase gene regions. Mammalian Genome *10*, 95-101.

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D., and Rastan, S. (1991). Conservation of Position and Exclusive Expression of Mouse Xist From the Inactive X-Chromosome. Nature *351*, 329-331.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. (1992). The product of the mouse *Xist* gene is a 15Kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., and Willard, H. F. (1991). A Gene From the Region of the Human X-Inactivation Center Is Expressed Exclusively From the Inactive X-Chromosome. Nature *349*, 38-44.

Brown, C. J., Carrel, L., and Willard, H. F. (1997). Expression of genes from the human active and inactive X chromosomes. American Journal Of Human Genetics *60*, 1333-1343.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., and Lawrence, J. (1992). The human XIST gene: analysis of a 17Kb inactive X-specific RNA that contains conserved repeats and is highly localised within the nucleus. Cell *71*, 515-526.

Brown, C. J., Powers, V. E., Monroe, D. M., Sheinin, R., and Willard, H. F. (1989). A gene on the short arm of the human X-chromosome complements the murine tsA1S9 DNA synthesis mutation. Somatic Cell and Molecular Genetics *15*, 173-178.

Brown, G. K., Brown, R. M., and Dahl, H. H. M. (1990). The Human Pyruvate-Dehydrogenase El-Alpha Genes - Location, Gene Organization and Expression. Genetical Research *55*, 125.

Brown, G. K., Brown, R. M., Scholem, R. D., Kirby, D. M., and Dahl, H.-H. M. (1990). The clinical and biochemical spectrum of human pyruvate dehydrogenase deficiency. Annals of the New York Academy of Science *573*, 360-368.

Brown, R. M., Dahl, H. H. M., and Brown, G. K. (1990). Pyruvate-Dehydrogenase E1-Alpha Subunit Genes in the Mouse - Mapping and Comparison With Human Homologs. Somatic Cell and Molecular Genetics *16*, 487-492.

Brown, R. M., Dahl, H.-H. M., and Brown, G. K. (1989). X chromosome localisation of the functional gene for the E1α subunit of the human pyruvate dehydrogenase complex. Genomics *4*, 174-181.

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

Carrel, L., Clemson, C. M., Dunn, J. M., Miller, A. P., Hunt, P. A., Lawrence, J. B., and Willard, H. F. (1996). X inactivation analysis and DNA methylation studies of the ubiquitin activating enzyme E1 and PCTAIRE-1 genes in human and mouse. Human Molecular Genetics *5*, 391-401.

Carrel, L., Hunt, P. A., and Willard, H. F. (1996). Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene. Human Molecular Genetics *5*, 1361-1366.

Cattanach, B. M. (1961). XXY mice. Genetical Research *2*, 156-160.

Cattanach, B. M., and Rasberry, C. (1972). Evidence of non-random X-chromosome activity in the mouse. Genetical Research *19*, 229-240.

Chang, D. C. (1989). Cell poration and cell fusion using an oscillating electric field. Biophysical Journal *56*, 641-52.

Chang, M., Naik, S., Johanning, G. L., Ho, L., and Patel, M. S. (1993). Multiple Protein-Binding Domains and Functional cis-Elements in the 5'-Flanking Region of the Human Pyruvate Dehydrogenase α-Subunit Gene. Biochemistry *32*, 4263-4269.

Chaudhary, J., and Skinner, M. K. (1998). Comparative sequence analysis of the mouse and human transferrin promoters: Hormonal regulation of the transferrin promoter in Sertoli cells. Molecular Reproduction and Development *50*, 273-283.

Chen, Z. Y., Hendricks, R. W., A., J. M., Powell, J. F., Breakfield, X. O., Sims, K. B., and Craig, I. W. (1992). Isolation and characterisation of a candidate gene for Norrie disease. Nature Genetics *1*, 204-208.

Chen, Z. Y., Kramm, C. M., Battinelli, E. M., Li, F., Schuback, D. E., and Corey, D. P. (1995). Neuron-Specific Expression of the Norrie Disease Gene (Ndp) in the Central-Nervous-System. American Journal of Human Genetics *57*, 747.

Chien, A., Edgar, D. B., and Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. Journal of Bacteriology *127*, 1550-1557.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry *162*, 156-159.

Chuang, P. T., Albertson, D. G., and Meyer, B. J. (1994). DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X-chromosome. Cell *79*, 459-474.

Chun, K., MacKay, N., Willard, H. F., and Robinson, B. H. (1990). Isolation, characterisation and chromosomal allocation of cDNA clones for the $E_1\beta$ subunit of the pyruvate dehydrogenase complex. European Journal of Biochemistry *194*, 587-592.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. Proceedings of the National Academy of Sciences (USA) *81*, 1991-1995.

Clark, S. J. (1997). Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite treated DNA. Nucleic Acids Research *25*, 4422-4426.

Clark, S. J., and Frommer, M. (1997). Bisulphite Genomic Sequencing of Methylated cytosines. In Laboratory Methods for the detection of mutations and polymorphisms in DNA., G. Taylors, ed.: CRC-Press), pp. 151-162.

Clark, S. J., Harrison, J., and Molloy, P. L. (1997). Sp1 binding is inhibited by (m)Cp(m)CpG methylation. Gene *195*, 67-71.

Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. Nucleic Acids Research *22*, 2990-2997.

Clemson, C. M., McNeill, L. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X-chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosomal structure. Journal of Cell Biology *132*, 1-17.

Collins, J., and Bruning, H. J. (1978). Plasmids useable as gene-cloning vectors in an in vitro packaging by coliphage lambda: "cosmids". Gene *4*, 85-107.

Collins, J., and Hohn, B. (1978). Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. Proceedings of the National Academy of Sciences USA *75*, 4242-4246.

Comb, M., and Goodman, H. M. (1990). CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. Nucleic Acids Research *18*, 3975-82.

Cooper, D. W., Johnston, P. G., Watson, J. M., and Graves, J. A. M. (1993). X-inactivation in marsupials and monotremes. Seminars in Developmental Biology *4*, 117-128.

Cormack, B. P., et al. (1996). FACS-optimised mutatnts of the green fluorescent protein (GFP). Gene *173*, 33-38.

Costa, G. L., Grafsky, A., and Weiner, M. P. (1994). Cloning and Analysis of Pcr-Generated Dna Fragments. Pcr-Methods and Applications *3*, 338-345.

Costa, G. L., and Weiner, M. P. (1994). Polishing With T4 or Pfu Polymerase Increases the Efficiency of Cloning of Pcr Fragments. Nucleic Acids Research *22*, 2423.

Cote, S., Sinnett, D., and Momparler, R. L. (1998). Demethylation by 5-aza-2'-deoxycytidine of specific 5-methylcytosine sites in the promoter region of the retinoic acid receptor beta gene in human colon carcinoma cells. Anti-Cancer Drugs *9*, 743-750.

Courtier, B., Heard, E., and Avner, P. (1995). *Xce* haplotypes show modified methylation in a region of the active X-chromosome lying 3' to *Xist*. Proceedings of the National Academy of Sciences (USA) *92*, 3531-3535.

Cross, S. H., Charlton, J. A., Nan, X. S., and Bird, A. P. (1994). Purification of Cpg Islands Using a Methylated Dna-Binding Column. Nature Genetics *6*, 236-244.

Cross, S. H., Lee, M., Clark, V. H., Craig, J. M., Bird, A. P., and Bickmore, W. A. (1997). The chromosomal distribution of CpG islands in the mouse: Evidence for genome scrambling in the rodent lineage. Genomics *40*, 454-461.

Cullen, C. R., Hubberman, P., Kaslow, D. C., and Migeon, B. R. (1986). Comparison of factor IX methylation on human active and inactive X-chromosomes: implications for X-inactivation and transcription of tissue specific genes. EMBO Journal *5*, 2223-2229.

Curran, T., and Franza, B. R. (1988). Fos and Jun - the Ap-1 Connection. Cell *55*, 395-397.

Dahl, H. H. M., Brown, R. M., Hutchison, W. M., Maragos, C., and Brown, G. K. (1990). A Testis-Specific Form of the Human Pyruvate-Dehydrogenase E1-Alpha Subunit Is Coded For By an Intronless Gene On Chromosome-4. Genomics *8*, 225-232.

Dahl, H.-H. M., Hunt, S. M., Hutchison, W. M., and Brown, G. K. (1987). The human pyruvate dehydrogenase complex: isolation of cDNA clones for the E1α subunit, sequence analysis and characterisation of the mRNA. Journal of Biological Chemistry *262*, 7398-7403.

deCapoa, A., Menendez, F., Poggesi, I., Giancotti, P., Grappelli, C., Marotta, M. R., DiLeandro, M., Reynaud, C., and Niveleau, A. (1996). Cytological evidence for 5-azacytidine-induced demethylation of the heterochromatic regions of human chromosomes. Chromosome Research *4*, 271-276.

Dickie, M. M. (1954). The tortoiseshell house mouse. Journal of Heredity *45*, 158-159.

Disteche, C. M. (1995). Escape From X-Inactivation in Human and Mouse. Trends in Genetics *11*, 17-22.

Disteche, C. M., Eicher, E. M., and Latt, S. A. (1981). Late replication patterns in adult and embryonic mice carrying Searle's X-autosome translocation. Cell Research *133*, 357-362.

Distesch, C. M. (1997). The Great Escape. American Journal of Human Genetics *60*, 1312-1315.

Duthie, S. M., Nesterova, T. B., Formstone, E. J., Keohane, A. N., Turner, B. M., Zakian, S. M., and Brockdorff, N. (1999). Xist RNA exhibits a banded localisation on the inactive X-chromosome and is excluded from autosomal material *in-cis*. Human Molecular Genetics *8*, 195-204.

Eden, S., Hashimshony, T., Keshet, I., Cedar, H., and Thorne, A. W. (1998). DNA methylation models histone acetylation. Nature *394*, 842.

Ehrlich, M., Gama-Sosa, M. A., Huang, L., Midgett, R. M., Kuo, K. C., McCune, R. A., and Gehrke, C. (1982). Amount and distribution of 5-methyl-cytosine in human DNA from different types of tissues or cells. Nucleic Acids Research *10*, 2709-2721.

Ehrmann, I. E., Ellis, P. S., Mazeyrat, S., Duthie, S., Brockdorff, N., Mattei, M. G., Gavin, M. A., Affara, N. A., Brown, G. M., Simpson, E., Mitchell, M. J., and Scott, D. M. (1998). Characterization of genes encoding translation initiation factor eIF- 2 gamma in mouse and human: sex chromosome localization, escape from X-inactivation and evolution. Human Molecular Genetics *7*, 1725-1737.

Eickbusch, T. H., and Moudrianakis, E. N. (1978). The histone core complex: an octamer assembled by two sets of protein-protein interactions. Biochemistry *17*, 4955.

Eils, R., Dietzel, S., Bertin, E., Schrock, E., Speicher, M. R., Ried, T., RobertNicoud, M., Cremer, C., and Cremer, T. (1996). Three-dimensional reconstruction of painted human interphase chromosomes: Active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. Journal of Cell Biology *135*, 1427-1440.

Ellison, J. W., Passage, M. B., Yu, L. C., Yen, P. H., Mohandas, T. K., and Shapiro, L. J. (1992). Directed Isolation of Human Genes That Escape X-Inactivation. Clinical Research *40*, A9.

Emili, A., Greenblatt, J., and Ingles, C. J. (1994). Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box binding protein. Molecular Cell Biology *14*, 1582.

Erickson, R. P., Zwingman, T., and Ao, A. (1993). Gene expression, X-inactivation, and methylation during spermatogenesis: the case of Zfa, Zfx, and Zfy in mice. Mol-Reprod-Dev *35*, 114-20.

Esposito, T., Gianfrancesco, F., Ciccodicola, A., Desposito, M., Nagaraja, R., Mazzarella, R., Durso, M., and Forabosco, A. (1997). Escape from X inactivation of two new genes associated with DXS6974E and DXS7020E. Genomics *43*, 183-190.

Falls, H. F. (1951). Sex-linked ocular albinism displaying fundus changes in the female heterozygote. American Journal of Opthalmology *34 (Suppl.)*, 41-50.

Feil, R., Charlton, J., Bird, A. P., Walter, J., and Reik, W. (1994). Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. Nucleic Acids Research *4*, 695-696.

Felgner, P. L., Sridhar, C. N., Wheeler, C. J., and Felgner, J. (1993). Enhanced Gene Delivery and Mechanism Studies With a Novel Series of Cationic Lipid Formulations. Journal of Cellular Biochemistry, 206.

Feng, D. F., and Doolittle, R. F. (1987). Progressive sequence alignment as a pre-requisite to correct phylogenetic trees. Journal of Molecular Evolution *25*, 351-360.

Fialkow, P. J. (1970). X-chromosome inactivation and the Xg locus. American Journal of Human Genetics *22*, 460-463.

Fisher, E. M. C., Beerromero, P., Brown, L. G., Ridley, A., McNeil, J. A., Lawrence, J. B., Willard, H. F., Bieber, F. R., and Page, D. C. (1990). Homologous Ribosomal-Protein Genes On the Human X-Chromosome and Y- Chromosome - Escape From X-Inactivation and Possible Implications For Turner Syndrome. Cell *63*, 1205-1218.

Ford, C. E., Jones, K. W., Polani, P. E., De Almeida, J. C., and Briggs, J. H. (1959). A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). Lancet *1*, 711-713.

Fouque, F., Marsac, C., and Benelli, C. (1998). The pyruvate dehydrogenase complex: from molecular organization to clinical issues. M S-Medecine Sciences *14*, 1366-1374.

Fraser, A. S., Sobey, S., and Spicer, C. C. (1953). Mottled: a sex-modified lethal in the house mouse. Journal of Genetics *51*, 217-221.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992). A genomic seqencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences USA *89*, 1827-1831.

Fuscoe, J. C., Fenwick, R. G. J., Ledbetter, D. H., and Caskey, C. T. (1983). Deletion and amplification of the HGPRT locus in Chinese hamster cells. Molecular Cell Biology *3*, 1086-1096.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. Journal of Molecular Biology *196*, 261-282.

Gaston, K., and Fried, M. (1995). CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes. Nucleic Acids Research *23*, 901-9.

Goldman, M. A., Reeves, P. S., Wirth, C. M., Zupko, W. J., Wong, M. A., Edelhoff, S., and Disteche, C. M. (1998). Comparative methylation analysis of murine transgenes that undergo or escape X-chromosome inactivation. Chromosome Research *6*, 397-404.

Gonzalgo, M. L., Bender, C. M., You, E. H., Glendening, J. M., Flores, J. F., Walker, G. J., Hayward, N. K., Jones, P. A., and Fountain, J. W. (1997). Low frequency of p16/CDKN2A methylation in sporadic melanoma: Comparative approaches for methylation analysis of primary tumors. Cancer Research *57*, 5336-5347.

Goodfellow, P. J., Mondello, C., Darling, S. M., Pym, B., and Little, P. (1988). Absence of methylation of a CpG-rich region at the 5' end of the MIC2 gene on the active X, the inactive X and the Y-chromosome. Proceedings of the National Academy of Sciences (USA) *85*, 5605-5609.

Grant, M., Zuccotti, M., and Monk, M. (1992). Methylation of CpG sites of two X-linked genes coincides with X-inactivation in the female mouse embryo but not in the germline. Nature Genetics *2*, 161-166.

Graves, J. A. M. (1983). Inactivation and reactivation of the mammalian X-chromosome. In Development in Mammals, M. H. Johnson, ed. (Amsterdam: North-Holland publishing company), pp. 251-295.

Greenfield, A., Carrel, L., Pennisi, D., Philippe, C., Quaderi, N., Siggers, P., Steiner, K., Tam, P. P. L., Monaco, A. P., Willard, H. F., and Koopman, P. (1998). The UTX gene escapes X inactivation in mice and humans. Human Molecular Genetics *7*, 737-742.

Grigg, G., and Clark, S. (1994). Sequencing 5-methylcytosine residues in genomic DNA. Bioessays *16*, 431-436.

Groudine, M., and Conkin, K. F. (1985). Chromatin structure and *de novo* methylation in sperm DNA: implications for activation of the paternal genome. Science *228*, 1061-1068.

Habener, J. (1990). Cyclic AMP response element binding proteins: a cornucopia of transcription factors. Molecular Endocrinology *4*, 1087.

Hansen, R. S., Ellis, N. A., and Gartler, S. M. (1988). Demethylation of specific sites in the 5' region of the inactive X-linked human phosphoglycerate kinase gene correlated with the appearance of nuclease sensitivity and gene expression. Molecular Cell Biology *8*, 4692-4699.

Hansen, R. S., and Gartler, S. M. (1990). 5-Azacytidine-Induced Reactivation of the Human X-Chromosome-Linked Pgk1 Gene Is Associated With a Large Region of Cytosine Demethylation in the 5' Cpg Island. Proceedings of the National Academy of Sciences of the United States of America *87*, 4174-4178.

Harrington, M. A., Jones, P. A., Imagawa, M., and Karin, M. (1988). Cytosine methylation does not affect binding of transcription factor Sp1. Proceedings of the National Academy of Sciences (USA) *85*, 2066-2070.

Heard, E., Clerc, P., and Avner, P. (1997). X-Chromosome inactivation in mammals. Annual Review of Genetics *31*, 571-610.

Hebbes, T. R., Thorne, A. W., Clayton, A. L., and Crane-Robinson, C. R. (1992). Histone acetylation and globin gene switching. Nucleic Acids Research *20*, 1017-1022.

Hendrich, B. D., Plenge, R. M., and Willard, H. F. (1997). Identification and characterization of the human XIST gene promoter: Implications for models of X chromosome inactivation. Nucleic Acids Research *25*, 2661-2671.

Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., and Baylin, S. B. (1996). Methylation specific PCR: a novel PCR assay for methylation status of CpG-islands. Proceedings of the National Academy of Sciences (USA) *93*, 9821-9826.

Hermann, R., and Doerfler, W. (1991). Interference with protein binding at AP-2 sites by sequence-specific methylation in the late E2A promoter. FEBS letters *281*, 191-195.

Herzing, L. K. B., Romer, J. T., Horn, J. M., and Ashworth, A. (1997). *Xist* has properties of the X-chromosome inactivation centre. Nature *386*, 272-279.

Higgins, D. G., and Sharp, P. M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. Computer of Applied Bioscience *5*, 151-153.

Hoffman, J. L. (1986). Chromatographic Analysis of the Chiral and Covalent Instability of S-Adenosyl-L-Methionine. Biochemistry *25*, 4444-4449.

Hohn, B., and Collins, J. (1980). A small cosmid for efficient cloning of large DNA fragments. Gene *11*, 291-298.

Holler, M., Westin, G., Jiricny, J., and Schaffner, W. (1988). Sp1 transcription factor binds binds DNA and activates transcription even when the binding site is CpG methylated. Genes and Development *2*, 1127-1135.

Holliday, R., and Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226-232.

Horn, J. H., and Ashworth, A. (1995). A member of the caudal family of homeobox genes maps to the X-inactivation centre region of the mouse and human X-chromosomes. Human Molecular Genetics *4*, 1041-1047.

Hornstra, I. K., and Yang, T. P. (1994). High-resolution methylation analysis of the human hypoxanthine phosphoribosyltransferase gene 5' region on the active and inactive X chromosomes: correlation with binding sites for transcription factors. Mol-Cell-Biol *14*, 1419-30.

Hsieh, C. (1994). Dependence of Transcriptional Repression on CpG Methylation Density. Molecular and Cellular Biology *14*, 5487-5494.

Hsieh, C. (1997). Stability of Patch Methylation and Its Impact in Regions of Transcriptional Initiation and Elongation. Molecular and Cellular Biology *17*, 5897-5904.

Hsu, D. R., and Meyer, B. J. (1993). X-chromosome dosage compensation and its relationship to sex determination in *C.elegans*. Seminars in Developmental Biology *4*, 93-106.

Hu, G. X. (1993). Dna Polymerase-Catalyzed Addition of Nontemplated Extra Nucleotides to the 3' End of a Dna Fragment. Dna and Cell Biology *12*, 763-770.

Huntriss, J., Lorenzi, R., Purewal, A., and Monk, M. (1997). A methylation-dependent DNA-binding activity recognising the methylated promoter region of the mouse *Xist* gene. Biochemical and Biophysical Research Communications *235*, 730-738.

Iaonnou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A., and de Jong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large DNA fragments. Nature Genetics *6*, 84-89.

Iguchi-Ariga, S. M., and Schaffner, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes And Development *3*, 612-9.

Iguchi-Ariga, S. M. M., and Scaffner, W. (1989). CpG methylation of the cAMP-responsive-element enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes and Development *3*, 612-619.

Jaenisch, R. (1997). DNA methylation and imprinting: why bother? Trends In Genetics *13*, 323-9.

Jahner, D., and Jaenisch, R. (1984). DNA methylation in early mammalian development. In DNA Methylation: Biochemistry and Biological Significance, A. Razin, H. Cedar and A. D. Riggs, eds. (New York: Springer Verlag), pp. 189-219.

Jeppesen, P., and Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell *74*, 281-9.

John, S. W. M., Weitzner, G., Rozen, R., and Scriver, C. R. (1991). A rapid procedure for extracting genomic DNA from leukocytes. Nucleic Acids Research *19*, 408.

Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature Genetics *19*, 187-191.

Kass, S. U., Pruss, D., and Wolffe, A. P. (1997). How does DNA methylation repress transcription ? Trends in Genetics *13*, 444-449.

Kay, G. F., Penny, G. D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. (1993). Expression of Xist During Mouse Development Suggests a Role in the Initiation of X-Chromosome Inactivation. Cell *72*, 171-182.

Keith, D. H., Singersam, J., and Riggs, A. D. (1986). Active X-Chromosome Dna Is Unmethylated At 8 Ccgg Sites Clustered in a Guanine-Plus-Cytosine-Rich Island At the 5' End of the Gene For Phosphoglycerate Kinase. Molecular and Cellular Biology *6*, 4122-4125.

Kenyon, J. R., and Craig, I. W. (1999). Analysis of the 5' regulatory region of the human Norrie's disease gene: evidence that a non-translated CT dinucleotide repeat in exon one has a role in controlling expression. Gene *227*, 181-188.

Kerppola, T. K., and Curran, T. (1994). Maf and Nrl can bind AP-1 sites and form heterodimers with Fos and Jun. Oncogene *9*, 3149-3158.

Kido, M., Rich, K. A., Lang, G., Barron, E., Kohn, D. B., AlUbaidi, M. R., and Blanks, J. C. (1996). Use of a retroviral vector with an internal opsin promoter to direct gene expression to retinal photoreceptor cells. Current Eye Research *15*, 833-844.

Kobayashi, T., Urakami, S., Cheadle, J. P., Aspinwall, R., Harris, P., Sampson, J. R., and Hino, O. (1997). Identification of a leader exon and a core promoter for the rat tuberous sclerosis 2 (Tsc2) gene and structural comparison with the human homolog. Mammalian Genome *8*, 554-558.

KokaljVokac, N., Zagorac, A., Pristovnik, M., Bourgeois, C. A., and Dutrillaux, B. (1998). DNA methylation of the extraembryonic tissues: an in situ study on human metaphase chromosomes. Chromosome Research *6*, 161-166.

Labella, F., and Heintz, N. (1991). Histone Gene-Transcription Factor Binding in Extracts of Normal Human-Cells. Molecular and Cellular Biology *11*, 5825-5831.

Larin, Z., Monaco, A. P., and Lehrach, H. (1991). Yeast artificial chromsome libraries containing large inserts from mouse and human DNA. Proceedings of the National Academy of Sciences (USA) *88*, 4123-4127.

Lee, J. T., Davidlow, L. S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nature Genetics *21*, 400-404.

Lee, J. T., Strauss, W. M., Dausman, J. A., and Jaenisch, R. (1996). A 450Kb transgene displays properties of the mammalian X-inactivation centre. Cell *86*, 83-94.

Leonhardt, H., Page, A. W., Weier, H. U., and Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865-73.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362-5.

Li, X. M., Alperin, E. S., Salido, E., Gong, Y., Yen, P., and Shapiro, L. J. (1996). Characterization of the promoter region of human steroid sulfatase: A gene which escapes X inactivation. Somatic Cell and Molecular Genetics *22*, 105-117.

Lifschytz, E., and Lindsley, D. L. (1972). The role of X-chromosome inactivation during spermatogenesis. Proceedings of the National Academy of Sciences (USA) *69*, 182-186.

Lin, C. S., Chang, C. H., and Huynh, T. (1997). The murine L-plastin gene promoter: Identification and comparison with the human L-plastin gene promoter. Dna and Cell Biology *16*, 9-16.

Linn, T. C., Pettit, F. H., and Reed, L. J. (1969). α-Keto acid dehydrogenase complexes, X-regulation of the activity of the pyruvate dehydrogenase complex from beef kidney mitochondria by phosphorylation and dephosphorlation. Proceedings of the National Academy of Sciences (USA) *62*, 234-241.

Liu, Q., Ji, X., Breitman, J. B., Hitchcock, P. F., and Swaroop, A. (1996). Expression of the bZIP transcription factor gene *Nrl* in the developing nervous system. Oncogene *12*, 207-211.

Lock, L. F., Takagi, N., and Martin, G. R. (1987). Methylation of the Hprt Gene On the Inactive-X Occurs After Chromosome Inactivation. Cell *48*, 39-46.

Loebel, D. A. F., and Johnston, P. G. (1996). Methylation analysis of a marsupial X-linked CpG-island by bisulphite genomic sequencing. Genome Research *6*, 114-123.

Lu, S., and Davies, P. J. A. (1997). Regulation of the expression of the tissue transglutaminase gene by DNA methylation. Proceedings of the National Academy of Sciences (USA) *94*, 4692-4697.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). X-ray structire of the nucleosome core-particle at 2.8 amstrong resolution. Nature *389*, 251-259.

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus L.*). Nature *190*, 372-373.

Lyon, M. F. (1961). Sex chromatin and gene action in the mammalian X-chromosome. , 135-148.

Lyon, M. F. (1998). X-chromosome inactivation: a repeat hypothesis. Cytogenetics and Cell Genetics *80*, 133-137.

Lyon, M. F., Searle, A. G., Ford, C. E., and Ohno, S. (1964). A mouse translocation suppressing sex-linked variegation. Cytogenetics *3*, 306-323.

Maizel, J. V., and Lenk, R. P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proceedings of the National Academy of Sciences (USA) *78*, 7665-7669.

Maragos, C., Hutchison, W. M., Hayasaka, K., Brown, G. K., and Dah, H. H. M. (1989). Structural Organization of the Gene For the E1-Alpha-Subunit of the Human Pyruvate-Dehydrogenase Complex. Journal of Biological Chemistry *264*, 12294-12298.

Marahrens, Y., Loring, J., and Jaenisch, R. (1998). Role of the *Xist* gene in X-chromosome choosing. Cell *92*, 657-664.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. Genes and Development *11*, 156-166.

Matsuzawa, T. (1991). Appearance of Autosome Linked G6pd Isozymes in Rat Sperm During Passage Through the Epididymis. Journal of Veterinary Medical Science *53*, 733-735.

McCarrey, J. R., and Thomas, K. (1987). Human testis specificPGK gene lacks introns and possesses characetistics of a processed gene. Nature *326*, 501-505.

McClelland, M., Nelson, M., and Raschke, E. (1994). Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Research *22*, 3640-3659.

McCullogh, K. C., and Spier, R. E. (1990). Chapters 2 and 3. In Monoclonal antibodies in biotechnology: Theoretical and practical aspects, J. Baddiley, N. H. Carey, I. J. Higgins and W. G. Potter, eds. (Cambridge: Cambridge University Press).

McDonald, L. E., and Kay, G. F. (1997). Methylation analysis using bisulfite genomic sequencing: Application to small numbers of intact cells. Biotechniques *22*, 272 (2 pages).

McFall, R. C. e. a. (1977). Characterization of a new continuous cell line derived from a human retinoblastoma. Cancer Research *37*, 1003-1010.

McMahon, A., and Monk, M. (1983). X-chromosome activity in female mouse embryos heterozygous for *Pgk-1* and Searle's translocation, T(X;16)16H. Genetical Research *41*, 69-83.

Meehan, R. R., Lewis, J. D., and Bird, A. P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Research *20*, 5085-92.

Meitinger, T., Meindl, A., Bork, P., Rost, B., Sander, C., Haasemann, M., and Murken, J. (1993). Molecular modelling of the Norrie disease protein reveals a cystine knot growth factor tertiary structure. Nature Genetics *5*, 376-380.

Meitinger, T., Meindl, A., Bork, P., Sander, C., and Murken, J. (1993). Norrie Disease - Mutation Analysis, Protein-Structure and Expression. American Journal of Human Genetics *53*, 1202.

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.

Migeon, B. R. (1994). X-chromosome inactivation: molecular mechanisms and genetic consequences. Trends in Genetics *10*, 230-235.

Millar, D. S., Krawczak, M., and Cooper, D. N. (1998). Variation of site-specific methylation patterns in the factor VIII (F8C) gene in human sperm DNA. Human Genetics *103*, 228-233.

Miller, A. P., and Willard, H. F. (1995). 5 Genes That Escape X-Chromosome Inactivation Are Clustered Within 1 Mb in Xp11.21. American Journal of Human Genetics *57*, 147.

Mizugaki, M., Itoh, K., Yamaguchi, T., Ishiwata, S., Hishinuma, T., Nozaki, S., and Ishida, N. (1996). Preparation of a monoclonal antibody specific for 5-methyl-2'- deoxycytidine and its application for the detection of DNA methylation levels in human peripheral blood cells. Biological & Pharmaceutical Bulletin *19*, 1537-1540.

Mohandas, T., Sparkes, R. S., and Shapiro, L. J. (1981). Reactivation of an inactive human X-chromosome: evidence for X-inacitvation by DNA methylation. Science *211*, 393-396.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extra-embryonic and germ cell lineages during mouse embryo development. Development.

Mullens, L. J., Veres, G., Caskey, C. T., and Chapman, V. (1987). Differential methylation of the ornithine carbamoyl transferase gene on the active and inactive mouse X-chromosomes. Molecular and Cellular Biology *7*, 3916-3922.

Nan, X., Campoy, F. J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471-81.

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex [see comments]. Nature *393*, 386-9.

Norrie, G. (1927). Causes of blindness in children. Acta Opthalmology *5*, 357-386.

Norris, D. P., Brockdorff, H. N., and Rastan, S. (1991). Methylation analysis of CpG rich bands on the active and inactive mouse X chromosomes. Mammalian Genome *1*, 78-83.

Norris, D. P., Patel, D., Kay, G. F., Penny, G. D., Brockdorff, N., and Sheardown, S. A. (1994). Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell *77*, 41-51.

Novotny, V., Prieschl, E. E., Csonga, R., Fabjani, G., and Baumruker, T. (1998). Nrf1 in a complex with fosB, c-jun, junD and ATF2 forms the AP1 component at the TNF alpha promoter in stimulated mast cells. Nucleic Acids Research *26*, 5480-5485.

O'Brein, T. P., Yang, G. P., Sanders, L., and Lau, L. F. (1990). Expression of cyr61, a growth factor inducible immediate early gene. Molecular and cellular Biology *10*, 3569-3577.

Ohno, S., and Hauschka, T. S. (1960). Allocycly of the X-chromosome in tumors and normal tissues. Cancer Research *20*, 541-545.

Ohno, S., Kaplan, W. D., and Kinosita, R. (1959). Formtion of sex-chromatin by a single X-chromosome in liver cells of *Rattus norvegicus*. Experimental Cell Research *18*, 415-418.

Ohno, S., and Makino, S. (1961). The single X-nature of sex chromatin in man. Lancet *1*, 78-79.

Okada, Y. (1993). Sendai virus-induced cell fusion. Methods In Enzymology *221*, 18-41.

Olek, A., Oswald, J., and Walter, J. (1996). A modified and improved method for bisulphite based cytosine methylation analysis. Nucleic-Acids-Res *24*, 5064-6.

Panning, B., Dausman, J., and Jaenisch, R. (1997). X-chromosome inactivation is mediated by *Xist* RNA stabilisation. Cell *90*, 907-916.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate *Xist* expression and silence X-linked genes. Genes and Development *10*, 1991-2002.

Park, J. G., and Chapman, V. M. (1994). CpG island promoter region methylation patterns of the inactive-X-chromosome hypoxanthine phosphoribosyltransferase (Hprt) gene. Mol-Cell-Biol *14*, 7975-83.

Paroush, Z., Keshet, I., Yisraeli, J., and Cedar, H. (1990). Dynamics of demethylation and activationof the α–actin gene in myoblasts. Cell *63*, 1229-1237.

Paul, C. L., and Clark, S. J. (1996). Cytosine methylation: quantitation by automated genomic sequencing and GENESCAN analysis. Biotechniques *21*, 126-33.

Paulin, R., Grigg, G. W., Davey, M. W., and Piper, A. A. (1998). Urea improves efficiency of bisulphite-mediated sequencing of 5'- methylcytosine in genomic DNA. Nucleic Acids Research *26*, 5009-5010.

Pazin, M. J., and Kadonaga, J. T. (1997). What's up and down with histone deacetylation and transcription ? Cell *89*, 315-328.

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., and Brockdorff, N. (1996). Requirement for *Xist* in X-chromosome inactivation. Nature *379*, 131-137.

Perey, B., Clermont, Y., and Leblond, C. P. (1961). The waves of the seminiferous epithelium in the rat. American Journal of Anatomy *108*, 47-77.

PerezCastro, A. V., Wilson, J., and Altherr, M. R. (1997). Genomic organization of the human fibroblast growth factor receptor 3 (FGFR3) gene and comparative sequence analysis with the mouse fgfr3 gene. Genomics *41*, 10-16.

Pfeifer, G. P., Steigerwald, S., Hansen, R. S., Gartler, S. M., and Riggs, A. D. (1990). Polymerase chain reaction aided genomic sequencing of an X-chromosome linked CpG-island: methylation patterns suggest clonal inheritance, CpG site autonomy and an explanation of activity state stability. Proceedings of the National Academy of Sciences (USA) *87*, 8252-8256.

Pfeifer, G. P., Steigerwald, S. D., Mueller, P. R., Wold, B., and Riggs, A. D. (1989). Genomic sequencing and methylation analysis by ligation mediated PCR. Science *210*, 604-610.

Pfeifer, G. P., Tanguay, R. L., Steigerwald, S. D., and Riggs, A. D. (1990). In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. Genes-Dev *4*, 1277-87.

Pierce, J. C., Sauer, B., and Sternberg, N. (1992). A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficiency. Proceedings of the National Academy of Sciences USA *89*, 2056-2060.

Pillet, N., Bonny, C., and Schorderet, D. F. (1995). Characterization of the promoter region of the mouse Xist gene. Proceedings of the National Academy of Sciences of the United States of America *92*, 12515-12519.

Prantera, G., and Ferraro, M. (1990). Analysis of methylation and distribution of CpG sequences on human active X-chromosomes by in-situ nick translation. Chromosoma *99*, 18-23.

Prestridge, D. S. (1991). SignalScan : A computer program that scans DNA sequences for eukaryotic transcriptional elements. Computer Applied Biosciences *7*, 203-206.

Quesnel, B., Guillerm, G., Vereecque, R., Wattel, E., Preudhomme, C., Bauters, F., Vanrumbeke, M., and PierreFenaux, P. (1998). Methylation of the p15(INK4a) gene in myelodysplastic syndromes is frequent and acquired during disease progression. Blood *91*, 2985-2990.

Radtke, F., Hug, M., Georgiev, O., Matsuo, K., and Schaffner, W. (1996). Differential sensitivity of zinc finger transcription factors MTF-1, Sp1 and Krox-20 to CpG methylation of their binding sites. Biological Chemistry Hoppe-Seyler *377*, 47-56.

Rahe, B., Erickson, R. P., and Quinto, M. (1983). Methylation of unique sequence DNA during spermatogenesis in mice. Nucleic Acids Research *11*, 7974-7962.

Raizis, A. M., Scmitt, F., and Jost, J. P. (1995). A bisulfite method of 5-methylcytosine mapping that minimises template degradation. Analytical Biochemistry *226*, 161-166.

Randle, R. J., Sugden, P. H., Kirbey, A. L., Radcliffe, P. M., and Hutson, N. J. (1978). Regulation of pyruvate oxidation and the conservation of glucose. Journal of Paediatrics *111*, 525-533.

Ransone, L. J., Kerr, L. D., Schmitt, M. J., Wamsley, P., and Verma, I. M. (1993). The bZIP domains of Fos and Jun mediate a physical association with the TATA box binding protein. Gene Expression *3*, 37.

Rappold, G. A. (1993). The pseudoautosomal regions of the human sex chromosomes. Human Genetics *92*, 315-324.

Rastan, S. (1983). Non-Random X-Chromosome Inactivation in Mouse X-Autosome Translocation Embryos - Location of the Inactivation Center. Journal of Embryology and Experimental Morphology *78*, 1-22.

Rastan, S., and Robertson, E. J. (1985). X-Chromosome Deletions in Embryo-Derived (Ek) Cell-Lines Associated With Lack of X-Chromosome Inactivation. Journal of Embryology and Experimental Morphology *90*, 379-388.

Razin, A., and Cedar, H. (1991). DNA methylation and gene expression. Microbiological Reviews *55*, 451-8.

Razin, A., Webb, C., Szyf, M., Yisraeli, J., Rosenthal, A., Naveh-Many, T., Sciaky-Gallili, N., and Cedar, H. (1984). Variations in DNA methylation during mouse cell differentiation *in vivo* and *in vitro*. Proceedings of the National Academy of Sciences (USA) *81*, 2275-2279.

Riggs, A. D. (1975). X-inactivation differentiation and DNA methylation. Cytogenetics and Cell Genetics *14*, 9-25.

Robinson, B. H., MacKay, N., Chun, K., and Ling, M. (1996). Disorders of pyruvate carboxylase and the pyruvate dehydrogenase complex. Journal of Inherited Metabolic Disease *19*, 452-462.

Rothberg, J. M., Jacobs, J. R., Goodman, C. S., and Artavanic-Tsakonas, S. (1990). Slit: an extracellular protein necessary for development of midline glia and commisural axon pathways contains both EGF and LRR domains. Genes and Development *4*, 2169-2187.

Rougeulle, C., and Avner, P. (1996). Cloning and characterization of a murine brain specific gene Bpx and its human homologue lying within the Xic candiate region. Human Molecular Genetics *5*, 41-9.

Rougier, N., Bourchis, D., Gomes, D. M., Niveleau, A., Plachot, M., Paldi, A., and ViegasPequignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes & Development *12*, 2108-2113.

Russell, L. B., and Bangham, J. W. (1960). Further analysis of variegated-type position effects from X-autosome translocations in the mouse. Genetics *45*, 1008-1009.

Russell, L. B., and Major, M. H. (1957). Radiation-induced presumed somatic mutations in the house mouse. Genetics *42*, 161-175.

Russell, L. B., and Mongomery, C. S. (1965). The use of mouse X-autosome translocations in locating the X-chromosome inactivation center. Genetics *52*, 470.

Sadri, R., and Hornsby, P. J. (1996). Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. Nucleic Acids Research *24*, 5058-5059.

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

Salido, L. J., Yen, P. H., Mohandas, T. K., and Shapiro, L. J. (1992). Expression of the X-inactivation associated gene Xist during spermatogenesis. Nature Genetics *2*, 196-199.

Saluz, H., and Jost, J.-P. (1993). Major Techniques to study DNA methylation. In DNA methylation: Molecular Biology and biological significance, H. Saluz and J.-P. Jost, eds. (Basel: Birkhauser Verlag), pp. 11-26.

Salvatore, P., Benvenuto, G., Caporaso, M., Bruni, C. B., and Chiariotti, L. (1998). High resolution methylation analysis of the galectin-1 gene promoter region in expressing and nonexpressing tissues. FEBS Letters *421*, 152-158.

Sanford, J., Forrester, L., Chapman, V., Chandley, A., and Hastie, N. (1984). Methylation patterns of repetitive DNA sequences in the germ cells of *Mus. musculus*. Nucleic Acids Research *12*, 2823-2836.

Sanford, J. P., Chapman, V. M., and Rossant, J. (1985). DNA methylation in extra-embryonic lineages of mammals. Trends in Genetics, 89-93.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences (USA) *74*, 5463-54367.

Schwemmle, S., deGraaff, E., Deissler, H., Glaser, D., Wohrle, D., Kennerknecht, I., Just, W., Oostra, B. A., Dorfler, W., Vogel, W., and Steinbach, P. (1997). Characterization of FMR1 promoter elements by in vivo - Footprinting analysis. American Journal of Human Genetics *60*, 1354-1362.

Sheardown, S., Norris, D., Fisher, A., and Brockdorff, N. (1996). The mouse Smcx gene exhibits developmental and tissue specific variation in degree of escape from X inactivation. Human Molecular Genetics *5*, 1355-1360.

Sheardown, S. A., Duthie, S. M., Johnston, C. M., Newall, A. E. T., Formstone, E., Arkell, R. M., Nesterova, T. B., Alghisi, G.-C., Rastan, S., and Brockdorff, N. (1997b). Stabilisation of Xist RNA mediates initiation of X-chromosome inactivation. Cell *91*, 99-107.

Sheardown, S. A., Newall, A. E. T., Norris, D. P., Rastan, S., and Brockdorff, N. (1997a). Regulatory elements in the minimal promoter region of the mouse Xist gene. Gene *203*, 159-168.

Shizuya, H., Birren, B., Kim, U., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor based vector. Proceedings of the National Academy of Sciences USA *89*, 8794-8797.

Simmler, M. C., Cattanach, B. M., Rasberry, C., Rougeulle, C., and Avner, P. (1993). Mapping the murine *Xce* locus with (CA)n repeats. Mammalian Genome *4*, 523-530.

Simmler, M. C., Cunningham, D. B., Clerc, P., and Vermat, T. (1996). A 94kb genomic sequence 3' to the murine Xist gene reveals an AT-rich region containing a new testis specific gene Tsx. Human Molecular Genetics *11*, 1713-1726.

Simmler, M. C., Heard, E., Rougeulle, C., Cruaud, C., Weissenbach, J., and Avner, P. (1997). Localization and expression analysis of a novel conserved brain expressed transcript, brx/BRX lying within the Xic/XIC candidate region. Mammalian Genome *8*, 760-6.

Singer-Sam, J., Grant, M., Le Bon, J. M., Okuyama, K., Chapman, V., Monk, M., and Riggs, A. D. (1990). Use of a HpaII-polymerase chain reaction assay to study DNA methylation of in the Pgk-1 CpG-island of mouse emnbryos at the time of X-inactivation. Molecular and Cellular Biology *10*, 4987-4989.

Slim, R., Levilliers, J., Ludecke, H. J., Claussen, U., and Nguyen, V. C. (1993). A human pseudoautosomal gene encodes the ANT3 ADP/ATP translocase and escapes X-inactivation. Genomics *16*, 26-33.

Solage, A., and Cedar, H. (1978). Organisation of 5-Methylcytosine in Chromosomal DNA. Biochemistry, 2934-2938.

Somma, M. P., Gambino, I., and Lavia, P. (1991). Transcription Factors Binding to the Mouse Htf9 Housekeeping Promoter Differ Between Cell-Types. Nucleic Acids Research *19*, 4451-4458.

Southern, E. M. (1975). Detection of species specific sequences among DNA fragments separated by electrophoresis. Jounal of Molecular Biology *98*, 503-517.

Sternberg, N. (1990). Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proceedings Of The National Academy Of Sciences Of The United States Of America *87*, 103-7.

Stoger, R., Kajimura, T. M., Brown, W. T., and Laird, C. D. (1997). Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. Human Molecular Genetics *6*, 1791-1801.

Strunnikov, A. V. (1998). SMC proteins and chromosome structure. Trends in Cell Biology *8*, 454-459.

Sultana, R., Adler, D. A., Edelhoff, S., Carrel, L., Lee, K. H., Chapman, V. C., Willard, H. F., and Disteche, C. M. (1995). The Mouse Sb1.8 Gene Located At the Distal End of the X-Chromosome Is Subject to X-Inactivation. Human Molecular Genetics *4*, 257-263.

Swaroop, A., Xu, J. Z., Pawar, H., Jackson, A., Skolnick, C., and Agarwal, N. (1992). A conserved retina-specific gene encodes a basic motif/leucine zipper domain. Proceedings of the National Academy of Sciences (USA) *89*, 266-270.

Tabor, S., and Richardson, C. C. (1989). Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and Escherichia coli DNA polymerase I. Proceedings of the National Academy of Sciences (USA) *86*, 4076-4080.

Tada, T., and Takagi, N. (1992). Early development and X-chromosome inactivation in mouse parthenogenetic embryos. Molecular Reproduction and Development *31*, 20-27.

Takagi, N. (1980). Primary and secondary nonrandom X chromosome inactivation in early female mouse embryos carrying Searle's translocation T(X;16)16H. Chromosoma *81*, 439-459.

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X-chromosome in the extraembryonic membranes of the mouse. Nature *256*, 640-642.

Tan, J., Yang, H. S., and Patel, M. S. (1998). Regulation of mammalian pyruvate dehydrogenase alpha subunit gene expression by glucose in HepG2 cells. Biochemical Journal *336*, 49-56.

Tan, S.-S., Williams, E. A., and Tam, P. P. L. (1993). X-chromosome inactivation occurs at different times in different tissues of the post-implantation mouse embryo. Nature Genetics *3*, 170-174.

Tate, P. H., and Bird, A. P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Current Opinion In Genetics And Development *3*, 226-31.

Toussaint, C., Bousquetlemercier, B., Garlatti, M., Hanoune, J., and Barouki, R. (1994). Testis-Specific Transcription Start Site in the Aspartate- Aminotransferase Housekeeping Gene Promoter. Journal of Biological Chemistry *269*, 13318-13324.

Tribioli, C., Tamanini, F., Patrosso, C., Minalesi, L., and Villa, A. (1992). Methylation and sequence analysis around Eag I sites: identification of 28 new CpG islands in Xq24-Xq28. Nucleic Acids Research *20*, 727-733.

Tugores, A., Magness, S. T., and Brenner, D. A. (1994). A Single Promoter Directs Both Housekeeping and Erythroid Preferential Expression of the Human Ferrochelatase Gene. Journal of Biological Chemistry *269*, 30789-30797.

Ura, K., Kurumizaka, H., Dimitrov, S., Almouzni, G., and Wolffe, A. P. (1997). Histone acetylation: influence on transcription, nucleosome mobility and positioning and linker histone-dependent transcriptional repression. EMBO Journal *16*, 2096-2107.

Viegas-Pequignot, E., Dutrillaux, B., and Thomas, G. (1988). Inactive X-chromosome has the highest concentration of unmethylated HhaI sites. Proceedings of the National Academy of Sciences (USA) *85*, 7657-7660.

Vila, V., Jimenez, O., Guell, A., and Castrillo, J. L. (1993). Pit-1 and Pit-2 Role in Growth-Hormone Gene-Regulation. Journal of Pediatric Endocrinology *6*, 225-228.

Voss, K. O., Roos, K. P., Nonay, R. L., and Dovichi, N. J. (1998). Combating PCR bias in bisulphite based cytosine methylation analysis. Betaine modified cytosine deamination PCR. Analytical Chemistry *70*, 3818-3823.

Walker, C. L., Cargile, C. B., Floy, K. M., Delannoy, M., and Migeon, B. R. (1991). The Barr body is a looped X-chromosome formed by telomere association. Proceedings of the National Academy of Sciences USA *88*, 6191-6195.

Warburg, M. (1966). Norrie's disease. A congenital progressive oculo-acoustico-cerebral degeneration. Acta Opthalmology *89*, 1-47.

Watt, F., and Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes And Development *2*, 1136-43.

Weiss, A., Keshet, I., Razin, A., and Cedar, H. (1996). DNA demethylation In Vitro: Involvement of RNA. Cell *86*, 709-718.

Welshons, J. F., and Russell, L. B. (1959). The Y-chromosome as the bearer of male-determining factors in the mouse. Proceedings of the National Academy of Sciences USA *45*, 560-566.

Widom, J. (1989). Toward a unified model of chromatin folding. Annual Review of Biophysical Chemistry *18*, 365.

Wigler, M., Levy, D., and Perucho, M. (1981). The somatic replication of DNA methylation. Cell *24*, 33-40.

Willard, H. F. (1995). The sex chromosomes and X-chromosome inactivation. In The metabolic and Molecular Basis of Inherited Disease, C. R. Scriver, A. F. Beaudet, W. S. Sly and D. Valle, eds. (New York: McGraw-Hill), pp. 719-737.

Williams, T., Admon, A., Luscher, B., and Tjian, R. (1988). Cloning and expression of AP-2, a cell-type-specific transcription factor that activates inducible enhancer elements. Genes and Development *2*, 1557.

Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T., and Migeon, B. R. (1984). Methylation of the Hypoxanthine Phosphoribosyltransferase Locus On the Human X-

Chromosome - Implications For X-Chromosome Inactivation. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences *81*, 2806-2810.

Wolffe, A. P. (1997). Sinful repression. Nature *387*, 16-17.

Wolffe, A. P., and Hayes, J. J. (1999). Chromatin disruption and modification. Nucleic Acids Research *27*, 711-720.

Wong, D. J., Barrett, M. T., Stoger, R., Emond, M. J., and Reid, B. J. (1997). p16INK4a promoter is hypermethylated at a high frequency in esophageal adenocarcinomas. Cancer-Res *57*, 2619-22.

Woodcock, D. M., Crowther, P. J., and Diver, W. P. (1987). The majority of methylated deoxycytidines in human DNA are not in the CpG di-nucleotide. Biochemical and Biophysical Research Communications *145*, 888-894.

Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P., and Warren, W. D. (1997). Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. J-Biol-Chem *272*, 7810-6.

Wu, J. S., Ellison, J., Salido, E., Yen, P., Mohandas, T., and Shapiro, L. J. (1994). Isolation and Characterization of Xe169, a Novel Human Gene That Escapes X-Inactivation. Human Molecular Genetics *3*, 153-160.

Xiong, Z. G., and Laird, P. W. (1997). COBRA: A sensitive and quantitative DNA methylation assay. Nucleic Acids Research *25*, 2532-2534.

Yang, T. T., et al. (1996). Optimised codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. Nucleic Acids Research *24*, 4592-4593.

Yievin, A., and Razin, A. (1993). Gene methylation patterns and expression. In DNA Methylation: Molecular Biology and Biological Significance, J. P. Jost and H. P. Saluz, eds. (Basel: Birkhauser Verlag), pp. 523-568.

Yisraeli, J., Frank, D., Razin, A., and Cedar, H. (1988). Effect of in vitro DNA methylation on beta-globin gene expression. Proceedings Of The National Academy Of Sciences Of The United States Of America *85*, 4638-42.

Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends In Genetics *13*, 335-40.

Young, L. J., Huot, B., Nilsen, R., Wang, Z. X., and Insel, T. R. (1996). Species differences in central oxytocin receptor gene expression: Comparative analysis of promoter sequences. Journal of Neuroendocrinology *8*, 777-783.

Zuccotti, M., and Monk, M. (1995). Methylation of the mouse Xist gene in sperm and eggs correlates with imprinted Xist expression and paternal X-inactivation. Nature Genetics *9*, 316-320.